

GhostWriter: Augmenting Collaborative Human-AI Writing Experiences Through Personalization and Agency

CATHERINE YEH, Harvard University, USA

GONZALO RAMOS, Microsoft Research, USA

RACHEL NG, Microsoft, USA

ANDY HUNTINGTON, Microsoft, UK

RICHARD BANKS, Microsoft Research, UK

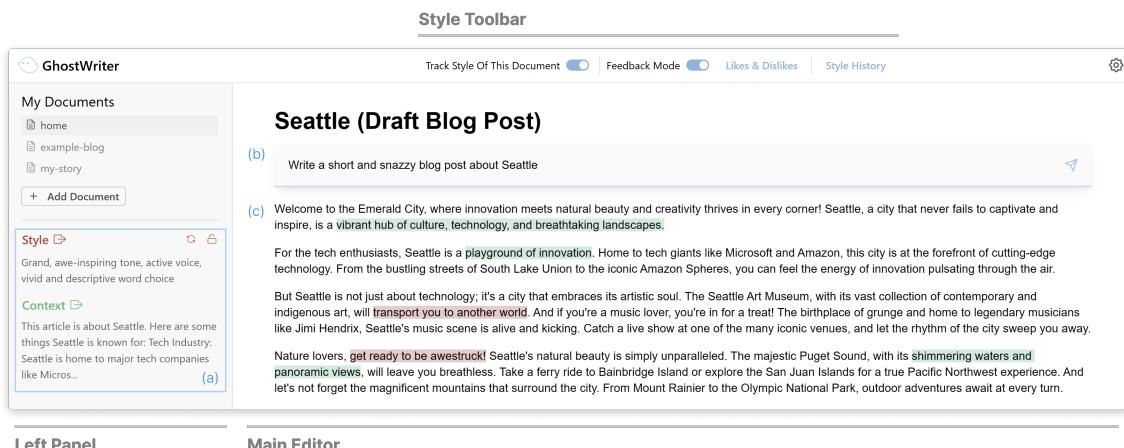


Fig. 1. Overview of GhostWriter, an AI-powered environment that personalizes the writing process through (a) editable *style* and *context* information. (b) Personalized content can then be generated using features such as inline LLM prompts. (c) Users can also explicitly teach the system about their style preferences by highlighting likes & dislikes.

Large language models (LLMs) are becoming more prevalent and have found a ubiquitous use in providing different forms of writing assistance. However, LLM-powered writing systems can frustrate users due to their limited personalization and control, which can be exacerbated when users lack experience with prompt engineering. We see design as one way to address these challenges and introduce GhostWriter, an AI-enhanced writing design probe where users can exercise enhanced agency and personalization. GhostWriter leverages LLMs to learn the user’s intended writing style implicitly as they write, while allowing explicit teaching moments through manual style edits and annotations. We study 18 participants who use GhostWriter on two different writing tasks, observing that it helps users craft personalized text generations and empowers them by providing multiple ways to control the system’s writing style. From this study, we present insights regarding people’s relationship with AI-assisted writing and offer design recommendations for future work.

CCS Concepts: • Human-centered computing → Interactive systems and tools; Natural language interfaces; HCI design and evaluation methods; • Computing methodologies → Artificial intelligence.

Additional Key Words and Phrases: AI-assisted writing, large language models, generative AI, co-creation, personalization, design probe

Authors’ addresses: Catherine Yeh, catherineyeh@g.harvard.edu, Harvard University, Cambridge, Massachusetts, USA; Gonzalo Ramos, Microsoft Research, Redmond, Washington, USA, goramos@microsoft.com; Rachel Ng, Microsoft, Redmond, Washington, USA, rng@microsoft.com; Andy Huntington, Microsoft, London, UK, anhuntn@microsoft.com; Richard Banks, Microsoft Research, Cambridge, UK, rbanks@microsoft.com.

1 INTRODUCTION

With the rapid advances and increasing ubiquity of large language models (LLMs), there is growing interest in exploring their impressive text generation capabilities [6], particularly in the context of writing. Many systems have emerged that leverage LLMs to assist with various writing tasks, from brainstorming ideas and drafting content [25, 38, 56] to summarizing documents [11, 57, 58] and refining existing text [46, 56].

For example, Sarah, a software engineer, may want to use artificial intelligence (AI) to help write a short introduction email to her new team at work. She might ask ChatGPT¹, a popular LLM-powered chatbot, for assistance and receive the following as output from the default GPT-3.5 model:

Subject: Introduction and Excitement for Joining the Team!

Dear Team,

I hope this email finds you all in good spirits. My name is Sarah, and I am thrilled to introduce myself as the newest member of the team. I am joining as a software engineer, and I couldn't be more excited to embark on this journey with all of you...

This example illustrates two core challenges that can arise when using LLMs for AI-assisted writing, which internal teams in our organization have consistently observed. First, the output can be too general, reflecting a *lack of personalization* in the generated text [9, 16, 28, 38, 56]. The LLM likely did not write in Sarah's own writing style, making the output feel, to those who know Sarah and Sarah herself, as coming from somebody else and overly formal. Additionally, the LLM may be missing relevant context about Sarah or the situation that may be helpful to know when writing this introduction email (e.g., her knowledge about a topic, what project she will be working on with the new team, etc.).

Second, the complexity of prompting and prompt engineering as a practice can make people feel limited control or *agency*² when working with LLMs [9, 10, 22, 36, 56]. Following the example above, when Sarah identifies something about the output that she wants to change, such as the writing style, it can be very difficult and time-consuming for her to make those changes, as this is often only achievable by end-users through iterative prompting where one understands what needs to change and why [56]. Furthermore, due to the stochasticity of LLM outputs, users like Sarah may feel like they lack control over how their changes will be reflected in the generated text and what parts of the prompt are influencing the model's behavior [32, 35].

The challenges of personalization and control in the use of LLMs are barriers limiting the benefits that AI writing systems powered by this technology can offer end-users. We see design and technology probes [5, 15, 17, 21] as key methods for addressing the aforementioned challenges, unlocking the use of LLMs as amplifications of human capabilities and as lenses through which we can understand how people interact with LLMs across different domains.

These challenges motivate us to work on the following research questions:

- **RQ1:** How can we improve and use personalization to increase the *alignment* between user intention and LLM-based machine writing?
- **RQ2:** How can we preserve and champion user agency in AI-powered writing interfaces?
- **RQ3:** How does increasing personalization and agency in intelligent writing environments impact user behavior and outcomes?

¹<https://chat.openai.com/>

²Interpreted as the action of steering a system behavior toward a desired outcome with confidence.

We address these questions through the design and development of GhostWriter, an AI-powered editor that allows users to personalize the LLM writing experience by providing agency in style and context specification (Figure 1). Users can teach the system about their target style by writing text, directly editing the style description, or highlighting parts of a document with “likes” and “dislikes.” GhostWriter also allows users to define and refine context information. By offering agency over a target style and context, users provide the underlying LLMs with information leading to personalized text generations.

GhostWriter is an artifact that explores the possibility of using AI to craft a personalized writing experience while preserving user agency and ownership in the creative process. We used GhostWriter as a design probe in a user study with 18 participants. This study helped us learn how participants leverage GhostWriter’s capacity as a personalized writing environment in two distinct writing tasks. Our results reveal that GhostWriter can help users exert control over steering LLM outputs and offers value in its flexibility in customizing style and context. Based on these findings, we share insights to guide future human-AI collaborative writing experiences.

The key contributions of our work include:

- GhostWriter, an AI-enhanced writing environment that offers content personalization through style and context.
- User study results that reveal the utility of GhostWriter in generating personalized text generations and championing user agency.
- A set of design recommendations and directions for future research on LLM-assisted writing.

