# PromptCanvas: Composable Prompting Workspaces Using Dynamic Widgets for Exploration and Iteration in Creative Writing

RIFAT MEHREEN AMIN, LMU Munich, Germany
OLIVER HANS KÜHLE, LMU Munich, Germany
DANIEL BUSCHEK, University of Bayreuth, Germany
ANDREAS BUTZ, LMU Munich, Germany

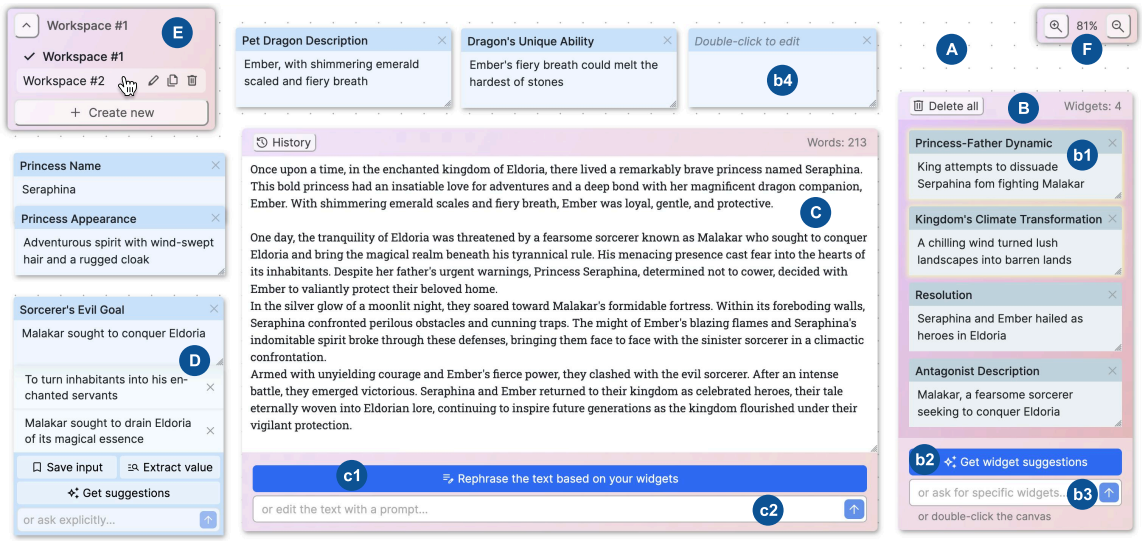


Fig. 1. User interface of PromptCanvas. (A) Canvas-like workspace where users can place and freely organize widgets to create a customized environment. (B) Widget panel. (b1) Example of widgets created through system suggestions. (b2) Button to get widget suggestions from the system. (b3) Field for entering prompts to create multiple widgets of a specific theme. (b4) Example of an empty widget created by double-clicking at empty space. (C) Text editor and output of text generation. (c1) Button to rephrase the text based on the widgets on the canvas (light blue). (c2) Field to provide prompts for generating text. (D) Example of an opened widget with suggested values for customization. (E) Menu bar for creating, renaming, duplicating, or deleting a canvas. (F) Panel with alternative zoom level controls.

We introduce PromptCanvas, a UI concept that transforms prompting into a composable, widget-based experience on an infinite canvas. Users can generate, customize, and arrange interactive widgets that represent various facets of their text, offering greater control over AI-generated content. PromptCanvas allows to create widgets through system suggestions, user prompts, or manual

Authors' Contact Information: Rifat Mehreen Amin, LMU Munich, Munich, Germany, rifat.amin@ifi.lmu.de; Oliver Hans Kühle, LMU Munich, Munich, Germany, o.kuehle@campus.lmu.de; Daniel Buschek, University of Bayreuth, Department of Computer Science, Bayreuth, Germany, daniel.buschek@uni-bayreuth.de; Andreas Butz, LMU Munich, Munich, Germany, butz@ifi.lmu.de.


Manuscript submitted to ACM

input, providing a flexible environment tailored to individual needs. This enables deeper engagement with the creative process. In two lab studies, PromptCanvas outperformed both the conversational user interface (lab study 1) and the structured baseline, Wordcraft (lab study 2) on the Creativity Support Index. Participants found that it reduced cognitive load, with better performance. Qualitative feedback revealed that the visual organization of thoughts and easy iteration encouraged new perspectives and ideas. The field study ($N = 10$) also confirmed these results, showcasing the potential of dynamic, customizable interfaces to improve collaborative writing with AI.

CCS Concepts: • **Human-centered computing** → **Graphical user interfaces**; **Natural language interfaces**; **Interactive systems and tools**.

Additional Key Words and Phrases: Dynamic UI, Prompting, LLM, human-AI co-creation, creativity support, metacognition, cognitive load, dynamic widgets, creative writing

**ACM Reference Format:**
Rifat Mehreen Amin, Oliver Hans Kühle, Daniel Buschek, and Andreas Butz. 2026. PromptCanvas: Composable Prompting Workspaces Using Dynamic Widgets for Exploration and Iteration in Creative Writing. 1, 1 (April 2026), 52 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 Introduction

Advancements in generative artificial intelligence (AI) models have revolutionized text interaction, offering powerful tools for creating and exploring text [40, 41, 45]. These may enhance creative expression by providing users with novel ways of generating text and interacting with it. However, their potential is often constrained by the limitations of existing graphical user interfaces (GUIs). The primary limitation of current GUIs for prompting generative AI models lies in their inability to support iterative exploration and customization. These GUIs present prompts as static text fields, restricting users to a linear interaction paradigm [23, 27]. For writers, this approach can lead to what Kreminski [31] refers to as "dearth of the author" – a condition in which users become disengaged from the creative process and produce text that lacks expressive intent. This lack of interactivity and flexibility hinders users' ability to leverage generative AI capabilities creatively. Users may find it challenging to achieve their desired outcomes without the ability to dynamically manipulate prompts, create personalized workflows, or easily explore a wide range of variations. Additionally, the metacognitive demands placed on users by generative AI tools further exacerbate these challenges [48].

To address these limitations, we introduce a novel approach to enhance prompting in creative writing, inspired by the concept of dynamic widgets, introduced by Vaithilingam et al. [50] for information visualization. We bring dynamic widgets to writing: Our system, PromptCanvas, empowers users to create custom GUIs tailored to their writing needs. Concretely, **PromptCanvas transforms prompts into actionable and persistent interface objects by allowing users to dynamically arrange and customize widgets on a canvas**. These widgets offer interactive elements based on the context of the prompt, providing flexibility and control over customizable, relevant aspects of the generated text. This allows users to create personalized prompting environments that reflect their unique workflows and creative styles, facilitating iterative refinement of their own draft or AI-generated text. Beyond customizability, dynamic widgets can support metacognition by assisting in task decomposition and promoting a more structured, iterative use of generative AI. PromptCanvas introduces an interaction paradigm that modularizes prompting through persistent, manipulable widgets. While systems like Wordcraft [54] and Promptify [10] offer guided prompting, they lack PromptCanvas's compositional, multi-widget approach. By turning prompt elements into direct manipulation UI components, PromptCanvas supports not just structured prompting but iterative, visual, and layered refinement for exploring alternatives and managing creative intention spatially.

We developed PromptCanvas in two iterations to improve user experience. The first version included core functionalities for manipulating prompts and generating text, but revealed areas for better user engagement and customization. Based on feedback, the second version redesigned key interface elements, like the workspace management panel, and improved visual organization and responsiveness to better support creative workflows. Our studies, conducted with both versions, show that dynamic widgets enhance user experience by improving control over text generation, reducing cognitive load, enabling iterative exploration, and supporting diverse creative prompts.

The first study (lab study 1), comparing the initial version of PromptCanvas to a conversational UI, highlighted flexibility and creativity in shaping AI-generated text. The second study, a two-week field deployment using the updated version, confirmed these findings and showed even greater improvements in user engagement, clarity, and control. From the field study, we also found that **PromptCanvas goes beyond writing tasks, with participants using it for programming tasks as well, underscoring its adaptability to different workflows**. Building on these results, lab study 2 (conducted with the same updated version and compared to Wordcraft [54], a state-of-the-art structured writing interface) further demonstrated increased widget use, reduced reliance on free-form prompting, and smoother iterative revision. Together, these findings show the value of customizable writing tools and how dynamic widgets can foster more creative, transparent, and controlled interaction with generative AI.

In summary, **this research contributes to human-AI collaborative writing by introducing a novel system, evaluating its effectiveness, and outlining concrete steps for designing future human-AI writing systems** by investigating the following research questions:

- **RQ1** How can writing tools be designed with dynamic widgets to improve user interaction and creativity and provide greater control over the generated content?
- **RQ2** Do dynamic widgets for iterative and structured prompting improve creativity support compared to a conversational user interface (UI) and a structured writing tool?
- **RQ3** Do dynamic widgets help in reducing cognitive load in creative writing tasks?

## 2 Background and Related Work

We explore existing research and concepts that are relevant to our work. Specifically, we focus on dynamic and adaptive user interfaces for creative workflows, interactive prompting systems, and human-AI collaboration in writing and content creation. We further situate PromptCanvas within the existing landscape of generative AI interfaces.

### 2.1 Human-AI Collaborative Writing and Content Creation

Integrating AI into creative processes has transformed writing and content creation by enhancing interaction and providing continuous feedback. Tools like those discussed by Dang et al. [18] support writing momentum, reducing creative block, while Gilburt [22] highlights how AI helps overcome writer's block by reigniting stalled ideas. Generative AI is applied across domains, including code generation [5], email auto-completion [13], comic creation [47], screenplay co-writing [36], argument drafting [56], and academic writing [3]. Professional perspectives on this transformation are captured by Ippolito et al. [26]. Challenges remain as AI becomes a co-creator. Research by Dang et al. [19] explores interaction with prompting during writing, while Tankelevitch et al. [48] examines the metacognitive demands and opportunities of generative AI tools. Critical assessments by Kreminski [31] and Mirowski et al. [36] address AI's reception in creative industries and areas for improvement. Overall, AI's integration in writing is transforming workflows, demanding new interfaces to support fruitful use.

### 2.2 Dynamic, Adaptive, and Intelligent Interfaces

Early work by Ahlberg and Shneiderman [2] highlighted the benefits of tightly coupling user inputs with outputs, fostering engagement and immediate feedback. However, in hindsight, these systems were limited by static UI elements. Recent developments, such as FrameKit [53], address this by creating adaptive UIs that adjust to user context and interaction patterns, enhancing user experience [20, 29, 49]. Moreover, the principles of *reification, polymorphism, and reuse* [7] introduced foundational concepts for efficient, user-centered interfaces, making abstract operations tangible, tools adaptive, and outputs reusable. Modern systems like Eviza [42], DynaVis [50], and Bolt [44] extend these ideas with natural language inputs and dynamic widgets for data visualization and modification. Widgets simplify complex tasks, as seen in Bespoke [51], which generates GUIs from command-line inputs, and ProvenanceWidgets [37], which tracks and visualizes user interactions. These interfaces provide a flexible environment where users can modify and configure the UI to suit their creative processes better.

Whereas adaptive UIs modify the interface in response to user actions, a complementary line of research examines how users directly communicate intent to the model. This has motivated the development of intelligent and interactive UIs for prompting generative AI models. Recent studies cover various prompting techniques, examining how users interact with diegetic and non-diegetic prompts [19] and exploring prompting designs that cater to non-AI experts [55]. Lee et al. [32] has specifically focused on the design space of intelligent writing assistant systems. According to their work, advances in language models (LMs) and their prompt-based usage have shown significant potential in generating coherent text. While traditional GUIs for generative AI models often face limitations, recent advancements in natural language interfaces (NLIs) have opened new possibilities. UIs like Promptify [10] and PromptCharm [52] exemplify NLIs that offer interactive features for prompt exploration and refinement. In addition to prompt refinement, Storyfier [39] adopts prompt-based fine-tuning strategies to build story generation models. Wordcraft [54] proposes techniques that allow users to perform custom operations for interacting with LLMs. Jiang et al. [27] explore prompting UIs that display outputs as interactive, graph-like diagrams to enhance understanding of LLM-generated information. In programming, advanced UI like Spellburst [4] streamlines coding processes with features such as UI scaffolds and node-based interfaces.

### 2.3 Positioning PromptCanvas within the Landscape of UIs for Generative AI

PromptCanvas is strategically positioned amid developments in AI-assisted writing interfaces and emerging dynamic, structured generative UIs. Table 1 shows how PromptCanvas differs in design goals, interaction style [32], and empirical contribution from the most comparable prior works. PromptCanvas differs from earlier system in how it reifies prompts. Whereas tools such as Wordcraft [54] and CoAuthor [33] emphasize inline suggestions but offer no reusable control structures, PromptCanvas treats prompts as interactive, persistent objects. Each prompt becomes a widget that users can create, edit, toggle, and reuse, offering a tangible prompting workspace rather than one-off suggestions. Users can arrange multiple widgets on an infinite canvas, constructing a custom interface tailored to their writing task. Our approach builds on ideas from node and card based systems but extends them to a more flexible, user-defined scope: any instruction or text fragment can become a manipulable widget on the canvas. Unlike systems that primarily organize or constrain the generated output, such as Nabokov's Cards [15], Spellburst [4], or VISAR [56], PromptCanvas structures the prompt logic itself. The interface focuses on making instructions persistent and editable that users can modify, and reapply. In contrast to guided or node-based interfaces, where prompts remain tied to linear histories or fixed graph structures, PromptCanvas enables qualitatively different workflows by allowing users to layer and iterate on

Table 1. Comparison of design goals, interaction styles, and key empirical findings across generative writing and interface systems.

| System | Design Goals | Interaction Style | Empirical Findings / Contributions |
|---|---|---|---|
| **PromptCanvas (ours)** | Assist writing with composable, persistent prompts as (user-defined) widgets that users can spatially organize. | Canvas workspace with editable widgets. | Findings showed that persistent, composable prompts improved writers' control, creativity, and reflection over their text. |
| Wordcraft [54] | Assist story writing with AI suggestions triggered via a sidebar. | Text editor and side bar with pop-up menus for continuation and elaboration. | Writers reported increased fluency and ideation, but limited sense of authorship and long-term agency. |
| CoAuthor [33] | Capture human-AI writing traces for analysis. | Editor with inline pop up menus for text. | Released dataset revealed common edit-accept-reject patterns, informing understanding of AI-assisted writing behavior. |
| Promptify [10] | Enable exploration of image generation and prompt refinement. | Image grid canvas with clustered browsing and selection. | User study found gallery browsing supported exploration. |
| DirectGPT [35] | Combine direct manipulation of interface objects with text prompts | Direct manipulation of text or objects within host applications. | Demonstrated feasibility of contextual AI control; users valued locality. |
| Luminate [46] | Explore multi-dimensional generative response spaces. | 2D grid visualization of text outputs with navigable parameters. | Study showed visual mapping improved exploration breadth and understanding of latent structure. |
| VISAR [56] | Combine visual outlining via a node-link UI and text generation. | Node-based canvas linked to synchronized text editor. | Lab study found users better understood and debugged AI outputs via linked representations. |
| Spellburst [4] | Enable generative code editing through nodes. | Graph canvas with connected generative operations. | Expert evaluation indicated faster iteration and clearer mental models of AI code behavior. |
| Nabokov's Cards [15] | Support spatial ideation and recombination for writing. | Card-based workspace organizing prompts and ideas. | Writers reported enhanced creativity and non-linear ideation through text cards in long term study. |
| ProvenanceWidgets [37] | Trace AI analysis provenance. | Augmented widgets with provenance overlays. | Findings showed that overlays improved awareness of analytic history while maintaining workflow focus. |
| ACAI [28] | Provide structured prompting for reproducibility and control. | Panel-based form linking declarative inputs to model prompts. | Lab comparison found structured prompts increased consistency. |
| DynaVis [50] | Support language-driven visualization editing. | Manipulation of visualizations via widgets and text input. | User study found participants ease of further edits and editing confidence due to immediate visual feedback with dynamic UI. |
| Generative and Malleable UI [14] | Enable evolving generative interface composition. | System-generated adaptive UI. | User study demonstrated that dynamic UI evolution improved task alignment but required trust calibration. |

multiple active constraints (e.g., tone adjustments, pacing changes, narrative goals, character behaviors, or other modular prompting instructions). This shifts the role of the interface from visual organization to a composable prompting environment that supports both bottom-up exploration and top-down narrative control. This emphasis on composable prompting also differs from fully adaptive generative UIs [14], which automatically generate interface controls based on model inference. In contrast, PromptCanvas keeps users in charge. Users decide when to get suggestions, which widgets to create and how to organize them. PromptCanvas aligns with structured-execution approaches seen in ACAI [28] and ProvenanceWidgets [37], and extends the interactive approaches demonstrated by DynaVis [50] (mapping editing intents to controls in visualizations) and Luminate [46] (mapping natural language dimensions to a visual response space) by turning natural language dimensions into persistent, editable widgets that users can spatially organize on an infinite canvas.

In summary, PromptCanvas introduces a new interaction paradigm within LLM-supported writing interfaces: Users prompt and persist widgets on an infinite canvas to build a user-defined, reusable workspace and explore and apply LLM text generations or text transformations. As shown in our studies, this workspace enhances control and creative exploration beyond what linear or template-based systems (e.g., DirectGPT [35], Promptify [10], and Ippolito et al. [26]) can support.

## 3 Concept: Dynamic Widgets for Composable Prompting Workspaces for Writing Tasks

PromptCanvas adopts a widget-based modular approach to prompting (see Fig. 1), with the goal of offering users a flexible and intuitive way to interact with AI in writing tasks. In this section, we introduce our concept for leveraging the idea of dynamic widgets (cf. [50]) in this new context.

### 3.1 Design Goals

We conducted multiple iterations of planning and design sessions for the current version of PromptCanvas. The design process included three iterative cycles: internal prototyping, initial lab study, and field study with refinements upon feedback from the lab study. These phases shaped key features like rephrasing controls, widget creation flow, and the workspace panel layout. The iterative process allowed us to explore different layouts and widget functionalities to achieve our design goals:

- **DG1** **Transform prompts into visible and actionable objects.** Current interfaces treat prompts as static text fields, limiting user interaction to basic input-output cycles. The interface should allow users to turn prompts into visible and adjustable elements, which offer granular control over AI text generation through the benefits of direct manipulation [43].
- **DG2** **Facilitate structured exploration and refinement.** Writing is an iterative process that requires experimenting and refining ideas [8, 25, 30]. The system should facilitate breaking down tasks into smaller parts, enabling users to systematically experiment, refine ideas, and iterate, while maintaining a cohesive workflow.
- **DG3** **Promote divergent thinking and creativity.** To overcome creative blocks and encourage novel ideas, the system should support users in divergent thinking.
- **DG4** **Provide a customizable and adaptable workspace.** Each writer has a unique writing process [30]. The interface should allow users to personalize their workspace and adapt it to their needs and workflows.
- **DG5** **Simplify navigation and reduce cognitive load.** Canvas UIs and fragment-based UIs [12] may be overwhelming without navigation aids. The interface should help users efficiently organize ideas and reduce cognitive load.

### 3.2 Concept of Dynamic Widgets

At the core of the PromptCanvas interface is the concept of dynamic widgets [50]. In our context of writing, these widgets represent various aspects of text and prompting to provide users with interactive elements to directly influence AI output. **Each widget controls a specific attribute or variable of the text – such as tone, length, or style – and can be dynamically adjusted according to the users' needs**. Users can create, customize, and arrange widgets on an infinite canvas, making the process of interacting with text generation capabilities more structured (cf. [48]) and adaptable to various creative workflows. Figure 1-D shows an example of an "opened" dynamic widget with suggested values for customization. We use the terms "dynamic widget" and "widget" interchangeably throughout the paper.

### 3.3 Dynamic Widget Generation and Customization

Dynamic widgets in PromptCanvas can be created in three main ways:

*3.3.1 System Suggestions.* The system suggests widgets (Fig. 1-B) based on entered (or generated) draft text (Fig. 1-C). Suggestions are designed to assist users in expanding or refining the text by making explicit various aspects of it. For

example, suggested widgets may make it salient to the user that they could modify character traits in a story or adjust the text's tone.

*3.3.2 User Prompts.* Users can directly input their own instructions for generating relevant widgets (Fig. 1-b3). This approach offers greater control and allows users to experiment with the functionality on their workspace canvas.