2 RELATED WORK

2.1 AI-Assisted Writing

Much work investigates how AI can enhance the human writing process [9, 11, 13, 16, 19, 25, 28, 54, 55]. For example, [10] explores how a machine-in-the-loop system can amplify creativity for short story and slogan writing tasks, and [51] presents a multimodal interface for creative writing powered by generative AI. Wordcraft [22, 56] is an LLM-powered editor for human-AI co-writing, offering features including rephrasing/continuing a text passage and just-in-time custom controls. Dramatron [38] also uses LLMs for creative text generation, targeting the domain-specific application of co-writing screenplays and theater scripts. However, a common theme in these systems is the felt lack of context-awareness and sense of autonomy when generating text, which we aim to address. Additionally, GhostWriter serves as a general-purpose editor, not constrained to creative writing domains.

Many products have integrated AI features into writing interfaces, for example, Gmail Smart Compose [7], Grammarly³, and Wordtune [59]. Last year, Tiptap introduced an extension with various AI-powered commands⁴, including tone adjustment, rephrase, and extend. While we build on this framework and implement similar capabilities, GhostWriter focuses specifically on personalization through writing style and context. Notion⁵ is a connected workspace with a landmark notebook component that has recently incorporated inline AI writing features in ways similar to how we envision AI writing assistance. Our work, developed in parallel, relates to this design pattern and takes it in a different direction focused on style capture and generation, while sharing results from users performing two types of writing tasks.

³<https://www.grammarly.com/ai-writing-tools>

⁴<https://tiptap.dev/blog/tiptap-ai>. This extension was announced on July 30, 2023, after our system implementation was already complete.

⁵<https://notion.so>

2.2 Personalized AI Experiences

Personalization has been extensively studied for recommender systems [3, 34, 39, 45], and we take inspiration from the idea of “natural language user profiles” [45] in our designs. To our knowledge, we have not yet seen this idea applied to the domain of AI-assisted writing, so our work explores the possibility of creating editable style and context profiles to guide the generation of personalized text. As with recommender systems, we propose that these natural language profiles can provide increased transparency and scrutability to users about AI-generated outputs. Impressona [4] explores the related idea of creating personas with different writing styles; however, we focus on personalizing the written *content* itself, rather than generating stylized *feedback* for writers.

With the rise of LLMs, various efforts have aimed to personalize downstream user experiences without retraining these neural models. FoundWright [42] and ForSense [44] use transformer models to personalize online research and information re-finding activities while retaining user agency in these machine-mediated processes. Our overarching goal is similar, and we also do not retrain any models, but we explore the themes of personalization and agency in the context of writing. Much LLM-based research focuses on personalizing text generation in particular, e.g., [49, 60]. Some work proposes incorporating social factor modeling to personalize the writing experience [29], while others draw on writing education [33] or suggest using methods like eye tracking to complement users’ cognitive states and situational awareness during writing [30]. LMCanvas [27] advocates transforming traditional text editors into canvas-based interfaces, where interactive “blocks” can be manipulated to create a personalized writing environment. We take a complementary approach with GhostWriter by studying personalization through the lens of writing style and context.

Our work also builds on ideas from interactive machine teaching (IMT) [47, 50], the paradigm in which a human teacher communicates information to the machine learner in an iterative, human-in-the-loop style process [1]. GhostWriter is inspired by IMT’s overarching objective to empower non-experts to participate in iteratively “teaching” an algorithmic “learner” about their target writing style and context, information that in turn is used to drive (personalized) text generations.

2.3 Working with Style and Context

A core concept of GhostWriter is extracting a specific writing style and applying it to generate or refine text. To extract style, we use LLM prompting, a human-interpretable alternative to neural methods such as style representation learning [43]. Another relevant idea from natural language processing is text style transfer (TST) [14, 20, 23], which aims to preserve the content of generated text while adjusting style attributes like tone or voice. However, while TSTs are limited to preexisting text, our system can apply existing styles to produce new text, similar to [33]. A related task is controllable text generation, where various aspects of generated text can be manipulated, such as context [52] or topic [12]. With GhostWriter, we explore how natural language can be used as a means for end-users to “perform” personalized text style transfers and controllable text generation with LLMs. Our work also draws from approaches such as context-faithful prompting [61] when incorporating the user’s writing style and context.

OpenAI recently released a custom instructions feature⁶ for ChatGPT, allowing users to specify information and instructions that the system should consider when generating output. We also allow users to tweak context and style in GhostWriter through natural language, but our system uses this information to generate personalized *writing* rather

⁶<https://openai.com/blog/custom-instructions-for-chatgpt>. Announced on July 20, 2023: also post-GhostWriter implementation.

than *chat output*. Additionally, GhostWriter’s context page provides information about document content, rather than the user themselves.

3 SYSTEM OVERVIEW

We address the aforementioned key challenges of *personalization* and *agency* in the context of AI-augmented writing with the design of GhostWriter, an LLM-powered writing environment. We use GhostWriter as a design probe [5, 17, 21] to explore the possibility of using AI to craft a personalized experience through writing *style* and *context* while preserving user agency. Design probes in human-computer interaction research are adapted from cultural probes [15], designed objects aiming to promote participant engagement in the design process [5]. We chose to create a design probe because using them has been shown to be effective in gaining insights from end-users on open-ended tasks during design ideation [18, 24, 42, 44]. With design probes, the focus is on identifying potential areas of value and improvement for novel ideas, rather than evaluating usability or comparing the probe with existing solutions [42]. As a probe, GhostWriter helps to elicit feedback on our design ideas aimed at personalizing the writing experience through giving users agency over style and context, and inspire future work on personalization in AI-assisted writing systems.

3.1 Design Principles

The design, development, and study of GhostWriter is guided by our research questions (section 1). These questions led us to ideate how we could improve agency and personalization in an LLM-powered writing experience. Some ideas included adding both progressive (i.e., as you write) and explicit (i.e., user-defined) forms of personalization, offering support for finer-grained refinements of LLM output, and allowing users to iteratively define the system’s context and writing style. Building on these ideas, we defined the following design principles (DPs):

- **DP1: Leverage Machine Capabilities While Championing Agency and Control.** We wish to explore opportunities for leveraging LLM capabilities in augmenting the writing process. Our goal is to use their text generation and semantic analysis abilities [6, 38, 56] to extract writing styles and produce customized content. Our work also aims to preserve and enhance user agency when co-writing with AI. We believe this will augment people’s writing capabilities and create a more productive, personalized writing experience that addresses concerns from previous work [9, 22, 56].
- **DP2: Use Familiar Editor Metaphors.** To reduce the cognitive load [53] of navigating a new system and adhere to the *consistency and standards* usability heuristic [41], we strive to utilize existing text editor metaphors in our design. By exploring how AI augmentations blend into familiar writing experiences, we can focus on distilling the effects of such augmentations and project our learnings into grounded, relatable scenarios.
- **DP3: Blend Into the Writer’s Existing Workflow.** We do not want users to have to work in unnatural or contradictory ways to their typical writing routine when using our system. We aim to prioritize simplicity and a non-fragmented user experience, similar to Notion or Microsoft Loop⁷, which offer integrated inline interfaces for writing. In this way, all writing and system interaction takes place in one central editor panel, reducing interface complexity, divided attention, and unnecessary context-switching [37].
- **DP4: Provide Transparency to Support Reflection and Discovery.** Our system strives to offer transparency into its internal state. We want users to know and be able to inspect what information about their writing style and context is given to the LLM to generate new content, particularly by leveraging the expressiveness and

⁷<https://www.microsoft.com/en-us/microsoft-loop>

Table 1. LLM-powered features for personalized text generation.

Feature	Description	Location
rewrite	rewrite text selection to match the system’s learned <i>style</i>	context menu
contextual prompt	apply LLM prompt to text selection to generate new content using the system’s learned <i>style</i>	context menu
continue	generate new text to continue the current document using the system’s learned <i>style</i> and <i>context</i>	slash menu
inline prompt	generate new content based on general LLM prompt using the system’s learned <i>style</i> and <i>context</i>	slash menu

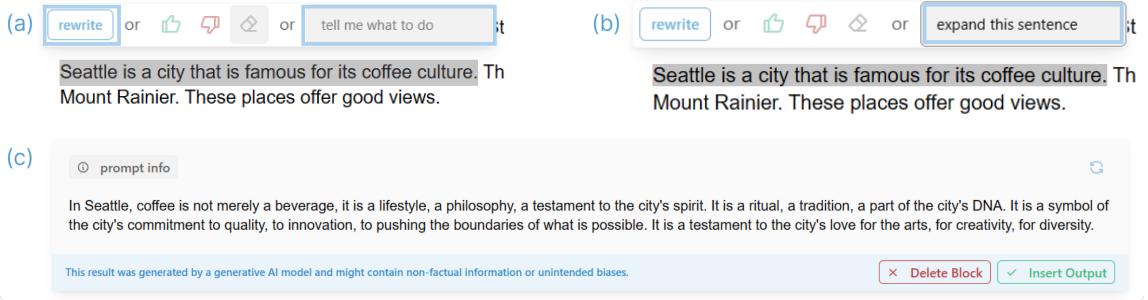


Fig. 2. Personalization through refining text. (a) Upon invoking the context menu, the user can choose the *rewrite* or the *contextual (“tell me what to do”) prompt* option. (b) The contextual prompt will apply the inputted text as a prompt to the current text selection. (c) Sample output from using a contextual prompt. Users can regenerate, delete, or insert the outputted text.

accessibility of natural language [45]. We also hope that transparentizing the system state encourages reflection and exploration during writing. For instance, perhaps the system can teach the user something new about their writing style, or allow experimentation with alternative styles and contexts.

3.2 Interface Design

Below, we outline the key components and features of GhostWriter’s interface, as informed by our DPs.

3.2.1 Main Editor. The central view of our interface is the *main editor* panel (Figure 1), which mirrors existing text editors [DP2] and provides a space for users to author documents. One way users can teach GhostWriter about their target writing style is simply by **writing text**. After each n (default: 100) new characters are written, the system will analyze the current document to extract its style [DP1].

In the main editor view, users can use LLM-powered features to generate customized text given the current writing style and context [DP1]. There are 4 options to craft personalized content (Table 1), which involve asking GhostWriter to *refine* existing text or *generate* new text. All features are embedded inline to support a non-fragmented writing experience and embody familiar workflows [DP2, DP3].

To refine existing text, users can invoke the context menu by selecting a portion of text in the document. Then, they can **rewrite** the selected text or use the **contextual “tell me what to do” prompt** to apply a prompt to the current

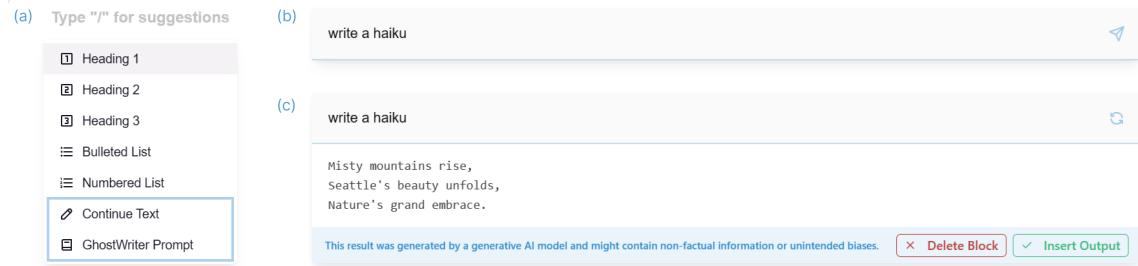


Fig. 3. Personalization through generating new text. (a) Upon invoking the slash menu, the user can choose the *continue text* or *inline (“GhostWriter”) prompt* option. (b) The inline prompt takes any general prompt as input. (c) Sample output from an inline prompt. Users can regenerate, delete, or insert the outputted text.



Fig. 4. Teaching style through likes and dislikes. (a) The user can like (thumbs up) or dislike (thumbs down) any text selection through the invoked context menu. (b) Once the corresponding icon is selected, the user can optionally provide feedback as to why they like or dislike the highlighted text.

selection (Figure 2a). Both operations apply the current style to generate personalized content (Figure 2b). Users can regenerate, delete, or insert LLM output (Figure 2c) [DP1].

To generate new text, users can invoke the slash menu by typing forward slash (“/”) anywhere in the document (Figure 3a). Then, they can **continue** the text from the current point or invoke the **inline “GhostWriter” prompt**. Both operations use the current style and context to generate personalized content (Figure 3b). Similar to Figure 2, users can regenerate, delete, or insert LLM output (Figure 3c) [DP1].

The user can also explicitly teach GhostWriter about their target style by **indicating likes and dislikes** [DP1]. To do so, they can highlight a portion of text in the document [DP2], which will invoke the context menu (Figure 4a). Then, they can press the like or dislike icon and optionally explain why they like or dislike the highlighted text to help the system learn more about their style preferences (Figure 4b).

3.2.2 Left Panel. On the left sidebar (Figure 1), users can (1) view their *document* list and (2) explore the system’s current *style and context*. (1) allows users to easily create new documents and switch between them, mirroring existing editor interfaces [DP2, DP3]. (2) offers a summarized view of the style and context information GhostWriter is using to generate personalized text [DP4]. The style summary is automatically refreshed when the system’s writing style updates (Figure 5a). Style updates can be manually triggered by pressing the refresh icon, e.g., to analyze the writing style of an existing document (Figure 5b) [DP1]. Users can disable all automatic style updates by pressing the lock icon.

To avoid the cold start problem, GhostWriter begins with a default writing style. The system is prompted to analyze at minimum the following five style characteristics: *tone*, *voice*, *word choice*, *sentence structure*, and *paragraph structure*. We chose these parameters to create a user-friendly language for communicating style, based on informal pilot studies and experimentation where we asked LLMs to characterize the style of a written document.



Fig. 5. GhostWriter automatically analyzes one's style as one writes. (a) When the style is updated, the new summary will appear on the left sidebar. (b) One can manually trigger a style update by pressing the refresh icon, or disable all automatic style updates by pressing the lock icon.

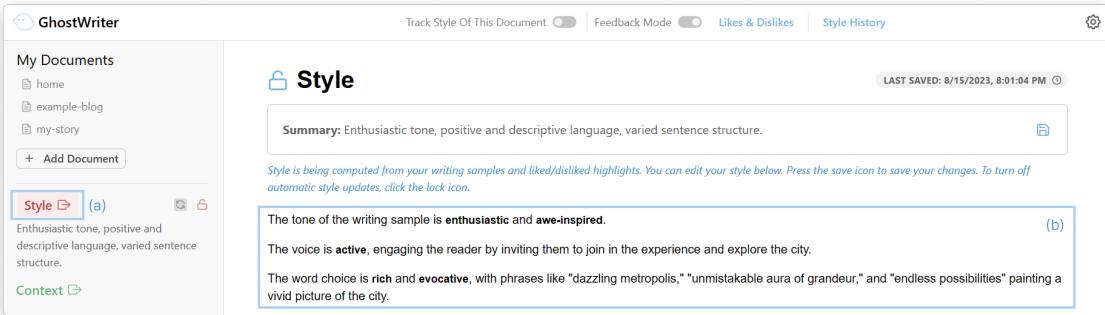


Fig. 6. Teaching style through manual edits. (a) The user can view and edit the full description of their current description by pressing the Style button. (b) The style description (excerpt shown) can be edited like a normal text file.