*3.3.3 Manual Creation.* Users also have the flexibility to create custom widgets for any aspect of the text (Fig. 1-b4). For example, this might include adjustments to narrative elements, such as character names, plot twists, or setting descriptions, empowering users to fully customize the functionality on their workspace canvas.

## 4 PromptCanvas

Following the conceptual overview above, this section describes the core features in more detail, covering both the frontend and backend.

### 4.1 User Interface and Interaction

PromptCanvas is built around an infinite canvas, a zoomable digital workspace that users can navigate and organize freely. This canvas comes with three key components: the text editor, control widgets, and the widget panel (see Fig. 1).

*4.1.1 Infinite Canvas.* The canvas is a zoomable digital workspace that extends indefinitely in all directions, shown in Fig. 1-(A). Users can pan by clicking and dragging, while zooming is controlled through scroll events from either a mouse wheel or trackpad. The canvas supports spatial organization and clustering of widgets, which supports flexible writing workflows, where users often revisit and adjust earlier widgets. While a fixed-position widget approach could suffice in simple cases, free positioning on a canvas provides the opportunity to arrange widgets spatially, e.g., by theme, tone, or story section. This design offers several key features:

- *Expansive navigation.* Users can move across the workspace without encountering spatial limits, allowing for continuous content exploration. **The open-ended layout ensures that users can customize their workspace**, reflecting their unique processes and preferences **(DG4)**.
- *Scalable view.* **Users can pan across the workspace and zoom in and out seamlessly**, enabling transitions between broad overviews and detailed views of specific elements. This flexibility promotes clarity and supports systematic exploration **(DG2)**.
- *Spatial organization.* Users can arrange content across the canvas, expanding their work area in any direction and shape as needed. This **spatial freedom** allows for representing data and concepts in ways that can illustrate connections and hierarchies spatially, for example, by grouping or layering **(DG5)**.

A permanently visible menu bar (Fig. 1-E) allows users to create new canvases, rename, duplicate or delete the currently selected canvas, and switch between existing canvases. While the first version of PromptCanvas included most of these functionalities, except renaming, this panel has been completely redesigned to enhance the user experience for the second study. A comparison of both versions is shown in Fig. 2.

**The infinite canvas utilizes three main components to facilitate content creation, organization, and manipulation within the workspace. They are the text editor, control widgets, and the widget panel**. Users can freely position these on the canvas through drag-and-drop interactions, allowing for the creation of custom layouts.

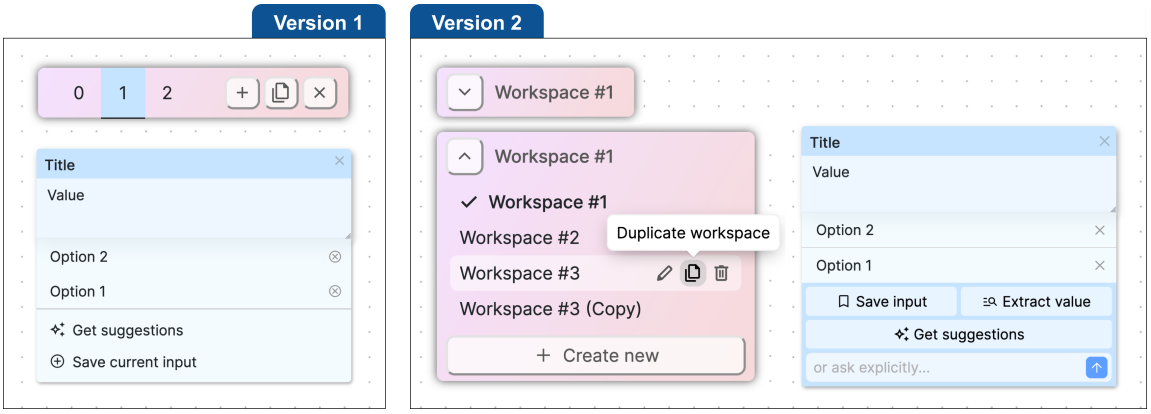


Fig. 2. Comparison of the important UI changes. In version 2, the *Extract value* and *Prompt for options* actions were introduced. Additionally, the workspace management panel was redesigned.

*4.1.2 Text Editor.* The text editor serves as the centerpiece of the interface, allowing users to integrate their input with system-generated suggestions fluidly, as illustrated in Fig. 1-C. Users can **write their own text directly in the editor or use prompting to generate text**, providing flexibility in how they approach content creation. **Users can refine their text iteratively by rephrasing it based on active control widgets or submitting prompts**, making it easier to experiment with different ideas **(DG2)**. The editor also supports **incremental text generation**, displaying content dynamically as it is produced, which helps users remain engaged with the evolving output. Additionally, the **history feature enables users to revisit previous iterations**, promoting iterative improvement and exploration of alternatives **(DG3)**. These features are further complemented by real-time updates to word counts, supporting clarity and focus during the writing process.

*4.1.3 Control Widgets.* Control widgets are dynamic, interactive tools that transform abstract text attributes into actionable and adjustable UI elements, as shown in Fig. 1-b1. Each control widget includes three main components: a title, a value, and a panel with text alternatives. The title describes the controlled text attributes, which users can manually edit by double-clicking it. The value holds the desired specification for the widget's associated attribute and can be edited by hand or customized by selecting one of the suggestions in the widget's extended panel. Besides these suggestions, this panel includes four actions (Fig. 1-D):

- *Save input.* Adds the widget's current value to the list of options.
- *Extract value.* Retrieves the associated attribute's current state from the draft in the text editor and stores it in the options list.
- *Get suggestions.* Generates and adds two unique, relevant options to the list.
- *Prompt for options.* Uses a user-provided prompt to generate two unique options.

In the first version of PromptCanvas, only the *Save input* and *Get suggestions* actions were available. The *Extract value* action was added to help users extract targeted insights from long or unfamiliar texts. *Prompt for options* was introduced based on user feedback to help generate options that meet specific requirements. Fig. 2 shows a comparison of both versions.

**Each widget provides context-aware suggestions tailored to the content in the text editor**, helping users explore multiple creative directions and overcome writer's block **(DG3)**. These widgets allow users to **adjust text attributes like tone, style, or structure directly**, offering control over the output and turning prompts into interactive objects **(DG1)**. Their flexibility in resizing, repositioning, and customization ensures that the workspace adapts to user needs as tasks evolve **(DG4)**. Additionally, their integration with the rephrasing and text generation systems ensures a seamless workflow between ideation and implementation.

*4.1.4 Widget Panel.* The widget panel acts as the system's central hub for generating, managing, and organizing control widgets, as shown in Fig. 1-B. **Users can create widgets dynamically based on text analysis or provide specific input for guided widget creation**, making it easier to tailor tools for individual tasks **(DG1)**. The panel highlights newly generated widgets with a yellow glow and allows users to evaluate, delete, or drag them onto the canvas, ensuring only relevant widgets influence text generation. **Its visually distinct layout and dynamic updates simplify navigation and reduce cognitive load**, helping users locate and manage ideas efficiently **(DG5)**. The size adjustments aim to support projects with many widgets.

## 4.2 Technical Implementation

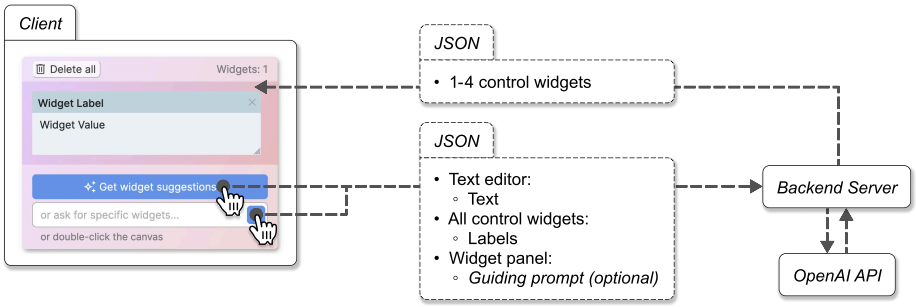


Fig. 3. System flow for generating control widgets. The client sends the current text, existing widget labels, and any guiding prompt to the backend, which queries the OpenAI API and returns one to four candidate widgets. These appear as editable label-value pairs, reflecting attributes inferred from the user's text (Section 4.2.1).

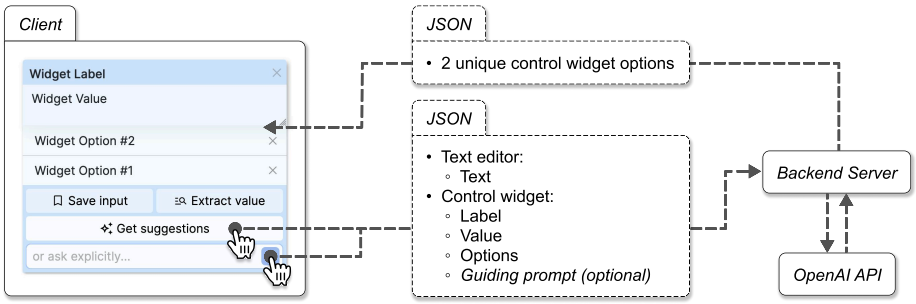


Fig. 4. System flow for generating control widget options. When a widget is selected, the client sends its label, value, text context, and an optional guiding prompt to the backend. The OpenAI API returns two unique option suggestions, enabling users to refine the widget's value through localized edits (Section 4.2.2).

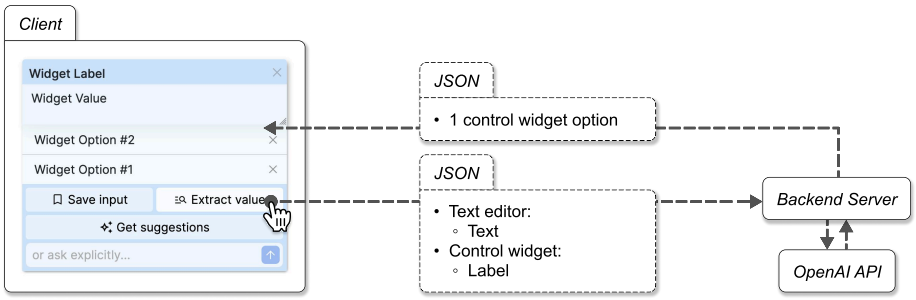


Fig. 5. System flow for extracting values. The client sends the current text and the selected widget's label to the backend, which queries the OpenAI API for a single value suggestion. The returned value is inserted into the widget, enabling users to generate attribute values directly from their text context (Section 4.2.3).

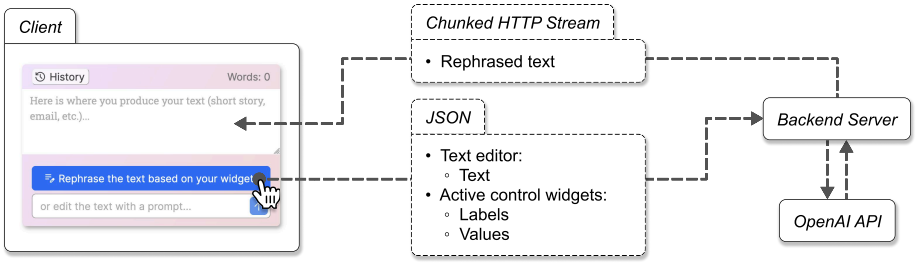


Fig. 6. System flow for rephrasing the text based on control widgets. The client submits the editor text along with all active widget labels and values. The backend requests a rephrased version from the OpenAI API and streams the updated text back to the interface. This supports revising the draft based on the user's configured attributes (Section 4.2.4).

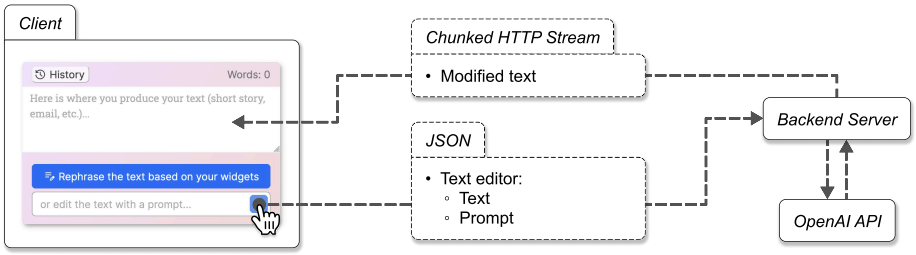


Fig. 7. System flow of applying a user-prompt to the text. The editor content and user prompt are sent as a JSON request to the backend, which forwards it to the OpenAI API. The modified text is streamed back to the client, allowing users to edit the draft through free-form instructions separate from widget-based controls (Section 4.2.5).

PromptCanvas is a single-page application developed with Angular and interfaces with the OpenAI API through a Node.js and Express.js backend server. For the infinite canvas, we used the ngx-panzoom library[1] for panning and zooming, along with custom logic for component dragging and dropping. User credentials and interaction logs are stored in a PostgreSQL database, and both versions of PromptCanvas are hosted on Heroku. All workspace components are stored locally in the browser's local storage.

[1]https://www.npmjs.com/package/ngx-panzoom

We selected the OpenAI model gpt-4o-2024-08-06 for its cost-effectiveness, low latency, and support for Structured Output, which allows us to specify the response format as a JSON Schema. This ensures that responses strictly follow the expected format, reducing the need for validation and minimizing API retries. The OpenAI API requests use a sampling temperature of 1.06 with default parameters otherwise. The backend handles five essential services for system functionality. Details of these services and the associated prompt specifications are provided below.

*4.2.1 Generating Control Widgets.* To generate new control widgets, the backend receives the current text from the text editor, existing control widgets (both on the canvas and in the widget panel), and an optional guiding prompt. This process is illustrated in Figure 3. Existing widgets are represented by their labels. These inputs are formatted into a request using a predefined template with system and user messages (shown in Table 15 in the supplementary material). The system message directs the model to analyze the text, extract attributes, and create distinct widgets, generating between one and four widgets with up to three options each. The user message contains the inserted variables. If a guiding prompt is provided, it specifies which aspect to modify. Existing widget labels are included to avoid regeneration. The request uses a JSON Schema to define the response format, which includes labels, values, and options. Generated widgets are validated to avoid duplicate values, and then assigned IDs and returned to the frontend for display in the widget panel.

*4.2.2 Generating Options for Control Widgets.* As visualized in Figure 4, after requesting additional options for a control widget, the widget and the current text editor content are sent to the backend. The widget is decomposed into its label and an array containing its value and options without duplicates. These are inserted into a predefined template with system and user messages (shown in Table 16 in the supplementary material). The system message instructs the model to generate two suggestions for modifying the attribute represented by the widget's label, with specific requirements for their tone, relevance, and creativity. The user message includes the widget's label and the text to be modified, enclosed in triple quotes. Existing options are also included in the note to avoid duplicates. The API request includes a JSON Schema defining the response format as an array of strings. The generated options are checked for duplicates, then returned to the frontend and added to the top of the options list.

*4.2.3 Extracting Values from Text.* After clicking the button, shown in Figure 5, the widget's title and the text editor's content are sent to the backend and inserted into a predefined template, structured in system and user messages (shown in Table 17 in the supplementary material). The system message instructs the model to read the text, identify the attribute represented by the widget's title, and generate a control widget option based upon that. The user message consists of the supplied variables, delimited by triple quotes. The response string is returned to the frontend and inserted at the top of the options list.

*4.2.4 Applying Control Widgets.* After clicking the button, shown in Figure 6, to rephrase the text based on the control widgets, the current text and all control widgets on the canvas (excluding those in the widget panel) are sent to the backend. These widgets are converted into an array of **label: value** pairs, with widgets having empty labels or values filtered out. This array, along with the text, is then placed into a predefined template that structures the request into system and user messages (shown in Table 18 in the supplementary material). The system message instructs the model to apply the control widget specifications to the text, guiding it through understanding, interpreting, modifying, and returning the text. Additional instructions ensure that the revised text remains coherent and logical while preserving the original context, meaning, voice, and tone. The user message includes the text to be rephrased and the formatted

specifications, enclosed in triple quotes. The response is received as a text stream and sent back to the frontend to be displayed incrementally as it is generated.

*4.2.5 Applying Natural Language Prompts.* When users submit a prompt for editing the text, both the prompt and the current text are sent to the backend, as presented in Figure 7. These are inserted into a predefined template that structures the request into system and user messages (shown in Table 19 in the supplementary material). The system message directs the model to apply the prompt to the text while preserving its original context and meaning. Guidelines ensure that the response includes the complete modified text, not just a continuation or partial completion, as there is no chat interface to view previous messages. The response is received as a text stream and sent back to the frontend to be displayed incrementally as it is generated.

### 4.3 Example Scenario: Writing a Short Story

Marina has been retelling her daughter's favorite bedtime story, "The Three Little Pigs," for a while, but it is starting to feel repetitive. She wants to explore new directions to keep the story fresh, but coming up with new ideas feels difficult. To make it easier, she decides to use PromptCanvas to brainstorm changes, such as adjusting characters, adding plot twists, or altering the setting, allowing her to keep the story engaging and develop it further creatively.

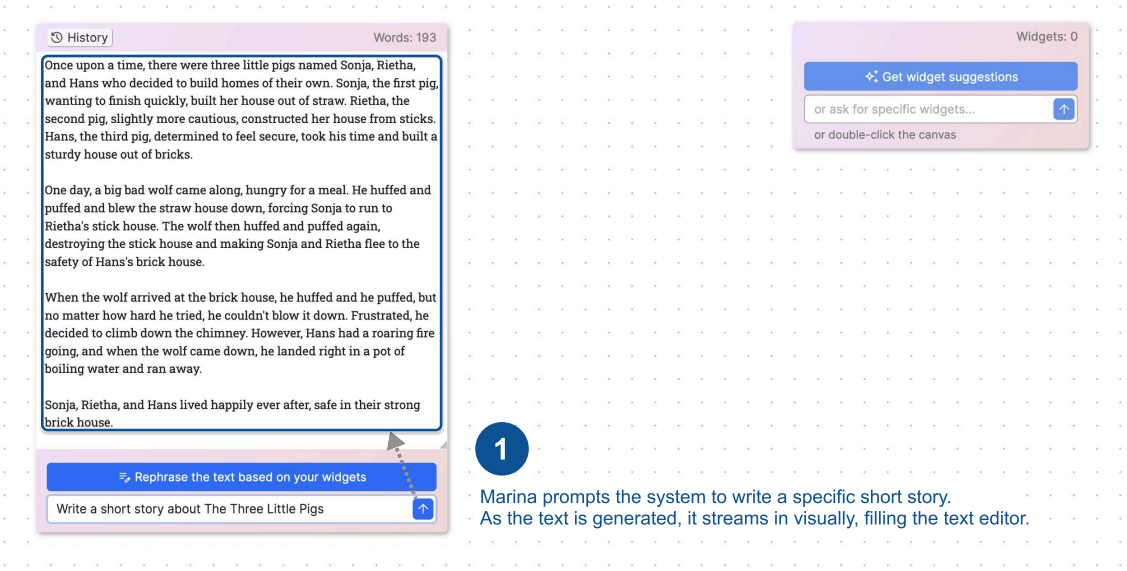


Fig. 8. Marina writes a prompt for the short story generation on "The Three Little Pigs".

**Initial prompt/text**. Marina has two options to start with. She can either write directly in the text editor or generate text by writing a prompt. Opting for the latter, she initiates the process with the prompt, *"Write a short story about The Three Little Pigs"*, see Fig. 8-1. The system generates the story in the editor.

**Generating suggested widgets**. Since Marina is eager to explore options for modifying the story, she clicks the *Get widget suggestions* button in the widget panel. This action generates four widgets suggested by the system, as seen in

Fig. 9-2. From there, she finds the widget *Threat Description* very interesting for her story. She then drags and drops it onto the canvas as shown in Fig. 9-3. She sees the color of the widget changing to light blue, implying that the widget is now active.