In addition to writing text, the user can **manually edit** the system's style – another way GhostWriter champions user agency [DP1] in defining a personalized style. This is achieved by pressing the **Style** button in the left sidebar [Figure 6a], which opens the full description of the system's learned style in the main editor [DP3, DP4]. This page can be edited like a normal text document and is not limited to the five default style parameters [Figure 6b]. In allowing users to inspect and modify the system's style, GhostWriter supports reflection and experimentation throughout the writing process [DP4].

Users can optionally provide additional context to ground the generation of text [DP1]. To do this, users can edit the **Context** page [Figure 7a] [DP4]. Like the style description, this document opens in the main editor [DP3] and can be edited like a normal text document; there are no constraints on what can be included as "context" [Figure 7]. Only the user has control over content in the context page.

Our implementation only stores one global style and context. In future system iterations, we want to add the capability of storing and selecting from multiple styles and contexts that could be differentially applied to documents. Additionally, simplifying style to five dimensions may not fully capture all aspects of the desired writing style. However, users are not limited to these initial characteristics, as we designed GhostWriter to allow direct editing of the system's style. This

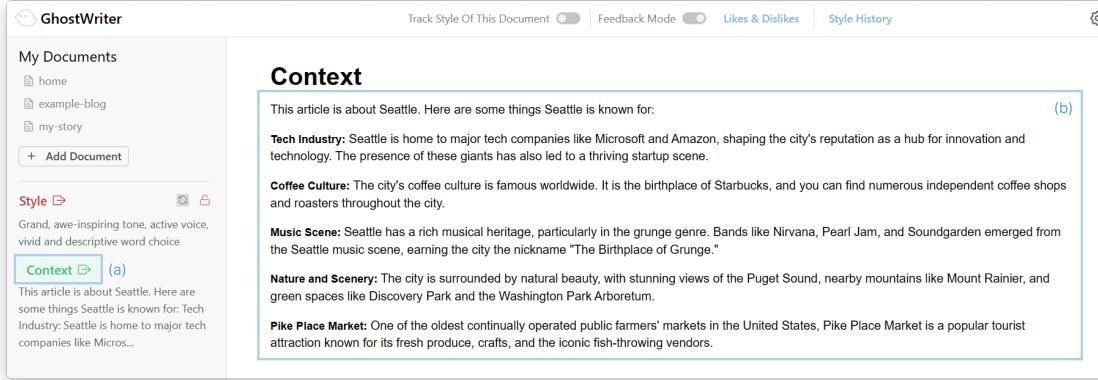


Fig. 7. (a) Users can view and edit the system’s contextual information by pressing the Context button. (b) Context can be edited like a normal text file.

permits the articulation of bespoke dimensions through natural language or by connecting aspects to explicit likes and dislikes. Other forms of style expression could also be adapted (see Section 6).

3.2.3 Style Toolbar. Above the main editor, there is a *style toolbar* (Figure 1), where the user can customize their GhostWriter experience [DP1] and further inspect the system’s style knowledge [DP4]. Users can toggle the **Track Style Of This Document** flag, which turns on/off automatic style updates triggered by the current document (vs. the *global* style lock in Figure 5b). They can also turn **Feedback Mode** on/off, which shows or hides all highlighted likes and dislikes in the current document.

Besides viewing the current style description on the Style page (Figure 6), another option for examining writing style is looking at the **Style History** page (Figure 8a), which displays a history of the system’s style changes in the main editor [DP3]. More recent styles are displayed at the top of the page. *Styles* are displayed in blue boxes on the left (Figure 8b), while *comparisons* between each pair of adjacent styles are shown in gray boxes on the right (Figure 8c). These LLM-generated comparisons [DP1] offer deeper insight into how the system’s style changed over time by providing a “difference rating” (Figure 8d) to quantify the difference between adjacent styles from 0 to 10 (0: identical, 10: entirely different) [DP4]. Users can revert to a prior style by clicking the *revert* icon in the corresponding blue style box (Figure 8e) [DP1].

Similarly, users can view their full collection of **Likes & Dislikes** in the main editor (Figure 9a) [DP3]. At the top of this page, there are LLM-generated summaries of the user’s style preferences based on all text selections they liked and disliked (Figure 9b) [DP1, DP4]. This interpretation by the system can encourage additional reflection [DP4] and help users assess whether GhostWriter is correctly understanding their highlights. Users can add additional likes and dislikes (Figure 9c) or toggle/delete highlights in this view (Figure 9d) [DP1]. Only active highlights are used to compute the like/dislike summaries, which are factored into the writing style used by the system for text generation. User feedback can also be edited on this page.

3.3 Implementation

As a system, GhostWriter consists of three main components:

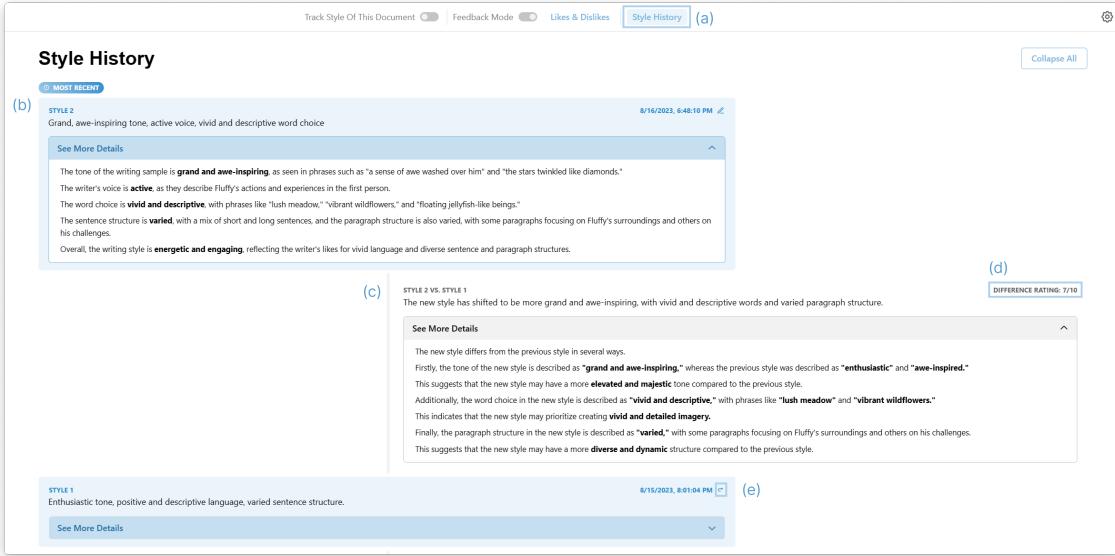


Fig. 8. (a) Users can view all their past writing styles on the Style History page. (b) The blue boxes on the left each display a writing style and (c) the gray boxes on the right display a comparison of each pair of adjacent writing styles. (d) Each comparison also includes the difference rating between adjacent styles. (e) Users can revert to prior writing styles as well.