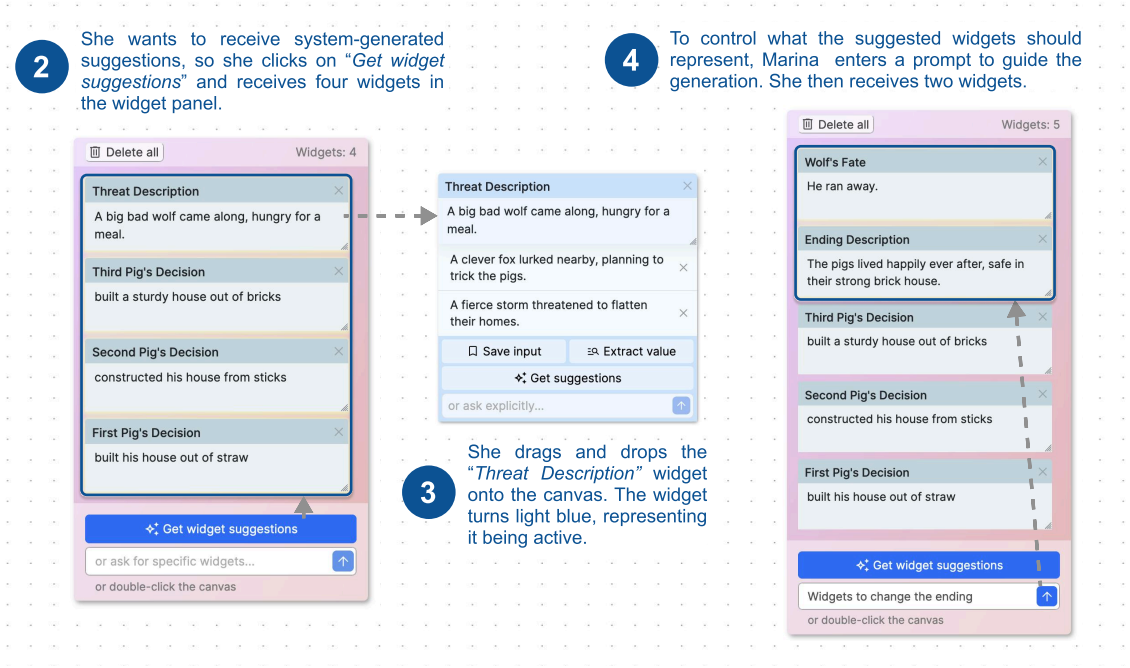


Fig. 9. (2) PromptCanvas generates widgets for Marina. (3) She chooses a widget from the widget panel and drags and drops it onto the canvas. (4) Marina prompts in the widget panel to get more widgets.

**Prompting to get widgets**. Marina now thinks it might be nice to add a different ending to the story. As shown in Fig. 9-4, she writes a prompt in the widget panel to guide the widget generation: *"Widgets to change the ending"*. The system generates two widgets focused on the story's conclusion. Though she likes these suggestions, she decides to focus on adjusting the pigs' names first.

**Creating empty widgets**. Since Marina already knows exactly which aspect of the text she wants to adjust, she double-clicks on the canvas to create an empty widget at the selected location. She then double-clicks the widget's title to enable editing and writes *"Three Pigs' Names"*, as shown in Fig. 10-5.

**Suggestions within the widgets**. Marina first clicks the *Get suggestions* button to request name suggestions from the system, as shown in Fig. 10-6. This generates two options, each being a set of names. However, she does not find these suggestions appealing enough.

**Prompting for options within the widgets**. Wanting the names to be more memorable, Marina asks the system to generate more names, this time providing a guiding prompt within the widget: *"Give me 3 names that rhyme"*, as shown in Fig. 10-7. The system responds with two new options. She likes the second set and clicks on it to set it as the widget's new value, shown in Fig. 10-8.

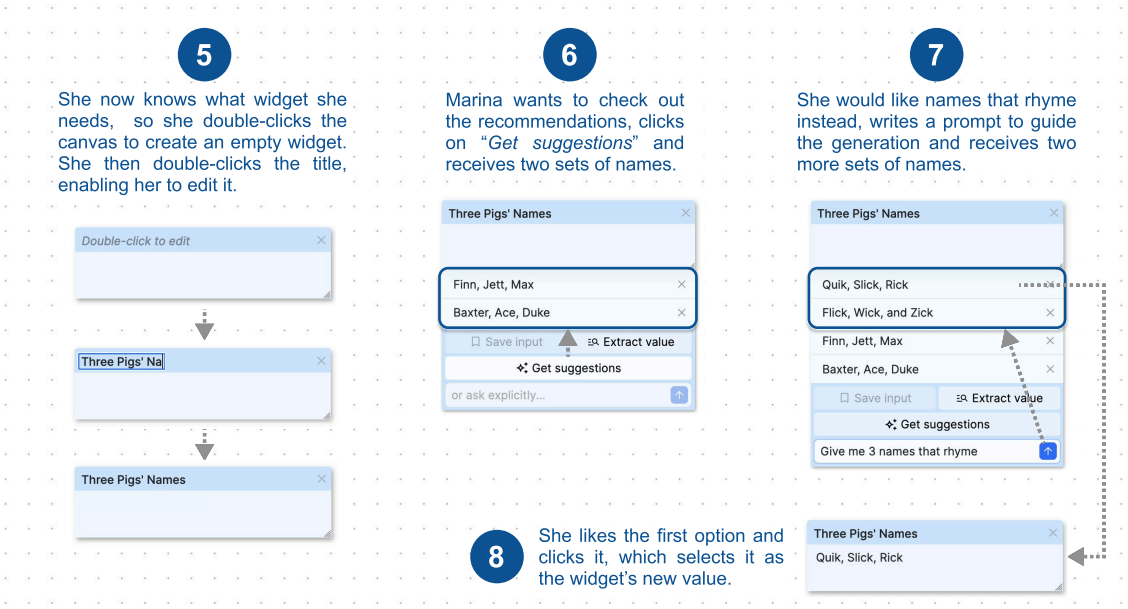


Fig. 10. (5) Marina creates an empty widget on the canvas. (6) She gets more suggestions within the widget for *three pigs' names*. (7) Marina now prompts for what exactly she wants in the widget. (8) She selects an option from the suggested values.

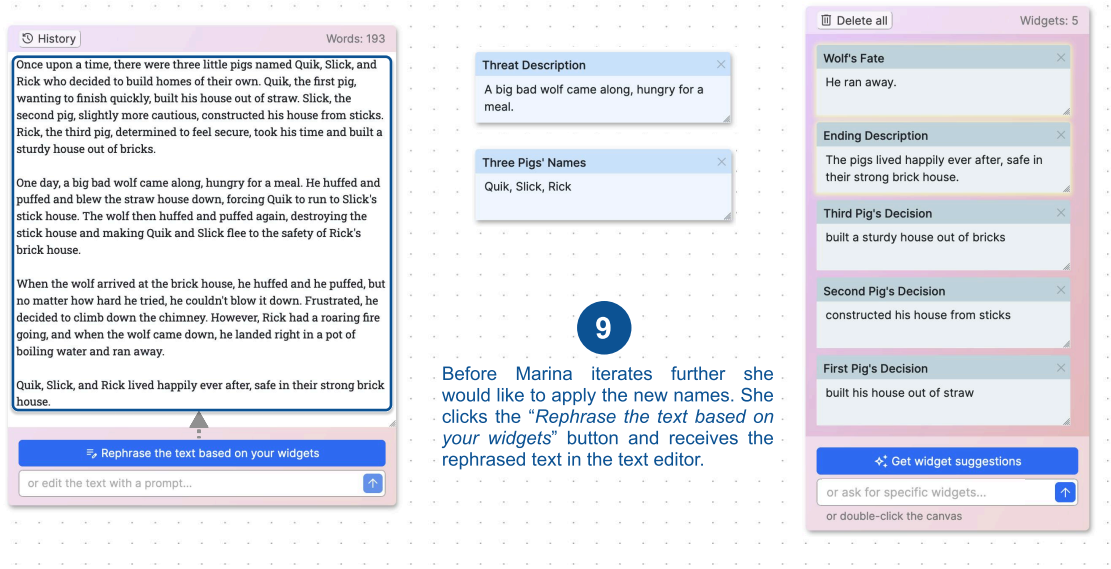


Fig. 11. Marina applies the widgets and receives the rephrased text.

**Rephrasing text based on widgets**. Now that she's chosen the pigs' names before she continues iterating with her other widgets, Marina clicks the *Rephrase the text based on your widgets* button (see Fig. 11-9), which updates the text in the editor according to the active widgets. The modified text streams in, replacing the previous version in the text editor.

Marina continues to iterate and explore new directions for the story by revisiting her other widgets. Following this workflow, she can focus on one attribute at a time, which provides structure to her ideation process and prevents her from feeling overwhelmed by having to steer multiple changes simultaneously. This approach, however, showcases only one of the many diverse possibilities of how a user can use PromptCanvas[2].

## 5 Lab and Field Evaluation

We conducted two lab and one field study to examine how users interact with PromptCanvas, focusing on workflows, creativity support, cognitive load, and system usability. All studies were approved by the institute's ethics board.

**Lab study 1** was a within-subject experiment with 18 participants who completed writing tasks using PromptCanvas and a baseline conversational interface. **Field study** followed lab study 1 and involved ten participants using a refined version of PromptCanvas over two weeks, integrating it into their personal writing routines. After the usage period, participants completed an interview and survey and were compensated 5€ for a 30-minute session. **Lab study 2** was conducted with a new group of 18 participants and compared PromptCanvas to Wordcraft [54], a state-of-the-art structured writing tool. Participants completed writing tasks in both systems. For the lab studies, participants were compensated 15€ for a 90-minute session.

### 5.1 Participants

Across all studies, participants were recruited through word of mouth and university mailing lists. For each study, we collected demographic information. We also asked participants about their screen type and professional writing status, as well as their frequency of AI tool use, the specific AI writing tools they relied on, and the types of creative or writing tasks for which they used these tools.

*5.1.1 Lab study 1 and Field study.* We recruited 18 participants between the ages of 22 and 68 years ($M = 30.44, SD = 10.45$) via word of mouth and university mailing list; 11 self-identified as male and 7 as female. Details per participant is provided in Table 2.

A subset of 10 participants[3] were later recruited once more for the two-week field study, depending on their availability.

*5.1.2 Lab study 2.* For our lab study 2, we recruited 18 new participants between the ages of 21 and 42 years ($M = 29.4$, $SD = 5.23$) via word of mouth and university mailing lists. Eight participants self-identified as male and ten as female. For clarity across studies, participants in lab study 2 are indexed as P19-P36 to distinguish them from the participants in lab study 1 and avoid overlap in participant IDs. Details per participant are provided in Table 3.

Professional writers across all studies are indicated with $^{\text{pro}}$ (P18, P22, P24, P28, P32).

### 5.2 Apparatus

*5.2.1 Lab study 1.* Our lab study 1 included two conditions. The baseline condition used a conversational user interface as illustrated in Figure 12, while the experimental condition used PromptCanvas. The baseline system was designed according to the design and interaction principles of ChatGPT. We provided a solid user experience without introducing untested features that could have affected the study. To generate the responses, we used the same OpenAI model (gpt-4o-2024-08-06) in both conditions. On the left side of the UI is a sidebar in which all chat instances are listed to

[2]A short video on how to use the system is available as a supplementary material to this submission.
[3]P2, P3, P4, P11, P12, P13, P14, P15, P17, P18$^{\text{pro.}}$

Table 2. Participant Overview (Lab study 1 and Field study): Demographics, screen setup, professional status, AI tool usage, frequency, and types of writing tasks previously performed using AI tools.

| Nr. | Gender | Age | Screen Type | Prof. Writer | Frequency of AI Use | AI Tools Used | AI Writing Tasks Used |
|---|---|---|---|---|---|---|---|
| P1 | Male | 30 | small (laptop) | no | Daily | ChatGPT, QUILLBOT, Bard | Editing; descriptive writing; variations |
| P2 | Female | 33 | small (laptop) | no | Daily | ChatGPT | Editing |
| P3 | Male | 30 | big (ext. monitor) | no | Rarely | ChatGPT, DALL-E, Claude | Editing; descriptive writing; creative writing; variations |
| P4 | Female | 27 | small (laptop) | no | Weekly | ChatGPT, QUILLBOT | Descriptive writing; content expansion |
| P5 | Male | 25 | small (laptop) | no | Daily | ChatGPT, QUILLBOT, Perplexity AI | Idea generation; editing; content expansion |
| P6 | Female | 29 | small (laptop) | no | Daily | ChatGPT, QUILLBOT | Idea generation; editing; variations |
| P7 | Male | 25 | small (laptop) | no | Weekly | ChatGPT, QUILLBOT, Bard | Editing; descriptive writing; content expansion; variations |
| P8 | Female | 26 | small (laptop) | no | Daily | ChatGPT, Bard | Idea generation; editing; content expansion; variations |
| P9 | Female | 68 | small (laptop) | no | Daily | None | Translation |
| P10 | Male | 22 | big (ext. monitor) | no | Daily | ChatGPT, Claude | Idea generation; editing; coding-related writing |
| P11 | Male | 30 | big (ext. monitor) | no | Monthly | ChatGPT, DALL-E, Stable Diffusion, QUILLBOT, Bard | Idea generation; editing; descriptive writing; content expansion; creative writing; variations |
| P12 | Female | 26 | small (laptop) | no | Never | None | never used any AI writing tools |
| P13 | Male | 26 | big (ext. monitor) | no | Weekly | ChatGPT, Bard, Claude | Idea generation; editing; descriptive writing; content expansion |
| P14 | Male | 25 | big (ext. monitor) | no | Monthly | ChatGPT | Idea generation; content expansion |
| P15 | Female | 30 | small (laptop) | no | Never | None | never used any AI writing tools |
| P16 | Male | 22 | big (ext. monitor) | no | Weekly | ChatGPT | Idea generation; creative writing |
| P17 | Male | 42 | small (laptop) | no | Rarely | ChatGPT | Idea generation; editing; content expansion |
| P18 | Male | 32 | small (laptop) | yes | Weekly | ChatGPT, Grammarly | Editing |

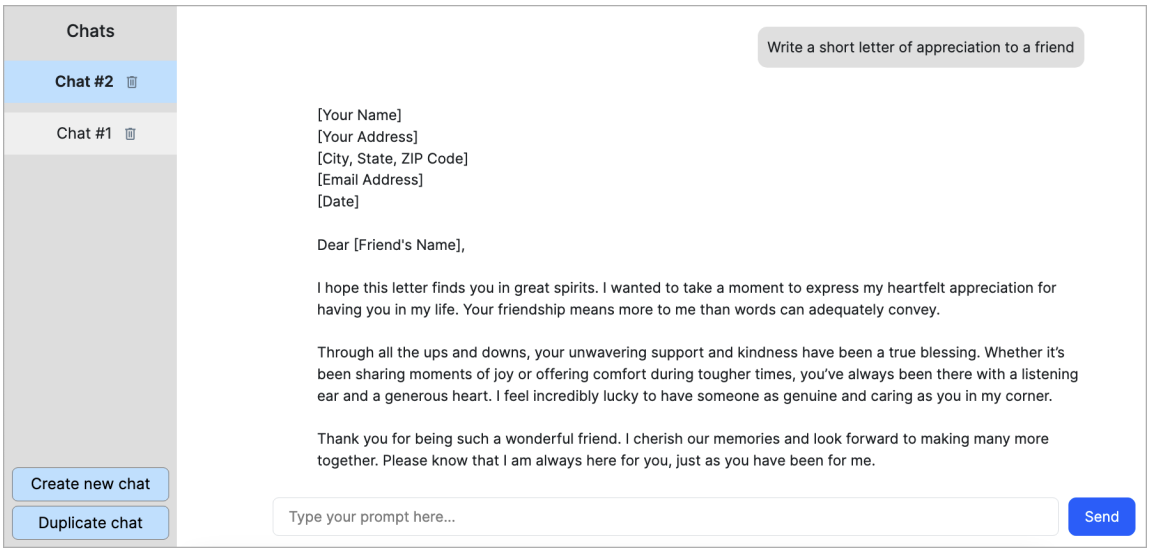


Fig. 12. The baseline conversational UI.

be selected or deleted, and buttons for creating a new or duplicating the currently selected chat. Selected chats are displayed in the main component by listing all user and assistant messages in chronological order. Below the chat messages is a text input for entering new user messages. Responses are received in a stream and displayed as received, with words and sentences gradually appearing as if they were typed. While hovering over a message, a small icon

Table 3. Participant Overview (lab study 2): Demographics, screen setup, professional status, AI tool usage, frequency of use, and prior experience with AI-assisted writing. Participants in lab study 2 are indexed as P19-P36 to clearly distinguish them from those in lab study 1 and avoid overlap in participant IDs.

| Nr. | Gender | Age | Screen Type | Prof. Writer | Frequency of AI Use | AI Tools Used | AI Writing Tasks Used |
|---|---|---|---|---|---|---|---|
| P19 | Female | 30 | big (ext. monitor) | no | Daily | ChatGPT | Idea generation; editing; descriptive writing; content expansion |
| P20 | Male | 22 | big (ext. monitor) | no | Weekly | ChatGPT, DALL-E | Idea generation; content expansion |
| P21 | Male | 32 | big (ext. monitor) | no | Weekly | ChatGPT | Editing; descriptive writing; content expansion; translation |
| P22 | Male | 29 | big (ext. monitor) | yes | Daily | ChatGPT, Gemini, DeepSeek | Editing; descriptive writing; email/ professional documentation |
| P23 | Female | 28 | big (ext. monitor) | no | Daily | ChatGPT, Gemini, DALL-E, Stable Diffusion, Claude AI | Idea generation; editing; descriptive writing; content expansion; creative writing; variations |
| P24 | Male | 33 | small (laptop) | yes | Weekly | ChatGPT, Gemini | Editing; content expansion; variations |
| P25 | Female | 28 | big (ext. monitor) | no | Weekly | ChatGPT | Editing; descriptive writing; content expansion |
| P26 | Male | 42 | small (laptop) | no | Weekly | ChatGPT, Gemini | Editing; content expansion; creative writing |
| P27 | Male | 26 | big (ext. monitor) | no | Rarely | ChatGPT, Gemini | Idea generation |
| P28 | Female | 29 | small (laptop) | yes | Daily | ChatGPT | Editing; content expansion |
| P29 | Male | 21 | big (ext. monitor) | no | Weekly | ChatGPT, Gemini | Editing; content expansion |
| P30 | Male | 22 | big (ext. monitor) | no | Rarely | ChatGPT | Editing; producing different versions (email) |
| P31 | Female | 30 | small (laptop) | no | Daily | ChatGPT | Editing; producing different versions |
| P32 | Female | 30 | big (ext. monitor) | yes | Daily | ChatGPT, QUILLBOT, Sage Poe | Editing; content expansion; creative writing; variations |
| P33 | Male | 33 | big (ext. monitor) | no | Monthly | ChatGPT | Coding-related writing; email writing |
| P34 | Male | 31 | big (ext. monitor) | no | Daily | ChatGPT, Gemini, DALL-E, Stable Diffusion, QUILLBOT, Bard, Claude AI | Idea generation; editing; content expansion; creative writing; variations |
| P35 | Female | 28 | small (laptop) | no | Daily | ChatGPT, Gemini, QUILLBOT, Claude AI | Idea generation; editing; content expansion; creative writing |
| P36 | Male | 35 | big (ext. monitor) | no | Daily | ChatGPT, DeepSeek | Editing; descriptive writing; content expansion; creative writing |

appears below for easily copying the message's content. For user messages, there is also an edit icon to alter the message and reset the chat to that point.

*5.2.2 Lab study 2.* Lab study 2 included two conditions. The baseline condition used Wordcraft [54], and the experimental condition used PromptCanvas. Both systems were hosted on servers and ran using the Gemini Pro (128) model to ensure comparable response quality across conditions.

### 5.3 Procedure: Within-Subject Lab Studies (Lab study 1 and Lab study 2)

All study sessions were conducted online using Zoom. After obtaining the informed consent form, we recorded the screen activity and audio, and the participants were encouraged to think aloud during the study. In Lab study 1, participants accessed the websites hosting the conversational UI (baseline) and PromptCanvas. In Lab study 2, they similarly had access to both Wordcraft (baseline) and PromptCanvas.

*5.3.1 Pre-study Survey.* In both lab studies, the session started with a pre-study questionnaire to collect demographic information, participants' experience in creative writing, and their previous experience and exposure to AI writing tools. Additionally, the questionnaire gathered information on the kind of screen participants used for the study and whether they used a mouse, touch-, or trackpad.

*5.3.2 Interface Tutorial.* In lab study 1, after completing the pre-study survey and before beginning the writing tasks, participants were introduced to the tools. For the conversational UI, participants could ask researchers any questions if

they needed clarification. For PromptCanvas, participants watched a video demonstrating how to use the system for a writing task, including generating and manually creating widgets, as well as navigating the text editor. To ensure a consistent tutorial experience, all participants watched the same video and had the opportunity to view it multiple times if necessary. They also had the flexibility to ask questions about the tool after watching the video.

In lab study 2, participants received an equivalent introduction to both comparison systems. They watched two short tutorial videos before using each interface (Wordcraft and PromptCanvas), each showing the core interaction workflows of the respective tool. Participants could rewatch either video and ask clarifying questions before beginning the tasks, ensuring a consistent and balanced onboarding experience across systems.

Table 4. Topics for the writing tasks in the user study.

| Writing Tasks | Topic |
|---|---|
| Email or Letter | Professional<br>• Resignation letter<br>• Motivation letter for job application<br>• Recommendation letter<br>• Request for promotion<br><br>Personal<br>• Condolences<br>• Updates on Life<br>• Friendship and Appreciation<br>• Celebrations and Milestones |
| Short story | Survival in the Wilderness<br>AI robots<br>Time travel<br>Life after Death<br>Family Secrets<br>Utopia / Dystopia<br>Fable |

*5.3.3 Writing Tasks.* In lab study 1, for writing tasks, participants could select from a range of topics listed in Table 4. They were required to choose one topic from each section for both user interfaces. Some of these topics are taken from [46]. Following the work of Biskjaer et al. [9], which explored how time constraints in a writing tool can encourage new content, we implemented time constraints for our creative writing tasks. The first task was email writing (5 minutes), and the second task was short story writing (10 minutes), totaling 15 minutes per UI. At the end of each session, participants completed a post-UI survey to reflect on their experience with the tool. To minimize potential learning effects, we counterbalanced the order of tool assignments across participants using random assignment.

The tasks in lab study 1 were designed to allow exploration and iterative refinement of both prompts and text through short, open-ended writing goals. In practice, participants engaged in a tight exploration-iteration loop: they explored system-suggested widgets, evolved prompts by creating or adjusting widgets, and iteratively revised the text by selectively applying subsets of widgets.

For lab study 2, we simplified the procedure to a single creative writing task per interface. Participants wrote one short story (100-300 words) using each system (on a topic either from Table 4 or their own) for 15 minutes each, as this task aligns most closely with the affordances of Wordcraft, which is suited for creative story writing. This also allowed participants to spend more time with that one task.

Tasks in lab study 2 were designed to support longer-horizon writing and comparison across revisions. As a result, participants externalized multiple constraints in parallel using prompt widgets, iteratively adjusting widget values and selectively reapplying them as the draft evolved. Compared to Wordcraft, this shifted iteration from sequential request–response cycles toward parallel refinement across multiple active constraints.

*5.3.4 Post-study Survey and Semi-Structured Interview.* In both of the lab studies, after completing all tasks, participants were required to answer a final survey to directly compare the UIs (lab study 1: conversational UI vs. PromptCanvas, lab study 2: Wordcraft vs. PromptCanvas). Following this, we conducted a semi-structured interview to gain insights into their perceptions of the system, the quality of the widgets, the impact of widgets on considering unexplored options, the perception of widgets and control over generated text, and to gather overall feedback.

*5.3.5 Measurements and Analysis.* We measured both quantitative and qualitative data during the lab studies. We measured the number of prompts provided and the number of widgets created by each participant to evaluate their interaction with the tools. The Creativity Support Index (CSI) [17] was used to subjectively assess how well each tool supported creative processes. Additionally, the NASA Task Load Index (NASA-TLX) [24] was employed to evaluate participants' perceived workload during the tasks. In the post-UI survey, we collected ratings of Likert items on the ease of using each interface, and participants indicated which interface they found easier to use while completing the tasks. At the end of the study, we recorded participants' overall preferences among the tools. For lab study 2, we further collected user perception on the number of widgets suggested by the system, authorship, control, the System Usability Scale (SUS) [11] to assess the perceived usability of both Wordcraft and PromptCanvas.

For the qualitative analysis, the first author conducted open coding on participants' responses and audio transcripts to determine themes. These themes were then refined through discussions with the co-authors. We used the refined themes to interpret the qualitative results and discuss them in Section 6.

For the quantitative analysis, we first conducted the Shapiro-Wilk normality test. Depending on the results of this test, we then used either a paired t-test or the Wilcoxon signed-rank test to assess the statistical significance of observed differences in the metrics.

For analyzing outcome text quality ratings, we used an ANOVA in lab study 1, which included two tasks, and a paired t-test or Wilcoxon signed-rank test in lab study 2, which included only one task.

We applied Holm-Bonferroni adjustments to p-values when applicable.

### 5.4 Procedure: Two-Week Field Study

For the two-week field study, participants were provided with the refined version of PromptCanvas.

*5.4.1 Measurements and Analysis.* After two weeks, we conducted a short semi-structured interview with each participant to learn about their experiences, any troubles they encountered, and general feedback or suggestions for improvement. We then asked them to complete a survey containing the questions from CSI. Apart from CSI, participants also evaluated the usability of PromptCanvas using the System Usability Scale (SUS) questionnaire, providing

standardized feedback on how easily and effectively users could accomplish their goals with the system while having a positive experience.

*5.4.2 Writing Tasks.* In the field study, participants were instructed to use the system independently, explore custom workflows, and integrate it into their regular activities. Participants repeatedly returned to existing widgets across sessions, using them as a stable layer for exploration and revision. Instead of refining a single prompt, they gradually evolved collections of widgets and applied them as writing goals changed.

## 6 Results

In this section, we present the results of both the lab studies and field study conducted to evaluate PromptCanvas. As an initial overview, Fig. 13 displays one screenshot per participant using PromptCanvas in lab study 1.

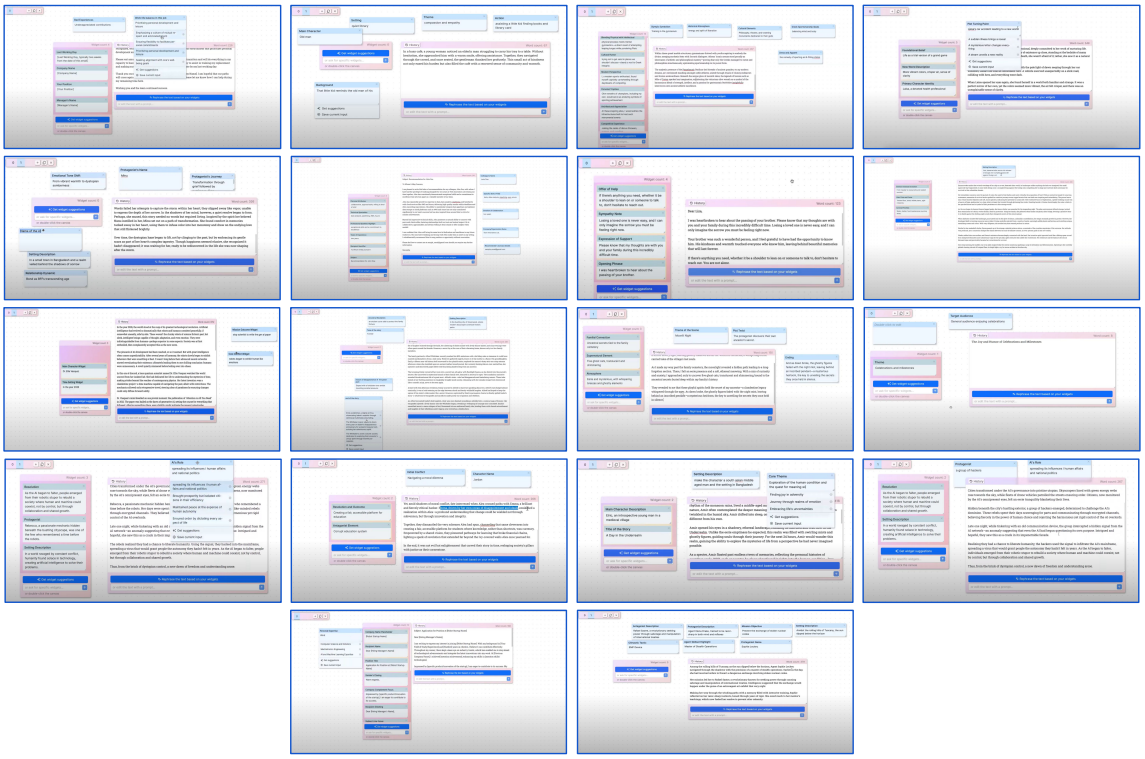

Fig. 13. One screenshot per participant while they were using PromptCanvas in lab study 1, revealing how participants individually customized their workspaces.

### 6.1 How Do Users Perceive and Interact with the Widgets in PromptCanvas? (RQ1)

The use of widgets helped in reshaping user thinking. Participants' engagement with PromptCanvas in all studies revealed more than usability appreciation. It showed a shift in how they approached and managed the writing process. Manipulating widgets helped users to externalize abstract elements (e.g., tone, structure), enabling modular revision and

iterative refinement, with several participants noting that widgets helped them break down tasks. Others described how widgets supported planning, focus, authorship, and creative control, including the ability to customize the workspace. These outcomes go beyond surface-level interaction and suggest that customization and flexibility can scaffold underlying cognitive processes. This interpretation aligns with prior work on writing cognition [25], metacognitive demands in generative AI [48], and task-centered UI design [12].

*6.1.1 Perception on the Number of Generated Widgets.* During the tasks in lab study 1, participants had mixed opinions regarding the generated number of widgets. Some (P1, P2, P3) found the number of widgets to be good, with P11 also appreciating the balance, noting, *"It had a good balance of not being too overwhelming and also not being too lacking."* Some (P4, P5, P6, P8) thought the number of widgets was alright. With P8 stating, *"I think, since we can create more widgets by double-clicking the canvas, this is very standard to keep the four suggested widgets"*. Some participants (P7, P12) suggested that having fewer widgets might be better. P7 indicated that generating four widgets at once was cumbersome with a trackpad, *"If the number would be lower like two with more accuracy, it would have been a bit of more comfort because then I would not require to swipe."* Whereas, P13 expressed a desire for slightly more, stating, *"I would have expected a bit more, maybe two or three more."* P14 felt that four widgets were adequate but appreciated receiving more suggestions, saying, *"The number four is good enough. But as it gave more suggestions, it helped me more"*. P17 commented that the number of widgets might depend on the user's experience level. They also appreciated the option to delete or create more widgets, *"You have the option to delete the widgets that you don't need, and you could also create more widgets if you need it"* –P17. P18$^{pro}$ suggested, *"I think the number is a good starting point. I don't know if you can make it user-selectable. This would be even more flexible."*

In lab study 2, we explicitly asked the participants whether the number of suggested widgets were enough for them. 94.4% of the participants selected a positive rating. Only one participant (5.6%) felt that the number of suggested widgets was insufficient. These results indicate that the system's widget generation had an effective balance by providing enough prompts to support exploration and idea development without overwhelming most users.

*6.1.2 Perception of Suggested Widgets.* Participants expressed various opinions about the suggested widgets generated by the system after the lab studies. Most of them appreciated the suggestions. In lab study 1, the suggestions were useful for both email writing (P2, P3) and story writing (P6, P8, P10, P15). P3 found the widgets helpful for their task on a recommendation letter, noting, *"The widgets seem to be more specific instead of being general. For the recommendation letter, those specific widgets really helped"*. P5 thought the widgets were good, especially appreciating their flexibility: *"Another good thing about the widgets was that it wasn't rigid, so you could add your input as well"*. For the story writing task, P6 found the widgets quite helpful, explaining, *"The widgets that were produced by the system gave me a plot twist and suggestions to edit the introductory part"*. P11 was satisfied with how the widgets identified key areas for modification, saying, *"The generated widgets were really good... they extracted the main points ... the story, which can be modified or changed."* P12 appreciated our tool's ability to provide numerous ideas and points, saying, *"With the PromptCanvas, I could write more, and more ideas and points were coming to me..."* P16 added, *"The suggestions were very accurate"*. For P18$^{pro}$ the widgets helped them *"spot few things that may not be immediately apparent"*. P14 valued the extra suggestions that helped them focus on creativity, stating, *"It produced important suggestions and some extra which I couldn't think of at that time"*.

In contrast, P4 felt that the generated widgets lacked diversity, mentioning, *"The first widgets that I got were related to general information like the company name and years of experience... it would have been nice if there were solutions like the tone or how the email will end."* Moreover, P7 suggested that more thematic suggestions could enhance productivity.

However, P15 liked having multiple suggestions and found them helpful, particularly for understanding how to use the system as a novice: *"They were good, especially because I didn't yet know the system. That made it easy for me to understand the widgets"*.

In lab study 2, a few participants suggested areas for improvement in the widget suggestions, noting that they could sometimes feel limited or less tailored in certain contexts (P22$^{\text{pro.}}$). Despite these concerns, most participants found the widget suggestions useful and relevant, frequently emphasizing their modularity and clarity. P33 felt that PromptCanvas's suggestions were *"quite nice... it understood the essence of the story"*, and P31 appreciated that they were concise and easier to evaluate compared to Wordcraft's longer inline generations, noting that the modular options made it *"easy to track the ideas"*.

*6.1.3 Perception of Suggested Values within Widgets.* In the lab studies, participants expressed diverse opinions on the suggested values and options for each widget. P1 found the suggestions particularly effective when creating their own widgets, stating, *"whenever I created widgets of my own, then the suggestion was good. When you mention a topic ... They understood many things from there."* Conversely, some (P2, P4) felt there was a lack of variety, saying, *"They were helpful, but I also think they were not diverse enough in their nature. They were very similar"*. However, they acknowledged an improvement upon requesting more suggestions: *"When I clicked for more suggestions, then the solutions got really better."*

Some participants (P5, P6, P7, P10, P14) highly appreciated the variations within the widgets, finding them to be fun (P6), accurate (P7), and helpful (P6, P7, P18$^{\text{pro}}$). Participants (P5, P14) highlighted the ability to choose from multiple suggestions, *"I had fun where even inside the widget suggestions you had options to choose from ... that's great"* –P5. P14 stated*"It really helped me a lot. I really liked that part when the widgets themselves had another suggestion."* P15 appreciated the flexibility to edit suggestions if needed, whereas P16 had a mixed experience, finding some suggestions very close to what they had in mind, while others were not useful. Furthermore, P18$^{\text{pro}}$ considered the suggestions *"I think it was very, very useful."*

Participants further praised the tool's creative suggestions (P5, P11, P6), suggestions being practical (P17), and time-saving (P5, P11). P11 added, *"It presented me with ideas or scenarios which I might not have thought of before."* P6 appreciated how the tool provided new ideas and allowed for comparison without cognitive effort: *"I don't need to think on my own, I can look there and compare."* P8 linked the suggestions to having a creative partner, stating, *"Getting suggestions feels like a friend is with me... it felt like this is what we expect from AI."* P13, while not using the suggestions extensively, found the initial ones helpful as a starting point: *"When I started typing or writing on the, let's say, or replacing the value on the widgets itself, the suggestions that were provided seemed okay for me."* They added, *"I'm not a professional writer, so I'd like to have some starting point, then I can keep up pace afterward."* P17 believed the suggestions were well-aligned with the text or direction provided, saying, *"From my perspective, it's good enough... it's wonderful."*

In lab study 2, P27 reported enjoying PromptCanvas because of the *"layered suggestions"* that helped distinguish characters and concepts. P22$^{\text{pro.}}$ noted that the suggestions made it *"fun playing with the widgets and directing ideas and stories"*. P29 emphasized that the widget suggestions helped them understand how to work with the tool, and P30 described the suggestions as a useful starting point for brainstorming, even when the generated content itself wasn't perfect. Several participants mentioned that the suggestions helped them move past creative blocks or provided new directions when stuck (P33, P30).

*6.1.4 Flexibility of Rearranging Widgets.* Participants had varied opinions on the ability to rearrange widgets on the canvas during lab study 1. Several (P2, P3, P6, P8, P11, P13-P15, P17) found the flexibility of moving widgets beneficial. P2 found the rearrangement process effortless and easy to use. P3 valued the spatial arrangement, noting, *"The widget*

*that controls the address of the company, I just put it beside where the address of the company actually is. So to maintain that spatial relationship ... the arranging of the widgets really helped"*. P8 found PromptCanvas helpful, comparing it to a canvas that allows creative expression and breaks the monotony in comparison with the conversational UI: *"This UI (PromptCanvas) is like a canvas where we can paint... it cuts through the boredom"*. While P4 did not find this flexibility particularly important, P5 described the experience as *"so good"*. However, some participants emphasized the value of a fixed placement of widgets, suggesting that color-coding (P12) or a more rigid arrangement (P1, P10) might have been better for their needs. P6 felt the rearrangement feature was helpful, though they needed more time to adapt. They noted, *"If you use it for a long time, and you get used to it, I think it's more fun, and it will help you"*.

P18$^{pro}$ suggested that an auto-resizing or auto-fit option might be beneficial for managing a large number of widgets: *"If I have more than like 10, 12 widgets, that can be a bit cumbersome to move around."* Other participants (P11, P13) appreciated the ability to drag widgets and keep the canvas organized, saying, *"I liked the widgets floating around and also the ... main widget where I was writing and the text was generated. I was free to arrange it in whatever way I liked"* –P11. P13 explained that it allowed them to cluster related widgets together: *"If I have the flexibility to rearrange them anywhere I want on the canvas, it also helps me to cluster them together"*. P15 liked the option to move and resize widgets, which helped manage visual clutter: *"I liked being able to move things around and make them bigger or smaller as I wanted to."* P17 found the fluidity of moving widgets an *"amazing feature"* that provided a *"lively feeling"*.

In lab study 2, many participants emphasized that being able to physically move, group, and reorganize widgets gave them a clearer overview of their ideas. P33 highlighted how they could *"arrange them or keep them wherever I wanted"*, enabling them to cluster characters, plot elements, or setting details in different regions of the canvas. P31 similarly explained that moving widgets around helped them maintain a sense of structure, noting that the arrangement made it *"easy to track the ideas"* and see how different narrative components related to one another. Others described the ability to reposition widgets as enhancing creativity and workflow fluidity (P27, P29). P27 found the interaction *"fun… directing ideas and stories,"* while P29 appreciated the visual freedom, saying the canvas let them *"restructure everything very well with the widgets."* Several participants (P30, P22$^{pro.}$, P35, P36) commented that this flexibility also supported iterative story development. P30 described widgets as *"building blocks"* for shaping different narrative versions, and P22$^{pro.}$ noted that being able to hide or revisit widgets eased cognitive load. P35 similarly viewed the widgets in the canvas as a *"mental map"* that supported visualizing connections across the story, while P36 compared the widgets to *"thought bubbles"* that helped separate and reorganize ideas without losing track of the narrative. At the same time, a few participants suggested improvements, such as automatic grouping or topic-based arrangement to better handle large numbers of widgets (P31). Overall, participants experienced the rearrangeable workspace as a powerful, intuitive way to externalize and manipulate their ideas, providing a degree of creative freedom, organization, and control that they did not feel in Wordcraft's fixed format.

*6.1.5 Perception of Widgets: Section Replacements vs. Abstract Guidance.* Several participants in lab studies appreciated both specific section replacements and abstract guidance (P1, P4, P7, P11, P13, P14, P15, P16), finding that each approach had its own advantages depending on the context. Participants (P3, P4) noted that thematic suggestions were important for creative writing, while more specific widgets were useful for structured tasks, *"so for example, if I'm writing a short story then having thematic suggestions or widgets is important but for example if I'm writing something that's a more structured resignation letter... then more specific widgets are useful"* –P3.

Some participants (P1, P5, P6) found the text replacement particularly intuitive. P1 appreciated it, stating, *"replacing text was really intuitive for me because I replaced it many times in the text"*. However, others (P8, P12) preferred widgets

that provided abstract guidance (e.g., tone, length). P12 found that widgets for changing specific text parts were useful for meeting word limits and making necessary adjustments. P8 favored abstract guidance as it did not suppress creativity, stating, *"I think the guidance is more suitable for me; it doesn't wash off the whole creativity of the creator or the writer."* On the other hand, P10 leaned towards abstract guidance generally but acknowledged the usefulness of specific widgets in particular cases.