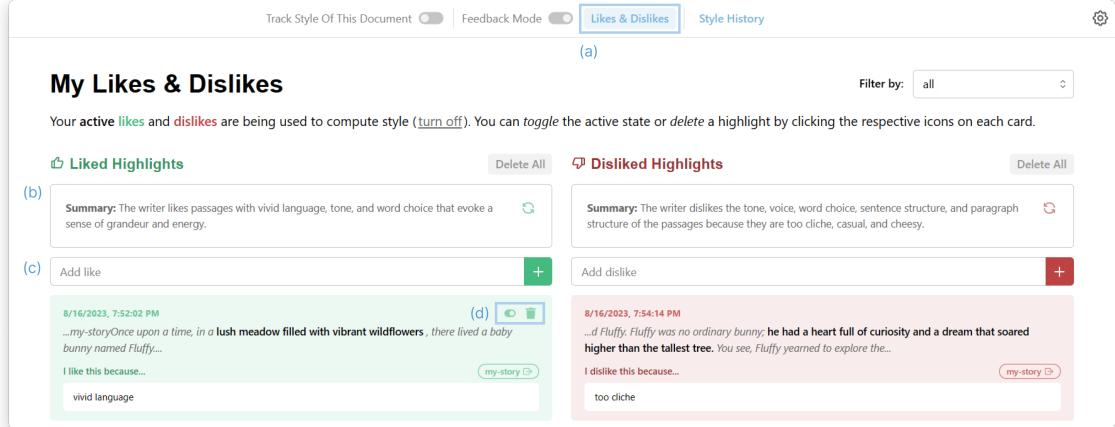


Fig. 9. (a) Users can view all their likes and dislikes on the Likes & Dislikes page. (b) A system-generated summary of the user's likes and dislikes is displayed at the top of the page. (c) The user can also manually add additional likes and dislikes. (d) Upon hovering on a like or dislike card, users have the option of toggling its active state or deleting it from their collection.

Frontend. GhostWriter's frontend is a web application written in React and TypeScript. To create the main editor interface, we build on Tiptap⁸, a headless editor framework.

Backend. The frontend communicates with a Python backend that provides AI services through a RESTful endpoint.

⁸<https://tiptap.dev/>

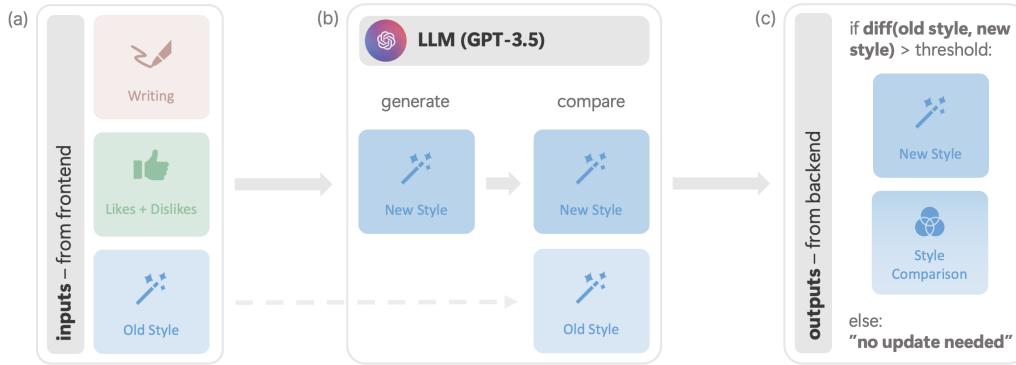


Fig. 10. How style updates are computed by GhostWriter. (a) The current document, likes & dislikes, and style description are passed as inputs from the frontend. (b) In the backend, we then ask the LLM to generate a new style description given this information. The LLM also generates a style comparison between the old and new styles, and computes a difference rating. (c) If the difference rating is greater than some threshold (e.g., 3 out of 10), the new style and comparison are passed as outputs back to the frontend. Otherwise, the user will be informed that there is no style update needed.

AI Services. All LLM operations in the backend are orchestrated with the LangChain⁹ framework. Most operations use OpenAI’s Chat GPT-4 model, but for computing style updates, we use Chat GPT-3.5-Turbo to reduce system latency. GPT-4 and 3.5 also produced subjectively similar results in this case during experimentation.

We crafted and refined system prompts through an iterative process. Given our goal of producing a design probe, we did not optimize prompts to achieve perfect outcomes. Instead, we informally used OpenAI’s GPT-4 to evaluate generated outputs as a quick sanity check to supplement qualitative inspection by the researchers. The model generally assigned high confidence scores (e.g., ≥ 8 out of 10) to LLM-generated style descriptions and comparisons/difference ratings, increasing confidence in the relative soundness of our prompts. People looking to deploy such systems in production should rigorously ensure the quality of output meets the specific needs of their use cases [26].

All prompt templates are included in Appendix A, and most are relatively straightforward. However, computing **style updates** requires additional logic (Figure 10). When a style update is requested (either by the user or the system), we pass the full *current style* description, user’s *likes & dislikes*, and *text content* of the current document as inputs to the backend. We then ask the LLM to produce a new style description based on the provided text selection and likes & dislikes. The LLM is also prompted to format the description as HTML (Figure 6) and produce a short style summary (Figure 5). Next, we ask the LLM to compare the new and old styles and produce a written analysis of how these styles differ, including a difference rating (Figure 8). This rating is used to determine whether a style update is necessary (i.e., the new style is sufficiently different from the old one). If the rating is greater than a threshold (currently, a value of 3 out of 10 that we settled on through pilot studies), we return the *new style* and *comparison* as outputs to the frontend. Otherwise, the user sees a “no style update needed” message.

4 USER STUDY

To evaluate our system and address our research questions, we conducted a two-part user study with 18 participants focusing on GhostWriter’s abilities to aid in text *editing* and *generation* tasks. These tasks were informed by background

⁹<https://langchain.com/>

research and discussions with collaborators about common user scenarios for AI-assisted writing systems, and designed to allow participants to explore different system features.

4.1 Participants

The participants were **18 employees** (16 female, 2 male) from a large technology company, recruited by mailing list. We targeted participants whose professions involve writing in significant ways. Our call led to 7 content designers, 6 UX researchers, 4 communications managers, and 1 executive assistant, all based in the United States.

4.2 Procedure

The study was organized into **two, 1-hour sessions** in which participants interacted with GhostWriter. Each session focused on a distinct writing task, so all participants experienced both tasks. Sessions took place virtually over Microsoft Teams. Participants were compensated with a **\$100** Amazon gift card upon study completion. This study was approved by our company's Institutional Review Board.

Throughout the study, we asked participants to think aloud. We also recorded their screens and used logged system events to gain deeper insight into their interactions with GhostWriter.

4.2.1 Pre-study Survey. Prior to **Session 1**, participants completed a pre-study survey in which we asked them about their experience with generative AI systems and prompting.

15 of the 18 participants reported working with generative AI systems for less than 1 year (1 had never interacted with them). Five participants reported interacting with such systems multiple times a week; another five reported multiple interactions per month, and the remainder reported less frequent interactions. 6 participants indicated that they generally write multi-sentence prompts, with the rest writing simpler prompts. 10 out of 18 participants noted that they typically do not attach contextual documents or references when prompting generative AI systems.

4.2.2 Tutorial & Practice. Upon beginning **Session 1**, participants were asked to watch a 5-minute tutorial video that provided an overview of the GhostWriter system. Afterwards, participants were given 10-15 minutes to orient themselves with GhostWriter and practice using the system's various features.

4.2.3 Task 1: Professional Editing. The main task in **Session 1** was to edit and refine a document to fit a particular writing style. We provided the following instructions to participants:

Imagine you are a freelance content writer working with a new client. The client runs a travel blog, *EpicWanderlust*, and wants you to help write a new post about Seattle. You are given a draft with some basic ideas, but the client feels it does not fit the “style” of their blog. Here is where you come in: Your job is to polish the draft (with GhostWriter’s help!) so that it feels more cohesive and consistent with the other posts on *EpicWanderlust*.

One additional request: the client wants you to help them reach younger audiences by tweaking the post’s writing style.