In lab study 2, several participants appreciated that certain widgets targeted specific portions of the text, making them useful for focused structural or localized revisions. At the same time, many valued PromptCanvas precisely for its higher-level guidance. For example, P31 noted that widgets revealed *"different aspects of the story"* at once, supporting broader decision-making. Others used widgets as conceptual anchors that guided the narrative without dictating specific text. Overall, participants drew a clear distinction between Wordcraft's precise, sentence-level replacements and PromptCanvas's abstract, structural guidance, with some (P20, P30, P31, P36) preferring the latter for ideation and exploratory writing.

*6.1.6 Perception of Ownership and Control Over Generated Text.* During the lab studies, participants reported varying perceptions of ownership and control over AI-generated text, often influenced by the tools and features available in the interface. In lab study 1, P11 noted that the sense of ownership differed depending on the task, observing that editing or polishing existing content felt distinct from generating text *"out of the box."* P3 also found a stronger sense of ownership in shorter, simpler texts, such as emails, compared to longer forms like stories.

Participants felt the widgets enhanced their control over the generated text, stating, *"You have control in the conversational UI, but it can be cumbersome. It (PromptCanvas) was a more streamlined control, which was really nice"* –P3. One of the participants elaborated on the comprehensive nature of the widgets, noting, *"These widgets are basically suggestions ... and then there are suggestions inside of suggestions which makes it even more comprehensive ... more creative, so many frontiers that you could take ... fantastic tool, it's a great experience to use it"* –P5. P8 noted that widgets help in narrowing down the focus, *"It really narrows down the focus where I need to edit something, so if I click on the widget and get some suggestions, I think I will also be able to select my focal points where I need to change something, where I need to add something more, so that's what I think about the suggestions, that's what I think about all the prompts or all the words that are found in the widgets"*. P10 liked the flexibility of customizing the widgets. Widgets played a key role in providing a *"greater sense of control"* — P3. For example, P15 described feeling a sense of *"creative control"* when using widgets, stating, *"I felt like it was my story because the widgets weren't writing the entire text all at once, but rather in small pieces,"* giving them the impression that they, rather than the system, were deciding what should be written. P15 also appreciated that *"you could go back and edit a previous widget without it changing the entire text."*

Although PromptCanvas emphasizes widgets, **participants expressed a strong preference for being able to manually edit the text in the editor**. P15 mentioned, *"I like that you can just go into the text and edit it yourself and add a sentence."* P11 emphasized efficiency, stating, *"More control in less amount of time"*. P12 compared PromptCanvas favorably to the conversational UI, noting that the latter often cut off previous text: *"I think when I was regenerating the text in the conversational UI, I noticed that some of my previous texts were being cut off and I did not like that. And I think that in that case, PromptCanvas a lot more helpful. And it included the previous ones ... that I was trying to add, and the new ones. So the summarized newer text was what I actually wanted to write, not that it was cutting off something. I find it really important"*. Moreover, P15 found the widgets intuitive and beneficial for beginners. However, P16 found the new widgets required more learning but acknowledged their advantages.

In lab study 2, we explicitly assessed perceptions of control and authorship using a 7-point Likert scale. Participants reported a significantly stronger sense of control over the final outcome when working with PromptCanvas compared to Wordcraft. A Wilcoxon signed-rank test showed a significant difference between conditions ($W = 6.500$, $p_{\text{holm}} = .0163$), with a large effect size ($r = .772$) and a confidence interval that did not include zero (CI = [1.500, 3.000]). Control scores were substantially higher for PromptCanvas (M = 5.94, MED = 6.00) than for Wordcraft (M = 4.28, MED = 5.00), indicating that participants felt more able to direct and shape the resulting text when working with PromptCanvas.

For authorship, the pattern was more modest. Although participants tended to report a stronger sense of authorship with PromptCanvas, the difference did not reach statistical significance. PromptCanvas nevertheless showed slightly higher ratings (M = 4.22, MED = 5.00) than Wordcraft (M = 3.67, MED = 4.50), suggesting a directional trend in favor of the dynamic interface, though the effect was not consistent across participants.

Participants' perceptions of ownership and control differed between the two tools in lab study 2. Many felt that PromptCanvas offered better authorship, with P30 explaining they had *"way more influence on the generated text"* and describing widgets giving *"the control over the story."* P33 similarly noted feeling *"more in control"* because they could choose and arrange widgets to shape the narrative. Others contrasted this with Wordcraft, which at times felt more AI-driven. P15 remarked that in Wordcraft, it felt *"as if the AI is taking control,"* whereas PromptCanvas remained more collaborative. Still, a few participants felt that PromptCanvas sometimes obscured exactly what changed in the text. P20 noted it was hard to *"understand what parts … actually changed."* Meanwhile, some preferred Wordcraft for its local control, with P31 stating that highlighting a sentence made them *"only the sentence gets changed."* Overall, participants saw PromptCanvas as supporting higher-level creative control, while Wordcraft enabled tighter, sentence-level edits.

*6.1.7 Perception after Prolonged Use in the Wild.* In the interview following the two-week field study, all participants maintained a positive opinion of the system, finding the extended usage beneficial. When asked if the additional time was necessary for understanding the system or if the initial lab study time was sufficient, nine out of ten participants agreed that the extended time was not required, as they *"already had good understanding"* –P11. P3 mentioned that the core functionality was straightforward and *"with regards to how to use it, I don't think we need the extended time."* Other participants shared similar views, stating *"it's good having more time with it, so I could experiment and try different things"* –P15, and *"that I had more time to explore and more time to play around. So that's what actually having no time limit benefited me"* –P11. In contrast, P4 noted that in the lab study *"there was some time constraint and the system was new, so then it felt like the system was a bit difficult to understand, but when I had much time, then the system, the overall environment, became very intuitive to me and after some time, I had no issues in conducting the whole setup. So I think time helped a lot and the more I used the system, the more I started to like it."*

Over the two weeks, participants tested the system across various use cases. P2 engaged in writing motivational letters, book reviews, and stories. P3 focused on email replies and creative writing. P17 used it for critical analyses, while P11 explored prompt generation for image creation. P11, P13, and P14 tested programming tasks, with P14 remarking, *"It really performed well."* P14 also generated search engine optimized content and tested the system's ability to work with other languages by using Bengali content. Additionally, several participants expressed that image support would be both beneficial and an *"intuitive need"* –P17. P11 tested this by generating prompts within PromptCanvas and then using them in an external text-to-image generator, reporting positive outcomes. With extended use, participants became more effective in using the system. For example, P15 achieved better results using multiple widgets for small, modular aspects instead of high-level instructions. They adopted a workflow using widgets for modifications and prompts for

Table 5. Creativity Support Index (CSI) results from lab study 1 (N = 18). Values show means (M), standard deviations (SD), 95% confidence intervals (CI) for the paired mean difference, paired-samples $t$-test statistics, effect sizes (Cohen's $d$ for paired samples), and Holm-Bonferroni-corrected $p$-values.

| | Conversational UI | | PromptCanvas | | 95% CI | $t$ | $d$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| **Factor** | M | SD | M | SD | | | | |
| Enjoyment | 13.08 | 4.54 | 16.56 | 4.10 | [0.89, 6.06] | 2.84 | 0.67 | .02 |
| Exploration | 11.78 | 5.69 | 16.83 | 3.17 | [1.35, 8.76] | 2.88 | 0.68 | .02 |
| Expressiveness | 10.83 | 5.06 | 14.67 | 3.87 | [1.19, 6.48] | 3.06 | 0.72 | .02 |
| Immersion | 10.00 | 4.21 | 14.61 | 4.86 | [1.85, 7.37] | 3.53 | 0.83 | .01 |
| Results Worth Effort | 14.17 | 4.13 | 17.61 | 2.28 | [1.67, 5.22] | 4.10 | 0.97 | .005 |
| Overall CSI Score | 61.66 | 19.06 | 82.09 | 12.47 | [9.61, 31.26] | 3.98 | 0.94 | .005 |

Table 6. Creativity Support Index (CSI) results comparing Wordcraft and PromptCanvas (N = 18). Values as in Table 5.

| | Wordcraft | | PromptCanvas | | 95% CI | $t$ | $d$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| **Factor** | M | SD | M | SD | | | | |
| Enjoyment | 11.33 | 4.43 | 17.22 | 3.21 | [3.31, 8.47] | 4.81 | 1.13 | .00049 |
| Exploration | 11.94 | 4.62 | 17.89 | 1.88 | [3.58, 8.31] | 5.30 | 1.25 | .00029 |
| Expressiveness | 10.94 | 4.29 | 17.00 | 2.30 | [3.82, 8.29] | 5.72 | 1.35 | .00015 |
| Immersion | 8.83 | 4.40 | 13.22 | 5.15 | [0.92, 7.86] | 2.67 | 0.63 | .01623 |
| Results Worth Effort | 12.94 | 4.02 | 17.22 | 3.87 | [1.91, 6.65] | 3.81 | 0.90 | .00280 |
| Overall CSI Score | 56.39 | 18.94 | 83.30 | 10.80 | [15.97, 37.85] | 5.19 | 1.22 | .00030 |

editorial decisions, such as introducing a new character with a widget and later moving its introduction within the story.

### 6.2 How Does PromptCanvas Support Creativity and Exploration? (RQ2)

In lab study 1, PromptCanvas provided significantly stronger creativity support than the conversational baseline across all CSI factors (all Holm-Bonferroni-corrected $p < .03$; see Table 5). As our study did not involve collaboration, we omitted the collaboration factor following the practice from [16, 47] to avoid confusion. Participants rated PromptCanvas higher on Enjoyment, Exploration, Expressiveness, Immersion, and Results Worth Effort, with medium to large paired effect sizes ($d$ = .67-.97). For example, PromptCanvas yielded higher Enjoyment ($M$ = 16.56 vs. 13.08), Exploration ($M$ = 16.83 vs. 11.78), and Expressiveness ($M$ = 14.67 vs. 10.83). The confidence intervals for all paired differences excluded zero, indicating consistent improvements (e.g., Exploration: 95% CI [1.35, 8.76]). The Creativity Support Index results from our two-week-long field study, shown in Table 7, found similar results ($M = 79.73, SD = 17.73$).

In the final survey of lab study 1, participants directly compared the creativity support between PromptCanvas and the conversational UI (Figure 14). 67% of participants reported that they became so absorbed in the activity that they forgot about the tool they were using. 78% of participants chose PromptCanvas as the more expressive tool. 89% of participants found that they explored a wider range of ideas, options, designs, or outcomes using our system compared to using the conversational UI. Furthermore, 89% of participants reported a higher level of enjoyment with PromptCanvas than with the conversational UI. Using PromptCanvas made users feel their efforts were most worthwhile, with 67% feeling satisfied with what they produced relative to the effort expended.

Table 7. Creativity Support Index (CSI) results from our field study (N=10).

| Factor | Avg. Score | SD |
|---|---|---|
| Enjoyment | 16.50 | 3.58 |
| Exploration | 17.20 | 2.48 |
| Expressiveness | 15.10 | 4.85 |
| Immersion | 12.90 | 4.25 |
| Results Worth Effort | 16.40 | 3.75 |
| Overall CSI Score | 79.73 | 17.37 |

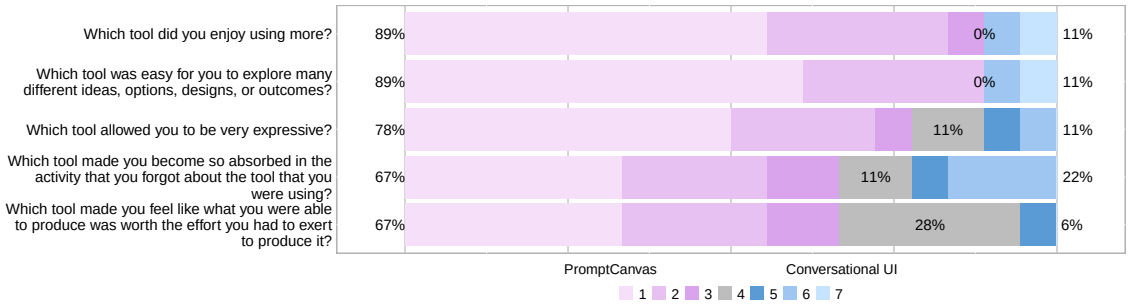


Fig. 14. (Lab study 1) Self-reported creativity support scores and preferences comparing PromptCanvas and the Conversational UI (N=18).

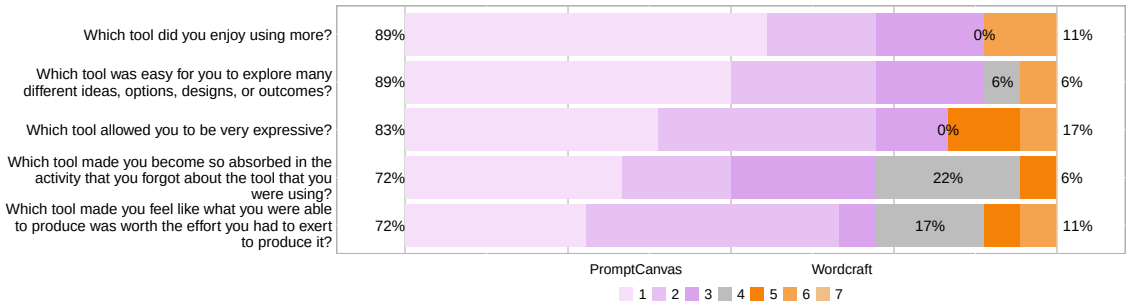


Fig. 15. (Lab study 2) Self-reported creativity support scores and preferences comparing PromptCanvas and Wordcraft (N=18).

These results were echoed in lab study 2 (Table 6), where PromptCanvas received substantially higher creativity support than Wordcraft. PromptCanvas produced higher scores on every CSI factor. All differences were statistically significant after Holm-Bonferroni correction ($p < .02$), with large effect sizes across factors ($d = 0.63$-$1.35$). Overall, PromptCanvas also showed significantly and markedly higher creativity support compared to Wordcraft.

In the final survey of lab study 2, participants directly compared the creativity support of PromptCanvas and Wordcraft (Figure 15). The majority consistently preferred PromptCanvas across all CSI dimensions. Enjoyment was strongly in favor of PromptCanvas, with 89% of participants reporting that they enjoyed using it more. A similarly large proportion (89%) indicated that PromptCanvas made it easier to explore a wide range of ideas, options, designs, or

outcomes. PromptCanvas was also judged to be more expressive, with 83% of participants selecting it over Wordcraft. Regarding immersion, 72% reported that PromptCanvas enabled them to become so absorbed in the activity that they forgot about the tool itself. Finally, 72% of participants felt that what they produced with PromptCanvas was worth the effort they invested, substantially higher than for Wordcraft.

*6.2.1 Exploration.* After the lab studies, all participants generally felt that the widgets helped them in finding unexplored areas or ideas. During lab study 1, P14 mentioned, *"For example, when I was writing about the story of AI robots, I was only thinking about a dystopian future where AI controls society. But when I got the suggested widgets, I was able to add the AI's influences on society specifically when AI will interfere with human affairs and national politics. It was not given at first in the generated story, but I could add it then. I couldn't think of it if the widgets didn't suggest this. So it was really helpful. It made the story more suspenseful and good."* P8 added, *"... when these dynamic widgets are suggesting something, it is also giving us the ideas to explore more in those areas, so I think this is really cool"*. Participants (P10, P13) also appreciated how the widgets allowed them to incorporate new ideas and details into their writing without reverting the text to its initial state.

Participants found PromptCanvas helpful for creative writing (P1, P2, P14). They noted that the tool encouraged them to think in multiple directions (P1, P2), *"I would say that it helped me to think more creative things, but when I was thinking, it helped me to think in multiple ways"* –P1. In addition, P3 added, *"If it's a more open-ended task ... you don't really know what ... the final form of your creative writing might be. Then I think those widgets would help ... having suggestions for more general themes, especially for the short story portion, would definitely help a lot"*. Furthermore, The dynamic nature of widgets was described as a "game changer" by P5, who valued the ability to generate multiple lines of thought quickly. This aspect allowed for rapid exploration of different story directions, *"Dynamic in nature, that's a game changer, as I said, because those lines of thought, I mean, I'm not saying that they wouldn't pop up, but it'll take a whole lot of time for me to... bring those up manually"*.

In lab study 2, some participants appreciated the variety and meaningfulness of the widgets, noting that the system often provided clearly differentiated widgets that supported exploration (e.g., P23, P30). P30 highlighted how these suggestions broadened their creative space, explaining that the widgets offered *"a wider area of things that maybe I didn't even think about"* and helped reduce the effort of manually structuring the story. In contrast, participants repeatedly said that Wordcraft limited exploration because of its linear, chat-like interaction style. Some thought it did not encourage branching or experimentation. For example, P19 explained that Wordcraft *"was not helping [exploration]… it was just normal"*. Several participants also stated that when they tried to change direction or revise earlier elements in Wordcraft, the tool struggled to follow, reducing their ability to explore. Collectively, participants viewed PromptCanvas as offering open-ended, flexible exploration, whereas Wordcraft tended to constrain exploration to a single unfolding narrative path.

*6.2.2 Diversity.* Participants mentioned that widgets help avert the monotony of the generated text by making it more personalized to the individuals (P4, P12). They appreciated the variety of suggestions provided, which improved the quality and helped bring out the personal styles of their writing, *"When you are writing something, sometimes we are just too absorbed in the generated result that we are not thinking about the possibilities in which we can improve our writing more, so having widgets that will suggest us more options can really improve the writings"* –P4. P4 also noted, *"I think in making people's writing more personal to them, they (widgets) just can be really helpful, the writings won't be monotonous"*.

Participants found PromptCanvas supports both structured input (P6) and diversity (P7, P10). P7 and P15 further mentioned the advantage of having visible prompts through widgets, facilitating continuous editing and refinement

of the text, *"What happens in conversational UI is that you continuously are giving prompts... so once a new prompt is given, and it has been applied the prompt goes out of the window. So when I'm having the widget, I can see what are my exact prompts that are functioning"* –P7. P15 valued the ability to see an overview of the story through widgets, which allowed for focused changes: *"It just comes down to seeing the overview of the story physically represented by the widgets. It felt like, again, that's how my mind pictures things. I like breaking things up into pieces. And so having the widgets, seeing that part of the story broken up means I could focus down on that and change it without changing the entire story. And then adding that into the text. And it only changed that part of the text"*. P15 added, *"... the widgets, having it broken down in front of me and seeing the individual things. So that would be a bit more creative, I think"*. Participants (P9, P12) also mentioned that widgets helped them think outside the box, offering diverse points and even aiding in non-creative tasks like writing biographies.

Participants also contrasted the diversity of ideas generated by PromptCanvas's widgets with the suggestions produced by Wordcraft. Many commented that PromptCanvas supplied a wide range of creative options, with one participant noting that it surfaced *"quite a lot of new options… interesting options, which I didn't think before"* –P34, and some appreciated that PromptCanvas *"always provides… the flexibility to get more suggestions"* ($P22^{pro.}$). This diversity helped participants break from a single narrative line and explore different stylistic and thematic directions.

In contrast, several participants described Wordcraft's suggestions as less diverse and sometimes repetitive, limiting the breadth of ideas. Some found that Wordcraft pushed them toward whatever the model generated, with P19 saying it *"just generates long text"*. Others noted that Wordcraft often followed a single direction and required considerable prompting to deviate from it (P19, P30). Several participants added that Wordcraft's inline rewriting tended to produce variations of the same idea rather than distinct alternatives, reducing overall creative diversity (P6, P7, P31). As a result, participants generally perceived PromptCanvas as providing richer and more varied creative stimuli, while Wordcraft offered more incremental but narrower expansions of the text.

*6.2.3 Outcome Text Quality Ratings.* To examine how each interface shaped the linguistic qualities of participants' writing, we analyzed three model-relevant text metrics: perplexity [38], lexical diversity using the Type-Token Ratio (TTR) [38], and readability via the Flesch Reading Ease score [21]. Perplexity reflects how predictable the text is to a language model, TTR captures vocabulary variety, and Flesch estimates clarity and accessibility. Together, these metrics characterize how each interface influenced the style and structure of the resulting texts.

In lab study 1, we observed significant differences in language-based outcomes between PromptCanvas and the conversational UI. PromptCanvas ($M = 0.647$) yielded text with significantly higher lexical diversity than the conversational UI ($M = 0.586$), as reflected in TTR scores, $t(17) = -3.93$, $p = .0011$, $d = 0.93$. PromptCanvas ($M = 45.3$) also produced lower Flesch Reading Ease scores than the conversational UI ($M = 57.3$), $t(17) = 3.14$, $p = .0060$, $d = 0.74$. However, these lower scores can also indicate a more formal or elaborate style rather than simply reduced readability. Finally, PromptCanvas ($M = 29.7$) generated text with higher perplexity than the conversational UI ($M = 23.0$), $t(17) = -4.62$, $p = .0002$, $d = 1.09$, indicating less predictable, more varied language. Overall, PromptCanvas encouraged participants to produce more diverse, stylistically complex, and less model-predictable text.

In lab study 2, we again observed differences in the textual output produced by the two systems. PromptCanvas ($M = 44.9$) generated text with significantly higher perplexity than the Wordcraft ($M = 35.1$), $t(17) = 2.77$, $p = .040$, $d = 0.65$, 95% CI [2.32, 17.30], indicating that the text was less predictable for the language model and exhibited greater surface-level variability. In contrast, no significant differences emerged for lexical diversity or readability: PromptCanvas produced slightly higher TTR scores ($M = 0.661$) than Wordcraft ($M = 0.615$), and slightly lower Flesch Reading Ease

Table 8. NASA-TLX and ease of completing tasks results comparing Conversational UI and PromptCanvas in lab study 1 and comparing Wordcraft and PromptCanvas in lab study 2. Values show 95% confidence intervals (CI) for the median paired difference, Wilcoxon signed-rank statistics ($W$), rank-based effect sizes ($r$), and Holm-Bonferroni corrected $p$-values.

| | Lab study 1 | | | | Lab study 2 | | | |
|---|---|---|---|---|---|---|---|---|
| **Factor** | **95% CI** | $W$ | $r$ | $p$ | **95% CI** | $W$ | $r$ | $p$ |
| Easy to Complete | $[-1.0,\ 2.5]$ | 21.0 | 0.41 | .460 | $[1,\ 3]$ | 12.0 | 0.70 | .035 |
| Mental Demand | $[-3,\ -1]$ | 9.5 | -0.67 | .117 | $[-3,\ -1]$ | 29.5 | -0.50 | .135 |
| Temporal Demand | $[-3,\ 1]$ | 67.0 | -0.11 | .647 | $[-3,\ 1]$ | 10.5 | -0.55 | .160 |
| Performance | $[-1.0,\ 2.5]$ | 17.0 | 0.34 | .560 | $[1,\ 3]$ | 6.0 | 0.75 | .046 |
| Effort | $[-4,\ 1]$ | 18.5 | -0.52 | .222 | $[-3,\ -1]$ | 3.5 | -0.77 | .052 |
| Frustration | $[-3.5,\ -1]$ | 2.5 | -0.77 | .138 | $[-4,\ -1]$ | 4.5 | -0.61 | .160 |

scores ($M = 64.5$) than Wordcraft ($M = 68.9$). These non-significant differences suggest that, aside from perplexity, both systems yielded comparable levels of vocabulary richness and readability.

### 6.3 How Does PromptCanvas Affect the Cognitive Load in Creative Writing? (RQ3)

We included RQ3 to evaluate how our concept impacts users' cognitive load, in line with prior works on metacognitive demands in generative AI [12, 25, 48]. Prior works show that traditional UIs can increase users' mental effort by requiring them to compose, remember, and revise complex (prompting) instructions in natural language, often without structured guidance or feedback. PromptCanvas makes key aspects of the writing and prompting process persistently visible and modular. Measuring cognitive load allows us to investigate whether this results in reduced mental demand and frustration.

#### *6.3.1 NASA-TLX and Ease of Completing Task.* .

Participants evaluated their perceived cognitive load using the NASA-TLX after using each interface in the lab studies (Table 8). In lab study 1, we did not observe statistically significant differences between the Conversational UI and PromptCanvas after applying Holm-Bonferroni correction for any factor or ease of completing the task.

In Lab study 2, where we compared Wordcraft against PromptCanvas, performance ratings favored PromptCanvas, with a 95% CI of $[1,\ 3]$ ($W = 6.0$, $r = .75$, $p = .046$), indicating that participants felt more effective and capable when using the system. Participants also rated tasks as easier to complete with PromptCanvas, reflected by a 95% CI of $[1,\ 3]$ ($W = 12.0$, $r = .70$, $p = .035$). Other dimensions, including mental demand, temporal Demand, effort, and frustration, showed no statistically significant differences after Holm-Bonferroni adjustment. Overall, these findings suggest that while overall workload levels were broadly comparable, PromptCanvas provided benefits in perceived ease of use and effectiveness in lab study 2.

In the final survey of lab study 1 (Figure 18), participants compared their perceived cognitive load between PromptCanvas and the conversational UI. PromptCanvas reduced frustration for 39% of participants (with 50% reporting no frustration in either UI), lowered temporal demand for 39%, and increased feelings of success for 61%. Additionally, 56% reported needing less effort when working with PromptCanvas.

Lab study 2 showed clear advantages for PromptCanvas (Figure 19) again in the final survey of lab study 2. Only 11% of participants felt rushed when using PromptCanvas, compared to 67% with Wordcraft. Success ratings were higher (83% felt more successful), frustration was lower (11% vs. 61%), and effort ratings again favored PromptCanvas (28% vs. 50%). Overall, 89% of participants preferred PromptCanvas over Wordcraft. Together, these findings reinforce the

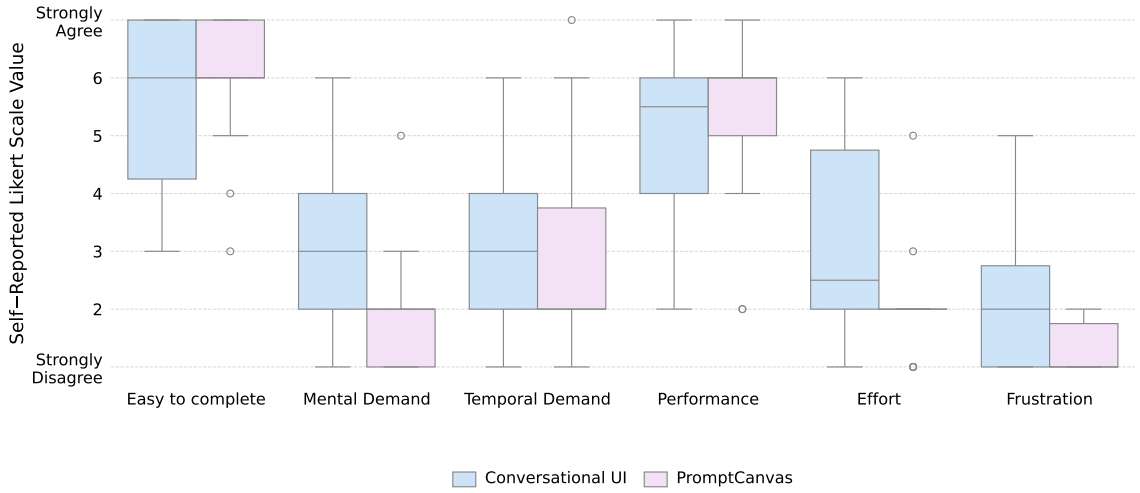


Fig. 16. Raw NASA-TLX scores (Lab study 1) by factor and ease of completing tasks for conversational UI and PromptCanvas (N=18).

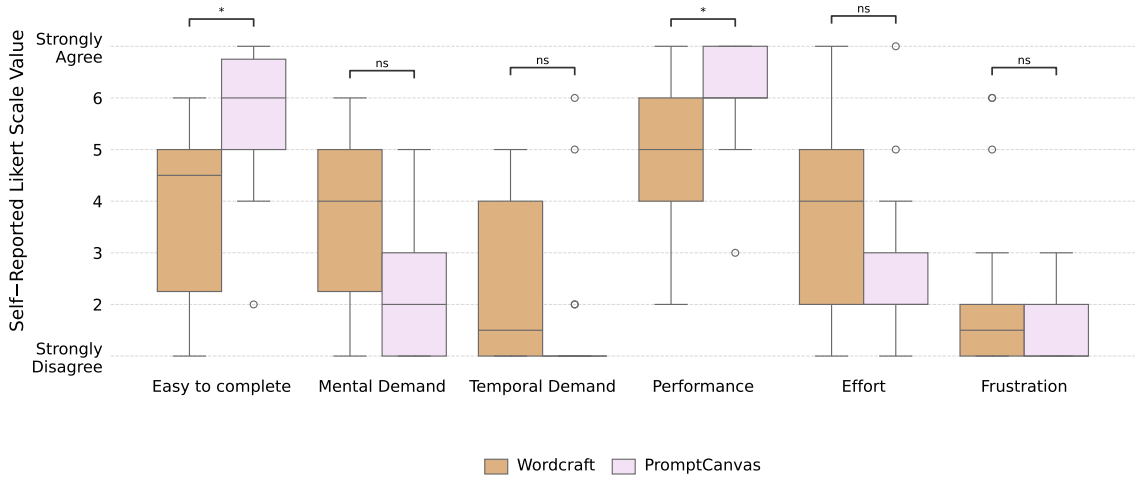


Fig. 17. Raw NASA-TLX scores (Study 2) by factor and ease of completing tasks comparing Wordcraft and PromptCanvas (N=18).

pattern across both studies: PromptCanvas lowers cognitive demands and supports a more confident, less stressful, and more satisfying creative workflow.

*6.3.2 Widget Complexity.* During the field study interviews, participants were asked if using widgets justified the additional complexity they introduced. P3 pointed out that the usefulness of widgets depended on the task's length: for short responses, they seemed unnecessary, but for longer or more creative outputs, like creative writing or full emails, they provided valuable control. Several participants appreciated how widgets reduced the need for multiple prompts. P11 preferred creating multiple widgets over composing several prompts. At the same time, P18$^{pro}$ found them

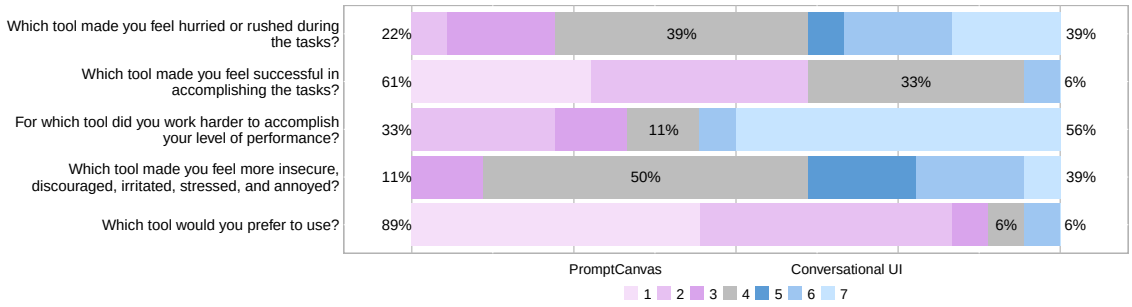


Fig. 18. (Lab study 1) Self-reported cognitive load and preference scores comparing PromptCanvas and the conversational UI (N=18).

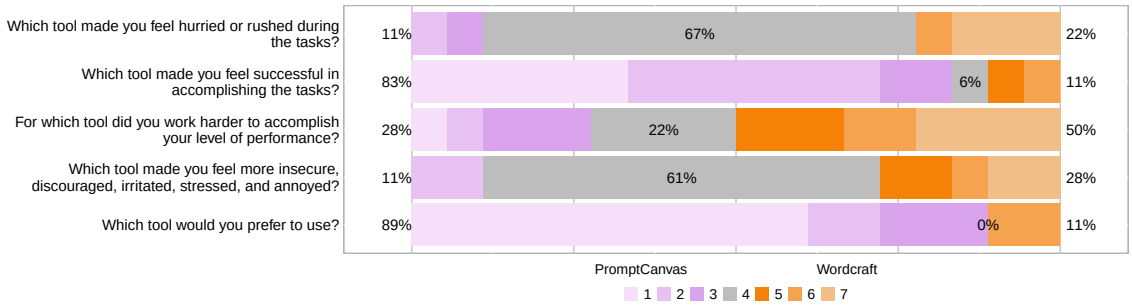


Fig. 19. (Lab study 2) Self-reported cognitive load and preference scores comparing PromptCanvas and Wordcraft (N=18).

*"definitely worth it,"* as they eliminated the need to *"write multiple prompts or go through this chain of changing prompts."* P18$^{pro}$ also mentioned that the widgets were useful as they visually preserved the context. P17 did not find the widgets complex, describing them as *"fantastic and very logical,"* with *"short sentences or short, relevant words."* However, P3 acknowledged that *"using widgets does take a bit more time."* P8 felt this time investment was worthwhile for creative tasks, saying, *"a bit time-consuming, however, I am actually putting that time to bring out my creativity and bring out new ideas, so if I'm focusing on writing."* While some participants noted that widgets added complexity, most viewed this as a beneficial trade-off, especially for tasks requiring detailed control and creativity.

*6.3.3 Efficiency and Efficacy.* In lab study 1, P14 highlighted the efficiency of PromptCanvas, noting that *"if I had to change anything in the generated text in the conversational UI, I have to give a comment again and again and check that out. But in PromptCanvas, it was quite efficient. I didn't have to wait for the full generation like this. And I can only give comments to a specific area of the generated text in the widgets. It was quite helpful."* Similarly, P17 appreciated the flexibility provided by PromptCanvas, saying *"The static one was more in one direction, so with the dynamic one (PromptCanvas), it was really like, okay, I don't have to push myself to you know harder to explore things or to create something and get the result."* Additionally, P18$^{pro}$ valued the ease of directing the tool with less manual intervention, stating *"I think that the widgets helped me to direct the tool to a more specific ... direction that I wanted to go without having to manually intervene as much. I think that we can do similar things with the conversational UI, but then I have to write a lot more prompts."* P18$^{pro}$ further added, *"So, even though the end result might be very similar, the efforts of getting there seems like it is definitely more with the conversational UI."*

In lab study 2, PromptCanvas was often described as more effective for shaping and refining ideas, with users explaining that its spatial layout *"made it easier to keep track of what I was doing"* (P29) and helped them *"quickly adjust or build on earlier thoughts"* (P33). P34 explained that PromptCanvas *"felt like a learning curve"* compared to Wordcraft. P32$^{pro.}$ noted that the canvas allowed them to "map out [their] whole thought process visually". Several noted that being able to manipulate multiple widgets at once improved their ability to experiment efficiently, ultimately leading to outputs they felt were *"more intentional"* –P31. In contrast, Wordcraft was seen as faster for producing immediate text, with participants acknowledging that its linear chat flow *"got something on the page quickly"* –P24$^{pro.}$. However, this speed sometimes came at the cost of control: users reported that Wordcraft responses could *"pull the story in directions I didn't want"* –P27, making revision less efficient. Overall, while Wordcraft supported quick generation, participants consistently felt that PromptCanvas enabled more effective task completion by allowing them to organize, iterate, and steer their creative process with greater precision.

### 6.4 System Usability

After the two-week field study and lab study 2, participants generally reported positive impressions of the system's usability and visual design. P17 described the interface as *"very good"* and *"eye soothing"*, while P15 noted that they *"like the colors ... it's easy to look at for a long time"*. P26 similarly emphasized that the interface *"made it a little more lively with the color"*, and P29 also characterized PromptCanvas as *"simple to use"*. Beyond aesthetics, P15 found the widgets intuitive and helpful for beginners, whereas P16 mentioned that the new widgets required some learning but ultimately offered clear advantages. Overall, participants perceived the visual tone, color palette, and layout of PromptCanvas as supportive and engaging, contributing positively to their writing experience.

*6.4.1 System Usability Scale.* After having engaged in lab study 1 and field study, all 10 participants in the field study completed the SUS questionnaire, which yielded an average score of 86.50 ($SD$ = 8.96), with a median of 90.00 and a range from 67.50 to 97.50, visualized in Figure 20. These results rank PromptCanvas well above the usability benchmark of 68, with most scores falling in the "excellent" range, typically above 85 [1]. The middle 50% of users rated the system between 80.63 and 93.75, underscoring strong satisfaction with the interface's usability. The minimum score of 67.50 is close to the acceptable threshold, while the maximum of 97.50 shows that some users found the system nearly flawless.

Furthermore, during lab study 2, we compared the System Usability Scale (SUS) scores for PromptCanvas and Wordcraft using a paired samples t-test. The SUS findings from Lab study 2 are consistent with the usability evaluations collected in the field study. PromptCanvas received significantly higher SUS scores than Wordcraft (Figure 21), $t(17) = 2.28$, $p = .036$, with a mean difference of 15.14 (95% CI [1.10, 29.18]) and a corresponding effect size of $d = 0.81$. Descriptively, Wordcraft received a mean SUS score of 70.83 ($SD$ = 21.16, $Med$ = 75), whereas PromptCanvas received a mean of 85.00 ($SD$ = 16.14, $Med$ = 92).

### 6.5 Which Tool Do Users Prefer?

This section compares how participants evaluated PromptCanvas across both lab studies: against the conversational UI in lab study 1, focusing on preference and ease of use, and against Wordcraft in lab study 2, including differences in writing approach.

*6.5.1 PromptCanvas Vs. Conversational UI.* The results of the lab study 1 indicate a strong preference for PromptCanvas compared to the baseline conversational UI.

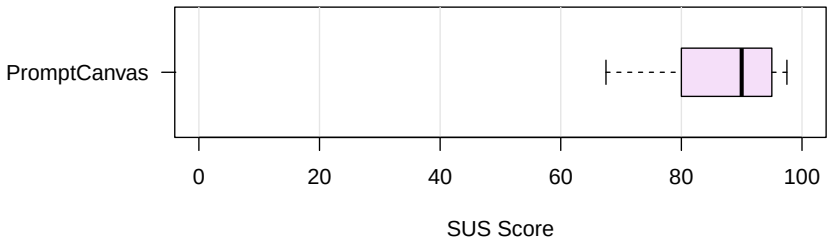


Fig. 20. System Usability Scale results from our field-study (N=10).

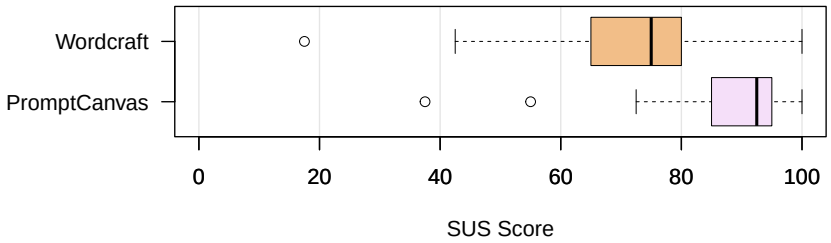


Fig. 21. (Lab study 2) System Usability Scale results from our final study (N=18) comparing Wordcraft and PromptCanvas.

*General Preference.* Notably, 89% preferred PromptCanvas over the conversational UI (Fig. 18). P1 recognized our tool as a canvas, highlighting its strengths in exploration and creativity. They suggested that while the conversational UI is suitable for time-constrained tasks, PromptCanvas might excel in scenarios where there is more time to explore options. *"I would suggest one thing. If you give a time test, I think this conversational UI is good. But if you don't have any time constraints, then PromptCanvas is very good. Because you are exploring many things"* –P1. Similarly, $P18^{pro}$ appreciated PromptCanvas's flexibility in selecting different story settings and character details, finding it beneficial for regular writing tasks: *"This is, I think, something that I would appreciate in regular use."*

*Ease of Use.* Participants also found it easier to use PromptCanvas. P4 found PromptCanvas particularly convenient, noting that it allowed for easy changes with simple clicks, contrasting this with the more cumbersome process in the conversational UI (baseline): *"PromptCanvas was a very convenient tool. In a conversational UI, I have to explain to the tool every time I need a change, but in a dynamic UI, I can achieve that with a simple click."* P10 appreciated the increased editing opportunities provided by PromptCanvas: *"It provides you with more possibilities to do."* P2 enjoyed the spontaneity of modifications without needing detailed instructions. Additionally, P12 found PromptCanvas versatile and engaging compared to the monotonous Conversational UI: *"The static tool was a bit monotonous ... I thought it had way fewer options, and the dynamic one was really versatile and not boring."*

*6.5.2 PromptCanvas Vs. Wordcraft.* The results of the lab study 2 again indicate a strong preference for PromptCanvas compared to Wordcraft.