We included an example blog post from *EpicWanderlust* to help participants get started and extract the desired style. A snippet of this post is shown in [Figure 5](#). Additionally, we provided some default context about Seattle ([Figure 7](#)). Participants were given ~30 minutes to complete this task.

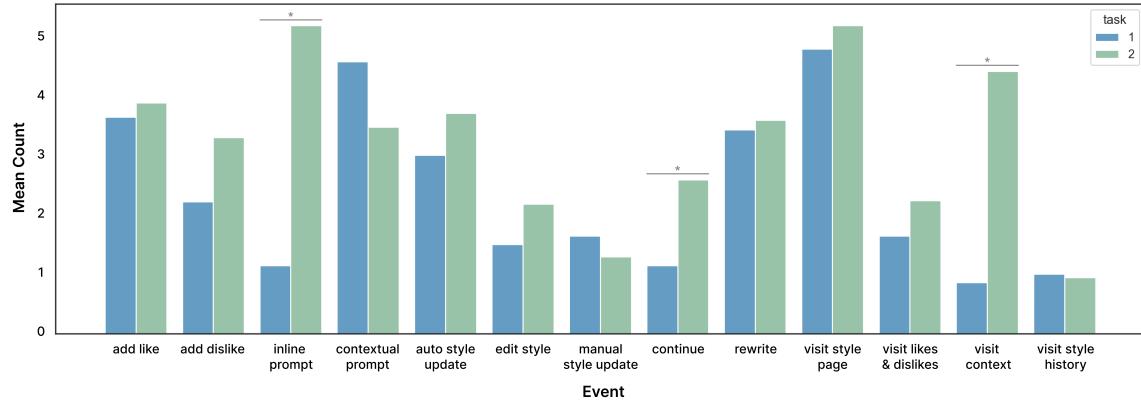


Fig. 11. System usage as illustrated by interaction log event counts. The mean counts for task 1 (editing) are shown in blue, and the mean counts for task 2 (generation) are shown in green. Statistically significant differences are denoted by an asterisk.

4.2.4 Task 2: Creative Writing. The main task in **Session 2** was to write a document following a writing style chosen by the participant. We asked participants to bring a short writing sample containing a style they wanted to emulate (written by them or someone else) and then asked them to write a story based on one of these instructions:

- Story 1:** “Write a short story about an intern at a tech company having an adventure on an alien planet.”
- Story 2:** “Write a short story about a group of friends who find themselves trapped in a haunted mansion.”
- Story 3:** “Write a short story about a musician who finds a magical device that can control the weather.”

Participants chose which prompt to work on, and we provided some default context about a possible setting and characters they could use in each case to get started. Participants were given ~45 minutes to complete this more open-ended, generative task.

4.2.5 Post-task Survey. In both sessions, after finishing **Task 1** or **Task 2**, participants filled out the same, ~10 minute post-task survey. We first included a series of Likert scale questions regarding various aspects of the user experience, including satisfaction, ease of use, perceived agency and output ownership, as well as system trust, responsiveness, and understanding. Then, we asked participants open-ended questions about their overall experience with GhostWriter, likes and dislikes, and, after completing both sessions, how they perceived their relationship with the AI system.

5 RESULTS

In this section, we first present a quantitative analysis of our event logs to assess how participants used and interacted with GhostWriter. We then synthesize themes from the qualitative survey data to collect participant impressions about GhostWriter as an AI-powered writing assistant.

5.1 System Usage

All participants successfully used GhostWriter and its features to complete both the *editing* and *generation* tasks (Figure 11). Participants iteratively crafted a target style through implicit or explicit methods. Their actions led to an

average of **6.71** style updates per task: **3.39** were **automatic**, **1.87** came from **direct edits**, and **1.45** were **manually requested**. During each task, participants inputted additional style feedback by adding an average of **3.77 likes** and **2.81 dislikes**. The **style history** page was viewed an average of **0.97** times and the **likes & dislikes** page an average of **1.97** times in each task. In the *editing* task, participants viewed the main **style** page **4.79** times on average, while in the *generation* task, they viewed it **5.18** times. Participants viewed the **context** page an average of **0.86** times during the *editing* task, compared to **4.41** times in the *generation task* ($t(18) = -6.2627, p < 0.0001$).

On average, participants used the **rewrite** feature **3.52** times per task. The **continue** feature was significantly less used in the *editing* task (**1.14** times) compared to the *generation* task (**2.59** times) ($t(18) = -2.9885, p = 0.0061$). For the *editing* task, participants used an average of **1.14 inline** prompts, while for the *generation* task, they used an average of **5.18** ($t(18) = -6.1252, p < 0.0001$). Conversely, participants used an average of **4.57 contextual** prompts in the *editing* task, compared to **3.47** in the *generation* task.

5.1.1 Patterns By Task Phase. Writing tasks are not monolithic, and we expect people's behaviors (like any story) could change between the first and last parts of the task. As such, we looked at differences in interaction patterns between the first and second half of each task. During the first half of tasks, participants requested **manual** style updates **1.19** times on average, while in the second half, this dropped to **0.29** ($t(36) = 5.8391, p < 0.0001$). The main **style** page was visited an average of **3.32** times in the first half of each task, compared to **1.68** in the second half ($t(36) = 2.8818, p = 0.0072$). Similarly, the **context** page was visited more often in the earlier half of each task (first half mean: **2.10** times vs. second half mean: **0.71** times) ($t(36) = 3.4847, p = 0.0015$). On the other hand, the **likes & dislikes** page was visited less frequently in the first half (mean: **0.58** times) compared to the second half (mean: **1.39** times) ($t(36) = -2.8816, p = 0.0072$).

Comparing task phases across our two tasks yielded additional insights about event frequencies over user study sessions. For instance, during the *editing* task, participants viewed the main **style** page **3.07** times less on average in the second half of the session vs. **0.47** times less for the *generation* task ($t(18) = 2.4286, p = 0.0221$). Conversely, participants visited the **context** page, on average, **0.29** times less during the second half of the *editing* task, compared to **2.29** times less during the *generation* task ($t(18) = -3.0141, p = 0.0069$).

5.1.2 Patterns By Prompt Type. To learn more about user intents while using GhostWriter, we used our log files to analyze LLM prompts written during both tasks. In total, participants composed **97** inline and **134** contextual prompts.

For **inline** prompts, the most common intent was *adding additional content* (**n=42**), e.g., “Elaborate on how John comes across PlayStation while exploring Starfield” or “Add a paragraph with additional places to visit.” Participants also frequently wanted to *generate full drafts* of documents (**n=30**), e.g., “Write a short horror story in the style of Edgar Allan Poe. The story should have a strong plot with a surprise twist” or “Write me a blog article with an introduction, 4 paragraphs with catchy titles and a summary to convince my friend to visit Seattle.” **10** inline prompts by participants asked for a document *introduction*, and **9** asked for a *conclusion*.

With **contextual** prompts, participants commonly asked GhostWriter to *expand* the selected text (**n=53**), e.g., “Add more details about his fellow researchers; reference a second character named Bob.” **43** prompts expressed an intent to *rewrite* the selected text, e.g., by changing the perspective (“Change to first person”), tone (“Make this more positive and enthusiastic”), or audience (“Rewrite for younger readers”). Participants also used contextual prompts to *condense* text selections (**n=6**), e.g., “Shorten this poem by half”, or request specific *formatting* (**n=7**), e.g., “Bullet form this paragraph.” In **3** cases, participants wanted to *add transitions* between paragraphs. **3** contextual prompts were used to ask for *suggestions or critique*, e.g., “Is this a well-written sentence?”