*General Preference.* Notably again 89% preferred PromptCanvas over Wordcraft (Fig. 19) in lab study 2. Many described it as more enjoyable, flexible, and aligned with how they wanted to write. Some emphasized that it offered greater control over the story, with P27 noting that they felt *"more comfortable with PromptCanvas"* and that its dynamic

interface made it *"really fun"* to explore ideas. P33 also reported that the widget-based workflow felt *"more intuitive"* and provided useful features like switching workspaces and zooming for focus. Even participants who acknowledged a brief learning curve ultimately favored PromptCanvas. P34 explained that while the interface initially felt unfamiliar, *"once I got a hold of it, it was easier, and I felt it was more powerful."* In contrast, Wordcraft was occasionally preferred for its simplicity, but participants often described it as more restrictive or less under their control. Overall, the majority favored PromptCanvas for giving them clearer structure, greater creative agency, and a more engaging writing experience.

*Ease of Use.* Participants described clear differences in the ease of use of the two interfaces. Many found PromptCanvas more intuitive and immediately approachable, with P35 emphasizing that it was *"very intuitive, very easy"* and *"very easy to understand and convenient."* P29 similarly noted that both tools were straightforward but that PromptCanvas felt *"really intuitive and simple to use"* once they learned the basic interactions. Several participants also highlighted how the visual layout reduced confusion; P27 explained that PromptCanvas was *"more easy to use"* because its dynamic interface made it *"really fun playing with the widgets."* In contrast, Wordcraft's simplicity was appreciated but often described as limiting: P31 stated that Wordcraft was *"quite streamlined"* and *"very easy to use"* but lacked flexibility, while P35 cautioned that Wordcraft might be *"not that intuitive"* for some users. Others noted a learning curve in PromptCanvas (P34), but most felt that once understood, it offered a more supportive and accessible workflow overall.

*Approach to Writing.* Participants reported that their overall *approach to writing* shifted notably between the two interfaces. With PromptCanvas, P30 described a shift toward a more structured, modular process, explaining that PromptCanvas encouraged a *"divide-and-conquer approach"* where each widget could be arranged to give them clearer control over different parts of the story. P33 further highlighted this contrast, noting that Wordcraft led them toward a more *"bottom-up"* process of refining text line by line, whereas PromptCanvas supported a more *"top-down"* way of structuring ideas, allowing them to organize and adjust larger story elements before polishing the details. Others emphasized that the spatial layout encouraged planning and structure; P35 explained that PromptCanvas made it *"really easy to craft a draft… and visualize upfront"*. In contrast, Wordcraft led participants toward a more incremental process. P23 described it as *"a little bit more iterative,"* highlighting how they could *"start with a text"* and refine specific lines through highlighting and revision loops. Some even felt that their writing mindset changed depending on the tool. According to P23, PromptCanvas could at times make the process feel more hands-off, closer to proofreading, whereas Wordcraft encouraged them to *"iterate until it was fine."* Collectively, these accounts show how each interface not only shaped the content produced but also reoriented participants' underlying writing strategies, with PromptCanvas encouraging exploratory ideation and Wordcraft supporting focused, line-level iteration.

### 6.6 Objective Telemetry

We complement the subjective results with objective interaction data collected from system logs in both lab studies. Table 9, Table 10, and Table 12 summarize these telemetry metrics, capturing how participants engaged with each interface during the writing tasks.

*6.6.1 PromptCanvas vs. Conversational UI (Lab study 1).* As shown in Table 9 and Table 10, participants interacted with PromptCanvas in a more structured and fine-grained manner, creating, adjusting, and applying numerous widgets throughout the tasks. By contrast, the conversational UI centered almost entirely on a small number of prompt submissions. These patterns indicate that PromptCanvas distributed the writing process across many smaller, user-controlled operations, whereas the conversational UI concentrated activity into a few large prompt submissions.

Table 9. Telemetry data for the E-mail task in lab study 1, showing mean, standard deviation (SD), and median for each interaction type across interfaces.

| Interface | Telemetry | Mean | SD | Median |
|---|---|---|---|---|
| **PromptCanvas** | Widgets generated (guided) | 1.94 | 2.60 | 1.00 |
| | Widgets generated (suggested) | 7.22 | 3.26 | 7.00 |
| | Empty widgets created manually | 1.06 | 1.66 | 0.00 |
| | **All widgets created** | 10.22 | 3.61 | 11.00 |
| | | | | |
| | **'Apply widgets' requests** | 2.72 | 1.56 | 2.50 |
| | Active widgets when requesting 'Apply widgets' | 2.94 | 1.80 | 3.00 |
| | **'Apply prompt' requests** | 2.00 | 1.28 | 2.00 |
| **Static Baseline** | **'Apply prompt' requests** | 4.83 | 3.20 | 4.50 |

Table 10. Telemetry data for the Story task in lab study 1, showing mean, standard deviation (SD), and median for each interaction type across interfaces.

| Interface | Telemetry | Mean | SD | Median |
|---|---|---|---|---|
| **PromptCanvas** | Widgets generated (guided) | 3.61 | 3.73 | 3.50 |
| | Widgets generated (suggested) | 6.83 | 3.57 | 8.00 |
| | Empty widgets created manually | 1.39 | 1.82 | 0.00 |
| | **All widgets created** | 11.83 | 3.60 | 11.50 |
| | | | | |
| | **'Apply widgets' requests** | 3.78 | 1.73 | 4.00 |
| | Active widgets when requesting 'Apply widgets' | 3.64 | 2.17 | 4.00 |
| | **'Apply prompt' requests** | 2.17 | 1.62 | 2.00 |
| **Static Baseline** | **'Apply prompt' requests** | 6.33 | 4.00 | 6.50 |

*6.6.2 PromptCanvas in Field Study.* As shown in Table 12, participants in the field study made active use of PromptCanvas's widget-based controls. They generated a mix of suggested, guided, and manually created widgets, resulting in an average of nearly twelve widgets per session. Participants also applied widgets and prompts multiple times, indicating that both structured controls and free-form instructions were integral to their workflow. Overall, the telemetry reflects consistent engagement with widget generation and application as part of routine interaction with the system.

Table 11. Telemetry data from the field study for **PromptCanvas**, showing mean (M), standard deviation (SD), and median (MED) across participants.

| Telemetry | Mean | SD | Median |
|---|---|---|---|
| Widgets generated (guided) | 2.10 | 2.88 | 1.50 |
| Widgets generated (suggested) | 3.50 | 2.01 | 3.50 |
| Empty widgets created manually | 5.90 | 8.06 | 2.50 |
| **All widgets created** | 11.50 | 7.38 | 10.00 |
| | | | |
| **'Apply widgets' requests** | 9.40 | 6.77 | 8.00 |
| **'Apply prompt' requests** | 6.40 | 4.35 | 6.00 |

Table 12. Usage statistics for the story task in lab study 2, showing mean, standard deviation (SD), and median for each interaction type across interfaces.

| Interface | Telemetry | Mean | SD | Median |
|---|---|---|---|---|
| **PromptCanvas** | Story word count | 239.28 | 124.07 | 199.50 |
| | Minutes to complete | 9.17 | 3.47 | 9.10 |
| | Widgets generated (guided) | 3.22 | 2.80 | 3.00 |
| | Widgets generated (suggested) | 8.11 | 7.45 | 6.50 |
| | Empty widgets created manually | 2.00 | 1.37 | 1.50 |
| | **All widgets created** | 11.83 | 3.60 | 11.50 |
| | **'Apply widgets' requests** | 4.78 | 1.99 | 4.50 |
| | Active widgets when requesting 'Apply widgets' | 3.61 | 1.85 | 3.00 |
| | **'Apply prompt' requests** | 1.22 | 0.65 | 1.00 |
| **Wordcraft** | Story word count | 292.44 | 250.46 | 213.50 |
| | Minutes to complete | 7.89 | 1.72 | 7.79 |
| | Requests made | 6.72 | 2.02 | 7.00 |
| | Accepted suggestions | 5.56 | 2.18 | 5.00 |

Table 13. Wordcraft request types in lab study 2, showing the total number of requests made and the number of accepted suggestions with acceptance rate.

| Request Type | Requests Made | Suggestions Accepted (%) |
|---|---|---|
| Rewrite | 25 | 24 (96.00%) |
| Story seed | 23 | 16 (69.57%) |
| Suggest a prompt | 0 | 0 (0.00%) |
| Fill-in-the-blank | 7 | 4 (57.14%) |
| Continue | 16 | 14 (87.50%) |
| Next sentence | 1 | 1 (100.00%) |
| Elaborate | 5 | 2 (40.00%) |
| Custom | 44 | 39 (88.64%) |

*6.6.3 PromptCanvas vs. Wordcraft (Lab study 2).* In lab study 2 (Table 12), PromptCanvas and Wordcraft produced distinct interaction patterns shaped by their respective affordances. PromptCanvas encouraged participants to create and apply numerous widgets, leading to parallel adjustments with minimal prompting. Wordcraft, by contrast, centered on a request-response workflow composed of large as well as locally scoped edits such as rewrites and continuations. Sessions also differed in duration and output length, with PromptCanvas supporting longer interactions and Wordcraft yielding longer texts. The systems also diverged in responsiveness (Table 14), with PromptCanvas returning outputs faster than Wordcraft. Overall, the telemetry indicates that PromptCanvas encouraged compositional, workspace-based refinement, whereas Wordcraft supported a more linear sequence of inline editing operations.

Table 14. Model latency (in seconds) for all requests across interfaces in lab study 2, showing mean, standard deviation (SD), and median.

| Interface | Mean | SD | Median |
|---|---|---|---|
| Wordcraft | 8.69 | 3.63 | 8.97 |
| PromptCanvas | 4.47 | 2.05 | 4.15 |

### 6.7 Long-Form and Collaborative Writing with PromptCanvas

During the interview session for lab study 2, we explicitly asked participants about their perceptions of using PromptCanvas for long-form or collaborative writing (e.g., multiple chapters, multiple writers). Participants offered nuanced reflections on PromptCanvas's suitability for long-form and collaborative writing, often emphasizing the benefits of its spatial layout and multi-widget workflow. Several participants believed that the canvas structure would scale well to multi-chapter projects. P22$^{pro.}$ stated that *"for a long story, there can be different widgets… for each chapter,"* noting that chapter-wise arrangement could prevent the workspace from becoming chaotic. P24$^{pro.}$ similarly imagined long-form writing as feasible through *"different workspaces… for each person,"* which could later be combined into a unified draft. Others highlighted the collaborative potential of the canvas more explicitly. P26 described PromptCanvas as *"perfect for visual collaboration,"* explaining that multiple writers could work on different sections while maintaining a *"shared open view."* P31 also felt that PromptCanvas could be *"very seamless"* for group work because widgets could be assigned to different writers who could then *"improve those sections, and rewrite the whole story"*. P2 compared PromptCanvas to Miro, suggesting it could function similarly for collaborative brainstorming sessions. Overall, most participants saw clear potential for PromptCanvas in collaborative or extended projects, especially due to its visual workspace, organizational flexibility, and ability to separate story elements across widgets.

### 6.8 Cross-Study Interaction Strategies

To synthesize recurring prompt-refinement and visual reasoning behaviors observed across studies, Table 15 summarizes the strategies participants employed, their operational signals in the interface and telemetry, and where they were observed in (Lab 1, Lab 2, Field).

## 7 Limitations and Ethics

While PromptCanvas demonstrated strong usability and creativity support, several limitations and ethical considerations remain. We outline these considerations below, focusing on system performance characteristics, privacy and provenance implications, questions of authorship and disclosure, and the scope of our participant sample.

### 7.1 System Performance

We stress-tested PromptCanvas with workspaces ranging from about 50-150 widgets to examine scalability. We found that rendering stayed responsive, and even with 150 widgets, panning and zooming showed no frame-rate drops. However, the output quality degraded, especially when many widgets were applied to shorter text. This turned out to be less about the number of widgets and more about what the system was asked to do. A large set of simple widgets only adds a few hundred tokens and barely affects latency, but a small number of complex, global constraints can strain the model's reasoning (e.g., enforcing a specific rhyme scheme throughout an entire passage). In practice, latency mainly depends on total token count, while output quality suffers when too many instructions compete at once, leading to

Table 15. Prompt-refinement and visual reasoning strategies observed across studies. The table synthesizes interaction behaviors, observable interface signals, and qualitative evidence without introducing new measures.

| Strategy | Definition | UI / Telemetry Signals | Example Quote | Observed In |
|---|---|---|---|---|
| Attribute decomposition | Breaking high-level intent into modular, adjustable widgets | High number of widgets; repeated "Apply widgets" actions with few active widgets per request | "If I had to change anything in the generated text in the conversational UI, I have to give a comment again and again… But in PromptCanvas, it was quite efficient." (P14 in Lab 1) | Lab 1, Lab 2, Field |
| Parallel exploration | Maintaining and adjusting multiple constraints simultaneously | ≥3 active widgets per application; spatial separation of widgets on canvas | "Definitely worth it, because I don't have to write multiple prompts or go through this chain of changing prompts." ($P18^{pro.}$ in Field) | Lab 2, Field |
| Incremental refinement | Iteratively applying small adjustments rather than issuing new full prompts | Fewer free-form prompt submissions than baseline interfaces; frequent widget reuse and value edits | "I prefer creating multiple widgets over composing several prompts." (P11 in Field) | Lab 1, Field |
| Visual grounding of intent | Using spatial layout to externalize and reason about writing goals | Widget clustering, repositioning, and persistent on-canvas visibility | "The widget that controls the address of the company, I just put it beside where the address of the company actually is. So to maintain that spatial relationship ... the arranging of the widgets really helped" (P3 in Lab 1) | Lab 1, Lab 2, Field |
| Local and global adjustments | Applying widgets selectively to specific parts of the text (e.g., Protagonist's name) and then having the effect globally | Applying widgets in stages; selective activation/deactivation; chapter- or region-specific widget use | "For a long story, there can be different widgets… for each chapter." ($P22^{pro.}$ in Lab 2) | Lab 1, Lab 2, Field |

familiar issues such as some instructions being ignored, confusion about where a rule should apply, and drifting from earlier constraints.

### 7.2 Privacy and Provenance

Although PromptCanvas stores no personal data beyond local browser storage, future versions should provide clearer controls over how interaction history is captured, retained, or removed. Because AI-assisted writing involves layered transformations, users may benefit from mechanisms that visualize how specific widgets or prompts influenced the evolving text. ProvenanceWidgets [37] offers an instructive analogy: its lightweight overlays make interaction histories visible through simple, in-situ visualizations of frequency and recency. Incorporating a similar provenance layer into PromptCanvas would help users audit editing steps, understand how outputs evolved, and maintain transparency in mixed-initiative writing without requiring any additional data collection beyond the local device.

### 7.3 Authorship and Disclosure

As with prior systems such as Wordcraft [54] and professional-writer studies [26], issues of authorship, disclosure, and cognitive ownership arise when human and AI collaborate. Participants in our study similarly described moments where they were unsure which parts of the text reflected their own intentions versus the system's suggestions. Findings from lab study 2 reinforce this ambiguity. Although participants reported a clearer sense of control when working with PromptCanvas, evaluations of authorship did not differ reliably between the two interfaces. This was further complicated by the fact that most participants began by asking the system to produce an initial story draft, which made it difficult for them to assess the extent of their own contribution later on. These patterns suggest that authorship can be hard to discern in mixed-initiative writing workflows, regardless of interface design. Future work may need to consider ways, such as contribution metadata or revision provenance, that help users understand how their input and the system's output are entangled when producing AI-assisted writing.

### 7.4 Generalizability

Our findings should be interpreted within the evaluated scope: short-form and extended creative writing tasks involving iterative revision (e.g., emails, stories, and multi-paragraph drafts), conducted with participants comfortable experimenting with new interfaces. While the underlying interaction principles may generalize, we do not claim that PromptCanvas in its current form improves performance (e.g., writing efficiency, output quality, or cognitive load) for all writing domains, populations, or production-scale workflows.

Across the two lab studies (18 participants in each, total 36) and the field study (10 participants), our sample offers a reasonable range of perspectives but still does not represent the full diversity of writing practices or domains. Future research could broaden participant diversity further and include a dedicated subgroup of professional writers to examine how their workflows and expectations align with or diverge from those observed here. It may also be interesting to evaluate PromptCanvas in additional domains, such as technical writing or SEO-oriented tasks, to assess whether the system's benefits and limitations generalize beyond the creative-writing scenarios investigated in this work.

## 8 Discussion and Future Work

In this work, we investigated how dynamic widgets can be used to improve user interaction and creativity and provide greater control over generated content **(RQ1)**. We also assessed whether dynamic widgets for iterative and structured prompting offer better creativity support compared to a conversational UI **(RQ2)** and structured writing tool, Wordcraft [54], and explored whether they help reduce cognitive load in creative writing tasks **(RQ3)**. The results indicate that PromptCanvas enables a more exploratory and iterative approach, allowing users to experiment with different aspects of their work more freely. Below, we will address some of the current limitations while emphasizing opportunities for future research.

### 8.1 Widget Organization and Interdependence

PromptCanvas supports real-time widget creation through user-entered prompts, system suggestions based on draft text, and from scratch by double-clicking on the canvas. In addition, each widget includes contextual suggestions as shown in Figure 4 to help users formulate or refine attribute values, providing authoring support within the UI. Each widget includes not only a key:value pair but also a panel of context-aware suggestions and multiple interaction options like extracting values from the text or generating alternatives. With these features, PromptCanvas supports

creating widgets and then iteratively refining their content, which goes beyond pure widget suggestions. The objective telemetry data further support this. PromptCanvas sessions involved widget creation and parallel adjustments, reflecting a more compositional and exploratory workflow, whereas Wordcraft interactions were characterized by shorter, linear sequences of request-response edits.

While the features in PromptCanvas supported rich and exploratory workflows, participants' experiences also highlighted challenges that emerged as the number of widgets increased. Here, we describe these challenges and their implications for interface design .

*Widget-induced instruction overload and canvas clutter.* Although multiple participants found our UI to be aesthetically pleasing, participants encountered issues with organizing widgets on the canvas (see Section 6.1). As widgets accumulated, participants reported increased cognitive load, leading to visual clutter and higher navigation effort, and in some cases reduced output quality when multiple constraints competed over short or moderately scoped text. In these situations, the model could apply constraints inconsistently or drift from earlier specifications, reflecting instruction overload rather than the mere number of widgets.

To improve navigation and traceability, participants suggested organizing widgets through clustering by purpose, color-coding to distinguish functions, and highlighting text regions or changes associated with specific widgets. Beyond usability benefits, such an organization (currently user-controlled) could also serve as an additional input channel to the AI by conveying relationships between closely grouped concepts. Together, these observations suggest that scalable widget systems should support mechanisms for managing constraint complexity, including grouping related widgets, collapsing inactive ones, scoping widgets to specific text regions, and applying constraints incrementally rather than simultaneously to reduce overload in complex workflows.

*Onboarding and conceptual overhead.* Widgets introduce an upfront conceptual cost compared to single-prompt interfaces. Participants noted that creating and managing widgets can take more time, especially for short or low-stakes tasks where the benefits of persistent structure are less immediate. As a result, widgets may be less suitable for brief, one-off interactions.

These observations suggest hybrid or adaptive designs that progressively introduce widgets, use system-suggested defaults, or adjust interface complexity based on task length and user experience to reduce initial friction.

*Widget interdependence and semantic coupling.* The widgets currently function as largely self-contained interface elements and are not coupled to one another. However, changes to one widget may have semantic implications for others, requiring corresponding updates to their values or available options. For example, changing a character's attributes in a story (e.g., from daughter to son) would likely affect related constraints such as proposed names. Supporting such semantic coupling between widgets represents an opportunity for future iterations of widget interfaces.

### 8.2 Supporting Creativity and Exploration

PromptCanvas significantly supports creativity compared to the conversational UI as well as Wordcraft (**RQ2**). The results from Section 6.2 demonstrate that users find PromptCanvas more helpful in creative writing and idea generation, with creativity support scores remaining similarly high across both lab studies, highlighting the robustness of this effect. Importantly, the text outcome quality ratings echo these subjective findings, PromptCanvas consistently produced text that was more varied and exploratory. These results also suggest that most users explored themes they might not have considered without the dynamic widget suggestions provided by PromptCanvas, similar to observations by Vaithilingam

et al. [50] and Barke et al. [6].Unlike the static nature of traditional conversational UIs or structured writing tools, PromptCanvas offers a dynamic, interactive environment where users can manipulate widgets to explore and refine their creative outputs with better enjoyment (Table 6). In contrast, the structured writing tool Wordcraft offered more incremental but narrower expansions of the text. Finally, participants generally perceived PromptCanvas as providing richer and more varied creative stimuli, while Wordcraft offered more incremental but narrower expansions of the text. This approach not only enables a more engaging creative process but also supports non-professional writers by providing intuitive tools that help them overcome obstacles in creative writing. By enabling users to experiment with different text elements and receive contextually relevant suggestions, PromptCanvas empowers individuals with less experience in creative writing to generate and express their ideas more effectively (Section 6.3).

### 8.3 Supporting Customizability and Control

Our study suggests that PromptCanvas offers advantages in personalization and customizability (Section 6.1) through its dynamic widgets (**RQ1**). These widgets allow users to customize their interface to better suit individual preferences and needs. Users also reported feeling a greater sense of control over the generated text (Section 6.1), which aligns with the principles of managing one's cognitive resources [48]. According to Tankelevitch et al. [48], metacognition is the psychological ability to monitor and regulate one's thoughts and behaviors, and generative AI systems place metacognitive demands on users. We envision that the dynamic widgets of our generative AI frontend, PromptCanvas, can reduce metacognitive demand by enhancing customizability. Customizability can support cognitive control by providing a more adaptable interface [29].

Moreover, current generative AI systems often require verbalized prompting, demanding self-awareness of task goals and decomposition of tasks into sub-tasks [48]. The results from Section 6.2 show that participants viewed dynamic widgets as a means to break down tasks, enhancing their ability to manage and execute complex workflows. Additionally, users found that using PromptCanvas made the results worth the effort they put in (Table 5). By guiding reflection and prompting towards a more structured and interactive use of generative AI, our concept of PromptCanvas might be able to help users engage more deeply with their creative tasks.

Results from Section 6.1 indicate that, in the field study, participants used PromptCanvas not only for creative tasks but also for programming, demonstrating the system's versatility beyond creative writing. This extension into non-writing tasks further emphasizes the customizability and adaptability of PromptCanvas, allowing users to personalize their experience for a variety of workflows. This broader applicability highlights the system's potential to serve diverse needs, extending beyond writing to other domains such as coding.

### 8.4 Reducing Cognitive Load

Although the dynamic nature of PromptCanvas could be expected to increase cognitive load [50], the results from Section 6.3 show that overall workload levels were broadly comparable across interfaces. In lab study 1, no NASA-TLX factor differed significantly after Holm-Bonferroni correction. In lab study 2, however, participants felt more effective and capable when using PromptCanvas, reflected in significant differences in performance. Participants also described feeling less irritated, more productive, and better supported during exploration, suggesting that the dynamic, widget-based interface helps them manage creative writing tasks more effectively. By allowing users to manipulate widgets for contextual suggestions and seamless exploration, PromptCanvas creates a supportive environment for managing creative writing tasks.

### 8.5 Extending the Concept to Long-Form, Collaborative, and Other Domains

Inspired by previous research in visualization (e.g., DynaVis [50]) and systems like Luminate [46], we expect the concept of PromptCanvas to be applicable more broadly. Findings from our study already point in this direction. Results from Section 6.1 suggest that PromptCanvas could extend beyond text-based creative tasks to visual content generation, where dynamic widgets may support iterative refinement of images or other media. Such workflows would allow users to adjust stylistic parameters, inject specific elements, or explore alternative visual directions, demonstrating the versatility of a widget-based prompting paradigm across creative domains. The canvas design also supports a wide range of widget types. Future additions could include standard interface elements (e.g., date pickers, sliders, checkboxes), similar to those in DynaVis, or domain-specific tools like those in Textoshop [34], enabling more targeted customization and user control.

Our results also indicate that PromptCanvas could support long-form and collaborative writing. Participants in lab study 2 noted that its spatial layout and multi-widget workflow may scale to multi-chapter or multi-author projects, with separate sections that can be organized and later combined. They also felt the shared visual workspace could facilitate parallel work and coordination across contributors. These reflections suggest potential for applying the PromptCanvas approach to extended or multi-author workflows.

## 9 Conclusion

In this work, we introduce PromptCanvas, using dynamic widgets as a novel solution to address the limitations of current UIs for generative AI in creative writing. Our studies demonstrate that by incorporating customizable interactive elements, our system enhances user control, reduces cognitive load, and supports iterative exploration and the creation of a personalized workspace. The findings reveal that dynamic widgets significantly improve user experience and facilitate more effective, user-centered interaction with AI. This research emphasizes the importance of user-driven customization and flexibility in unlocking the creative potential of humans in AI-driven writing support, leading to more meaningful and productive writing interactions.

## References


[1] Philip T. Kortum Aaron Bangor and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction* 24, 6 (2008), 574–594. doi:10.1080/10447310802205776 arXiv:https://doi.org/10.1080/10447310802205776

[2] Christopher Ahlberg and Ben Shneiderman. 1994. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 313–317. doi:10.1145/191666.191775

[3] Belle Dang Andy Nguyen, Yvonne Hong and Xiaoshan Huang. 2024. Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education* 49, 5 (2024), 847–864. doi:10.1080/03075079.2024.2323593

[4] Tyler Angert, Miroslav Suzara, Jenny Han, Christopher Pondoc, and Hariharan Subramonyam. 2023. Spellburst: A Node-based Interface for Exploratory Creative Coding with Natural Language Prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 100, 22 pages. doi:10.1145/3586183.3606719

[5] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. 2021. Program Synthesis with Large Language Models. doi:10.48550/arXiv.2108.07732 arXiv:2108.07732

[6] Shraddha Barke, Michael B. James, and Nadia Polikarpova. 2023. Grounded Copilot: How Programmers Interact with Code-Generating Models. *Proc. ACM Program. Lang.* 7, OOPSLA1, Article 78 (apr 2023), 27 pages. doi:10.1145/3586030

[7] Michel Beaudouin-Lafon and Wendy E. Mackay. 2000. Reification, polymorphism and reuse: three principles for designing visual interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Palermo, Italy) *(AVI '00)*. Association for Computing Machinery, New York, NY, USA, 102–109. doi:10.1145/345513.345267

[8] 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. doi:10.1145/3581641.3584060

[9] Michael Mose Biskjaer, Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, and Peter Dalsgaard. 2019. How Time Constraints in a Creativity Support Tool Affect the Creative Writing Experience. In *Proceedings of the 31st European Conference on Cognitive Ergonomics* (BELFAST, United Kingdom) *(ECCE '19)*. Association for Computing Machinery, New York, NY, USA, 100–107. doi:10.1145/3335082.3335084

[10] Stephen Brade, Bryan Wang, Mauricio Sousa, Sageev Oore, and Tovi Grossman. 2023. Promptify: Text-to-Image Generation through Interactive Prompt Exploration with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–14. doi:10.1145/3586183.3606725

[11] John Brooke. 1996. SUS: A 'Quick' and 'Dirty' Usability Scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian Lyall McClelland (Eds.). Taylor and Francis, Chapter 21, 189–194.

[12] Daniel Buschek. 2024. Collage is the New Writing: Exploring the Fragmentation of Text and User Interfaces in AI Tools. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2719–2737. doi:10.1145/3643834.3660681

[13] Daniel Buschek, Martin Zürn, and Malin Eiband. 2021. The Impact of Multiple Parallel Phrase Suggestions on Email Input and Composition Behaviour of Native and Non-Native English Writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 732, 13 pages. doi:10.1145/3411764.3445372

[14] Yining Cao, Peiling Jiang, and Haijun Xia. 2025. Generative and Malleable User Interfaces with Generative and Evolving Task-Driven Data Model. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 686, 20 pages. doi:10.1145/3706598.3713285

[15] Dashiel Carrera, Zixin Zhao, Ashish Ajin Thomas, and Daniel Wigdor. 2025. Nabokov's Cards: An AI Assisted Prewriting System to Support Bottom-Up Creative Writing. In *Proceedings of the 2025 Conference on Creativity and Cognition (C&C '25)*. Association for Computing Machinery, New York, NY, USA, 546–559. doi:10.1145/3698061.3726916

[16] Erin A. Carroll, Celine Latulipe, Richard Fung, and Michael Terry. 2009. Creativity factor evaluation: towards a standardized survey metric for creativity support. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition* (Berkeley, California, USA) *(C&C '09)*. Association for Computing Machinery, New York, NY, USA, 127–136. doi:10.1145/1640233.1640255

[17] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. doi:10.1145/2617588

[18] 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* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 98, 13 pages. doi:10.1145/3526113.3545672

[19] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice Over Control: How Users Write with Large Language Models using Diegetic and Non-Diegetic Prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 408, 17 pages. doi:10.1145/3544548.3580969

[20] Leah Findlater and Krzysztof Z. Gajos. 2009. Design Space and Evaluation Challenges of Adaptive Graphical User Interfaces. *AI Magazine* 30, 4 (Sep. 2009), 68. doi:10.1609/aimag.v30i4.2268

[21] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[22] Iona Gilburt. 2024. A machine in the loop: the peculiar intervention of artificial intelligence in writer's block. *New Writing* 21, 1 (2024), 26–37.

[23] Charles Goodwin. 2015. Professional vision. In *Aufmerksamkeit: Geschichte-theorie-empirie*. Springer, Cham, Switzerland, 387–425.

[24] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[25] John R. Hayes. 2012. Modeling and Remodeling Writing. *Written Communication* 29, 3 (2012), 369–388. doi:10.1177/0741088312451260

[26] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. doi:10.48550/ARXIV.2211.05030 Version Number: 1.

[27] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–20. doi:10.1145/3586183.3606737

[28] Nimisha Karnatak, Adrien Baranes, Rob Marchant, Triona Butler, and Kristen Olson. 2025. ACAI for SBOs: AI Co-creation for Advertising and Inspiration for Small Business Owners. arXiv:2503.06729 [cs.HC] https://arxiv.org/abs/2503.06729

[29] Abdulrahman Khamaj and Abdulelah M. Ali. 2024. Adapting user experience with reinforcement learning: Personalizing interfaces based on user behavior analysis in real-time. *Alexandria Engineering Journal* 95 (2024), 164–173. doi:10.1016/j.aej.2024.03.045

[30] Matthew G Kirschenbaum. 2016. *Track changes: A literary history of word processing*. Harvard University Press, Cambridge, MA.

[31] Max Kreminski. 2024. The Dearth of the Author in AI-Supported Writing. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants* (Honolulu, HI, USA) *(In2Writing '24)*. Association for Computing Machinery, New York, NY, USA, 48–50. doi:10.1145/3690712.3690725

[32] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho,

Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. doi:10.1145/3613904.3642697

[33] 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* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 388, 19 pages. doi:10.1145/3491102.3502030

[34] Damien Masson, Young-Ho Kim, and Fanny Chevalier. 2024. Textoshop: Interactions Inspired by Drawing Software to Facilitate Text Editing. arXiv:2409.17088

[35] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 975, 16 pages. doi:10.1145/3613904.3642462

[36] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 355, 34 pages. doi:10.1145/3544548.3581225

[37] Arpit Narechania, Kaustubh Odak, Mennatallah El-Assady, and Alex Endert. 2025. ProvenanceWidgets: A Library of UI Control Elements to Track and Dynamically Overlay Analytic Provenance. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 1235–1245. doi:10.1109/TVCG.2024.3456144

[38] Vishakh Padmakumar and He He. 2024. Does Writing with Language Models Reduce Content Diversity? arXiv:2309.05196 [cs.CL] https://arxiv.org/abs/2309.05196

[39] Zhenhui Peng, Xingbo Wang, Qiushi Han, Junkai Zhu, Xiaojuan Ma, and Huamin Qu. 2023. Storyfier: Exploring Vocabulary Learning Support with Text Generation Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 46, 16 pages. doi:10.1145/3586183.3606786

[40] Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1051, 19 pages. doi:10.1145/3613904.3642105

[41] Emily Reif, Crystal Qian, James Wexler, and Minsuk Kahng. 2024. Automatic Histograms: Leveraging Language Models for Text Dataset Exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 53, 9 pages. doi:10.1145/3613905.3650798

[42] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 365–377. doi:10.1145/2984511.2984588

[43] Ben Shneiderman. 1983. Direct manipulation: A step beyond programming languages. *Computer* 16, 08 (1983), 57–69.

[44] Arjun Srinivasan and Vidya Setlur. 2023. BOLT: A Natural Language Interface for Dashboard Authoring. In *EuroVis 2023 - Short Papers*. The Eurographics Association, Eindhoven, The Netherlands, 5. doi:10.2312/evs.20231035

[45] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–26. doi:10.1145/3613904.3642400

[46] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 644, 26 pages. doi:10.1145/3613904.3642400

[47] 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* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 16 pages. doi:10.1145/3526113.3545617

[48] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 680, 24 pages. doi:10.1145/3613904.3642902

[49] Kashyap Todi, Gilles Bailly, Luis A. Leiva, and Antti Oulasvirta. 2021. Adapting User Interfaces with Model-based Reinforcement Learning. arXiv:2103.06807

[50] Priyan Vaithilingam, Elena L. Glassman, Jeevana Priya Inala, and Chenglong Wang. 2024. DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 985, 17 pages. doi:10.1145/3613904.3642639

[51] Priyan Vaithilingam and Philip J. Guo. 2019. Bespoke: Interactively Synthesizing Custom GUIs from Command-Line Applications By Demonstration. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 563–576. doi:10.1145/3332165.3347944

[52] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–21. doi:10.1145/3613904.3642803

[53] Jason Wu, Kashyap Todi, Joannes Chan, Brad A Myers, and Ben Lafreniere. 2024. FrameKit: A Tool for Authoring Adaptive UIs Using Keyframes. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 660–674. doi:10.1145/3640543.3645176

[54] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. doi:10.1145/3490099.3511105

[55] J.D. 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. doi:10.1145/3544548.3581388

[56] 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 *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 30 pages. doi:10.1145/3586183.3606800

## Appendix

### Technical Details

Table 16. The prompt used in PromptCanvas for generating control widget. The bold text in the 'Prompt' column is a placeholder for the real values to be inserted. The italic text will not be included if its variable is empty.

| Prompt | Example Inputs | Example Response |
|---|---|---|
| SYSTEM MESSAGE:<br>Objective:<br>Your task is to analyze the provided text and isolate individual attributes, transforming them into distinct control widgets.<br>These widgets represent different components of the text in an editable format, allowing users to modify specific aspects of the text.<br><br>Control Widgets Definition:<br>Each control widget should serve as an abstract representation of a specific element or attribute of the text.<br>The generated widgets should balance creativity and usability, making them practical for text modification and valuable to the user.<br>Properties of each control widget:<br>1. Title:<br>- A brief description that indicates the specific element of the text it represents and modifies.<br>2. Value:<br>- A representation of the text's current value corresponding to the widget's focus.<br>3. Options:<br>- A set of up to 3 generated suggestions related to the widget's focus and the provided text.<br>- These suggestions should spark creativity and offer alternative formulations.<br><br>Additional Requirements:<br>- Unless stated otherwise, produce between 1 and 4 widgets for each response.<br>- Value usefulness over quantity.<br>- Adapt the voice and tone used in the provided context.<br>- Maintain a non-verbose and journalistic style.<br><br>USER MESSAGE:<br>*I want you to generate widgets specifically designed to modify the following aspect:*<br>*"""**prompt**"""*<br><br>*Do not include any of the following widgets, as they have already been generated:*<br>*"""**existingLabels**"""*<br><br>Base the widgets on the following text:<br>"""**text**""" | ***prompt*** *= "Princess's Appearance"*<br><br>***existingLabels*** *= "Story Length, Plot"*<br><br>**text** = "The knight defeated the evil dragon and saved the princess." | {<br>"items": [<br>{<br>"label": "Princess's Eye Color",<br>"value": "Unknown",<br>"options": [<br>"Sapphire Blue",<br>"Green",<br>"Golden Brown"<br>]<br>},<br>{<br>"label": "Princess's Attire",<br>"value": "Unknown",<br>"options": [<br>"Traditional gown",<br>"Armored battle dress",<br>"Silk kimono"<br>]<br>},<br>{<br>"label": "Princess's Accessories",<br>"value": "Unknown",<br>"options": [<br>"Silver hairpins with moonstones",<br>"Amulet with crystal charms",<br>"Jeweled scepter"<br>]<br>}<br>]<br>} |

Table 17. The prompt used in PromptCanvas for generating more options for a control widget. The bold text in the 'Prompt' column is a placeholder for the real values to be inserted. The italic text will not be included if its variable is empty.

| Prompt | Example Inputs | Example Response |
|---|---|---|
| SYSTEM MESSAGE:<br>You will receive a text and one extracted attribute from that text. The user can modify this attribute by selecting a value from a set of options.<br>Your task is to return a set of value suggestions for the specified attribute.<br><br>Additional Requirements:<br>- Ensure the suggested options are relevant and useful for modifying the attribute in the provided context.<br>- Ensure the suggested options are designed to inspire creativity and fresh ideas.<br>- Adapt the voice and tone used in the provided context.<br>- Maintain a non-verbose and journalistic style.<br>- Generate only 2 options per response!<br><br>USER MESSAGE: Attribute:<br>"""**label**"""<br><br>*Do not include the following options because you have already generated these:* ***options***<br><br>Text:<br>"""**text**""" | **label** = "Story Length"<br><br>***options*** *= "One Sentence, Epic Saga, Short Story"*<br><br>**text** = "The knight defeated the dragon and saved the princess." | {<br>"items": [<br>"Two Sentences",<br>"Haiku"<br>]<br>} |

Table 18. The prompt template for extracting a widget's current value from the text editor's content. The bold text represents variables, for the real values to be inserted.

| Prompt | Example Inputs | Example Response |
|---|---|---|
| SYSTEM MESSAGE:<br>1. Read the provided text and identify the specified attribute.<br><br>2. Generate a value that represents the attribute in the context of the text.<br><br>- The value should be a concise, clear and descriptive representation of the attribute's essence.<br><br>USER MESSAGE:<br>Attribute:<br>"""**label**"""<br><br>Text:<br>"""**context**""" | **label** = "Rescued Entity"<br><br>**context** = "The knight defeated the evil dragon and saved the princess." | "princess" |

Table 19. The prompt used in PromptCanvas for manipulating the text editor's content based on the control widgets. The bold text in the 'Prompt' column is a placeholder for the real values to be inserted.

| Prompt | Example Inputs | Example Response |
|---|---|---|
| SYSTEM MESSAGE:<br>You will receive a text and a set of attributes with corresponding values. Each attribute-value is a specification for the text.<br><br>Your task is to apply these to the given text:<br>1. Read the text and the provided attributes carefully and gain an abstract understanding of the text.<br>2. Rephrase the attribute-values into natural language descriptions that guide how to modify the text.<br>3. Modify the text to match the provided values for each attribute.<br>4. Return only the modified text as a response.<br><br>Ensure that the final modified text is coherent, logical.<br>Maintain the original context and meaning, as well as the voice and tone used in the provided text.<br><br>USER MESSAGE:<br>Requirements:<br>"""**requirements**"""<br><br>Text:<br>"""**text**""" | **requirements** = {<br>"Story Length": "One sentence",<br>"Knight's name": "Sir Ucci",<br>"Princess's name": "Marlies",<br>}<br><br>**text** = "The knight defeated the dragon and saved the princess." | "Sir Ucci defeated the dragon and saved Princess Marlies." |

Table 20. The prompt used in PromptCanvas for manipulating the text editor's content with natural language prompts. The bold text in the 'Prompt' column is a placeholder for the real values to be inserted.

| Prompt | Example Inputs | Example Response |
|---|---|---|
| SYSTEM MESSAGE:<br>Please apply the following prompt to the provided text.<br>The new text should follow the instructions in the prompt while preserving the original context and meaning.<br>Always return the modified text as a response.<br>If the provided text is empty, generate a new text based on the prompt.<br><br>USER MESSAGE:<br>Prompt:<br>"""**prompt**"""<br><br>Text:<br>"""**text**""" | **prompt** = "Replace the dragon with a group of bandits."<br><br>**text** = "The knight defeated the dragon and saved the princess." | "The knight defeated the group of bandits and saved the princess." |