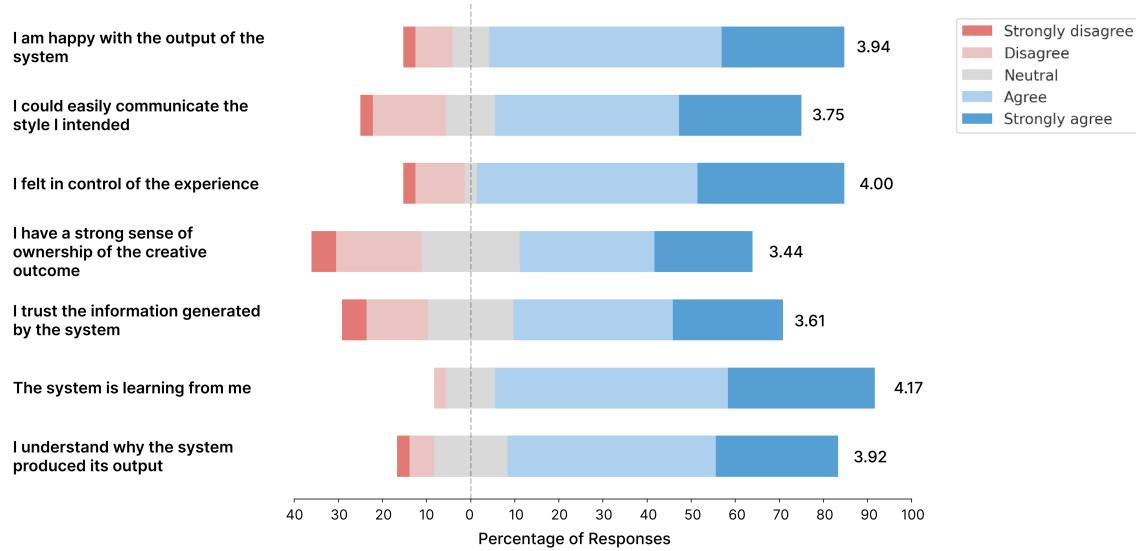


Fig. 12. Aggregated participant responses to survey questions regarding system usability and satisfaction after completing each user study task with GhostWriter. Each question was scored on a 5-point Likert scale (1: *strongly disagree* - 5: *strongly agree*). Each horizontal bar contains 36 responses (2 per participant), with average values displayed on the right.

5.2 Qualitative Results

Participants gave positive responses about their experience with GhostWriter (Figure 12). The only significant difference observed between tasks was for “I trust the information generated by the system,” where the mean trust rating increased after task 2 from 3.39 to 3.83 ($t = -2.406, p = 0.028$)¹⁰. This statement, along with “I have a strong sense of ownership of the creative outcome,” yielded the most variation and lowest mean participant ratings (mean trust: 3.61, mean ownership: 3.44). In each case, however, 61.1% and 52.7% of responses were still “agree” or “strongly agree,” respectively.

The statements “I felt in control of the experience” and “The system is learning from me” received the highest mean scores from participants for both tasks (mean control: 4.00, mean learning: 4.17), reflecting positively on our goals of personalization and agency. 83.3% of responses were “agree” or “strongly agree” for the former statement, and 86.0% showed agreement with the latter. Overall, survey responses indicate that most participants were able to communicate their intended style (mean rating: 3.75), felt satisfied with GhostWriter’s customized text generations (mean rating: 3.94), and understood the system’s behavior (mean rating: 3.92).

5.2.1 Overall Impressions. When asked about general impressions of their experience, participants provided responses consistent with their Likert ratings. Participants were impressed and excited by the *ease of style learning and application* with GhostWriter (**n=18**). **P06** said, “I was really impressed with how the system understood and incorporated my style, both from the sample text and [my own] writing.” **P12** mentioned our system was “helpful for real time feedback on the style of writing,” and **P15** noted that the “style extraction [feature] is new and exciting!”

Our system was also viewed as *easy to use and intuitive* as a whole (**n=15**). **P04** described GhostWriter as “fun and robust,” and **P14** reported, “This was an incredibly positive experience. It was easy to build and iterate on the story. The

¹⁰As our user study targets two inherently different writing tasks, and learning effects may have occurred between the two sessions, we did not find it particularly meaningful or fair to compare ratings across tasks.

tool has an easy learning curve, and a simple interface.” Participants enjoyed having different ways to *interact with style and give feedback* to the system (**n=9**), with 7 responses specifically mentioning the *likes & dislikes* feature. The *context page* (**n=7**) and *contextual prompt* (**n=4**) were also common favorite features. In particular, **P08** said the former “helped organize my thinking,” and **P10** thought that “context adds value into the prompt.”

8 responses described the potential of GhostWriter to *augment human creativity and productivity*. **P17** viewed our system was “extremely interesting, I would definitely use this kind of tool to help me be more productive in a work setting.” **P02** added that “for non-writers, I think the system offers a good starting point for any kind of writing,” and **P07** loved “seeing how creative it can be and how it boosted my own imagination.”

Participants also shared ideas for areas of improvement. The most frequent request was having *more fine-grained options* for interacting with LLM outputs (**n=11**). For example, **P06** said, “I wish I could have accepted only parts of the text it generated [and] provided feedback so that it could learn what I liked and didn’t like about it.” Participants also wanted GhostWriter to *store multiple styles and offer preset style templates* (**n=8**).

6 responses mentioned the desire for *alternative forms of style expression*. **P04** wanted to “make style more structured: voice, tone, audience, etc.” while **P08** and **P10** proposed offering “suggestions on styles (drop-downs of various options to spark individual creativity).” Another common issue raised by participants was the occasional *lack of contextual awareness* from GhostWriter (**n=5**). **P02** noted how “the system was unable to continue writing content from the previous sentence. It just wrote more [unrelated] content,” and similarly, **P08** hoped that inline generations could start “at a midpoint versus always having to restart.” **P03** also wished “the system would NOT take the context literally but use it to create a story.”

Other requests included having *side-by-side comparisons and version control* for documents and prompts (**n=5**), allowing *multimodal outputs* (**n=3**), and adding *more specific prompts* to rewrite and continue (**n=3**). 5 participants suggested incorporating a way to *validate style accuracy* as well. As **P15** expressed, it may be helpful if GhostWriter had “some kind of a score - how close are you with the style intended.”

6 DISCUSSION AND DESIGN RECOMMENDATIONS

We use our findings to address our research questions and present a series of design recommendations for practitioners creating similar solutions. We also share insights on how people view the role of AI in the context of AI-assisted writing. Session quotes are included to enhance the discussion.

6.1 RQ1: How can we improve and use personalization to increase the alignment between user intention and LLM-based machine writing?

Our observations and survey responses showed that participants were effective in using GhostWriter’s style extraction feature and building on it to achieve their desired writing style. Participants enjoyed having various ways to control the system’s writing style and, in general, felt like GhostWriter was learning from them (Figure 12). **P09** appreciated GhostWriter’s “responsiveness to the style guide” and **P18** was “pleasantly surprised at how much content I was able to generate in a style I preferred.” Many participants thought GhostWriter could help boost their productivity and creativity in real world settings.

Providing multiple personalization paths is valuable. One of GhostWriter’s perceived main strengths was having many ways to personalize the system’s style “that are optimal for different contexts and objectives” (**P09**). Because users may differ in their writing workflows and preferences, it makes sense to offer flexibility in how they can interact with AI writing systems such as GhostWriter by including options to teach style **implicitly** and **explicitly**. This

flexibility is also beneficial, as users' preferred mode of controlling style may change throughout the writing process. Interaction logs showed that style update requests and manual style edits generally decreased between the first and second halves of both tasks, suggesting these forms of style tweaking may be more useful in earlier writing phases. However, the likes & dislike feature continued to be used by participants throughout each task, indicating this form of explicit personalization was valuable regardless of writing phase – perhaps due to the lightweight aspect of the interaction, and the observed persistent need to make small adjustments to the style the system has captured so far.

Explicit teaching moments can be worth the effort. Participants regularly engaged with the **likes & dislikes** feature (Figure 11), which was unexpected as adding annotations to the document can add unnecessary effort to the writing process. However, most participants valued being able to capture likes & dislikes, and several noted that this was their favorite feature. **P08** said, “the likes and dislikes [were] extremely easy to use and effective” for tailoring LLM outputs. Participants also frequently filled out the optional feedback field, i.e., specifying why they like or dislike the selected text. Rather than feeling burdensome, participants enjoyed the opportunity to provide feedback to the system, illustrating the value of leveraging explicit teaching moments in AI writing systems. Participants even requested additional opportunities to give GhostWriter feedback throughout the writing experience, by “provid[ing] feedback directly on the text it generated” (**P08**) and “at all input levels” (**P06**). This finding also aligns with guideline 15, “Encourage granular feedback,” from [2] and appears to be key toward the goal of system personalization. Lastly, explicit feedback gives participants the opportunity to use their own insights to identify and underscore personal style dimensions that the system’s style extraction process might miss, as we observed in our user studies.

Having one style is good. Having many is better. Several participants noted that the system could be even more powerful and useful if it allowed them to define and opportunistically select **different styles and contexts**. Going from having one style and context to many is a trend in the emerging capabilities of independent agents like OpenAI’s GPTs¹¹. Having the ability to invite these different styles and contexts into the same document is the next natural step to satisfy our users. Defining multiple styles maps well to cases where users may complete different writing tasks and craft diverse documents. Here, it could be helpful to “apply context/style to individual documents rather than overall” (**P05**) or “allow different styles for different parts of the” document (**P08**). **P10** also suggested having a style “library” to facilitate applying different styles, similar to document templates in Microsoft Word or Google Docs.

Substance and style are important. So is format. Many participants viewed document **formatting as an integral part of style**, wishing to use GhostWriter’s LLM-powered features to personalize and affect the format of generated text. For example, several participants used contextual prompts to reformat text selections and wanted additional formatting support (e.g., bullet points, section headings, etc.). This was a highly requested feature during the *editing* task, as participants wanted to mimic the bold paragraph labels from the sample article (Figure 5). **P08** noted that “easier formatting options” would be helpful for generating different document “templates (Press Releases, Blogs, To Do Lists, emails)” that require different styles. Although we instruct the LLM to analyze sentence and paragraph structure when updating style, the system does not incorporate other aspects of formatting. This is a tractable problem in the domain of rich text or markdown documents as formatted text can be reduced to code. Formatting was also used as a non-literal part of the (teaching) language participants used when explicitly editing the system’s style (Figure 6b). We frequently saw participants bolding certain words, so as to underscore a term’s importance. This type of weight specification is consistent with the observation by [40] regarding teaching languages.

¹¹<https://openai.com/blog/introducing-gpts>

6.2 RQ2: How can we preserve and champion user agency in AI-powered writing interfaces?

Preserving user agency was top of mind in our GhostWriter design, being a crucial point in creating an effective personalized writing experience.

Natural language as an interface can enhance agency. It can also add effort. We chose to promote user agency by allowing style expression through natural language. Our initial hypothesis was that freeform style specification would provide users with empowering expressive flexibility. However, some participants mentioned this freedom and open-endedness could be a drawback. **P09** said “it was difficult to articulate a style that I wanted to replicate.” Similar comments alluded to having a **more structured specification format**—even at the cost of reduced agency. When describing their target style, **P08** said, “I was having a hard time, maybe some type of prompt or dropdown would help.” **P15** proposed “quantifying” style as “variables that can be defined or fed, for example, audience or venue.” While we considered these issues during system conceptualization, we chose to first probe expressiveness from freeform natural language, which performed well in our pilots. From our choice, we can gather insights as to what should be simplified instead of added. We see room to investigate different style specification languages, which can depend on factors connected to a user’s role and background, and the particular writing scenario.

Consider expectations regarding system behavior after learning a new style. GhostWriter never rewrites documents without explicit user action. Such behavior could affect unseen parts of documents, leading to a lack of transparency and erosion of user agency. However, our rationale ran counter to many participants’ expectations. **P14** thought that when the system’s style updates, it would automatically ensure “that the current text gets updated” as well. **P03** added, “Refreshing the style, applying it to the document was not very easy to wrap my mind around.” The idea of explicitly refreshing style (Figure 5b) confused some: “I’m not sure if I hit refresh, is it going to extract the style of this [document] or apply the style I pasted?” (**P15**). This mismatch between expected and actual behavior underscores an **agency / transparency vs. responsiveness trade-off**. On one hand, automatically refreshing documents after style updates could reduce users’ perceived sense of control and awareness when using GhostWriter. On the other hand, people want immediacy in their actions and to avoid the repetitive steps of style change and application.

Agency is desired at different levels. Our work focuses on providing agency for style and context. Having experienced such agency, participants requested similar control over **refining LLM generated outputs**. For example, **P06** said, “I wish I could have accepted only parts of the text it generated,” and many participants wanted to further manipulate the generated text before inserting it into the document. Common intents captured through participant prompts and feedback reveal that additional desired LLM-powered features include expanding/condensing a text selection, or adding specific prompts to make the operations of rewrite and continue more specific (e.g., “Rewrite for a Gen Z audience” or “Continue on this thread and come up a scenario that leads them to explore the mansion”). Our observations and the intrinsic iterative nature of writing underscore the importance of providing user agency at different levels of text generation and style crafting: draft, intermediate, and final.

Agency over the writing process influences one’s sense of outcome ownership but is not sufficient to define it. Our probe allowed us to investigate the relationship between perceived ownership and process agency. We were curious if preserving agency throughout the system could potentially increase **users’ perceived sense of ownership** over the results of collaborative AI writing. In some cases, this was true: “I do [feel ownership] more here because there were so many different things that I could inform to make [it] my own” (**P18**), or “I didn’t actually create” the document, but “I created the prompt. Yeah, I’m putting my name on it” (**P01**). Others were hesitant to claim ownership over the task outcome: “I feel like I architected it, but I didn’t build it” (**P09**), or “There’s something about not feeling like the owner

when using a product like this - more like a partner I should split the by-line with” (**P13**). **P04** shared that the only way they would feel ownership “is if I weren’t using the tool.” This relationship between ownership and agency raises questions about how ownership should be perceived and (possibly) redefined in the age of generative AI. Those looking to preserve it should explore assigning writers a more central role, empowering them to be more than mere spectators or managers when writing with LLMs.

6.3 RQ3: How does increasing personalization and agency in intelligent writing environments impact user behavior and outcomes?

The perceived border between context and style can be blurry, and needs explanation. Many participants explored personalization through editing both **style** and **context** information ([Figure 11](#)). **P07** was “fascinated by how [GhostWriter] combines the sample, style and context to find the right words.” However, some participants initially did not fully grasp the role of style and context: “[I am] trying to understand how style impacted context when they were conflicting in tone” (**P15**), or “[I am not] sure if I should put [information] in the context or the style” (**P04**). Others like **P16** generated hypotheses about whether “the style has more influence than the context” in steering text generations. Ultimately, as style and context specifications are incorporated into LLM meta-prompts, it can be challenging to determine how one piece of information will reinforce, support, or contradict one another. Our goal with GhostWriter was not to arrive at optimal prompts, but to explore the challenges and opportunities in helping users develop helpful mental models of their AI counterparts [[48](#)]. Providing end-users with ways to confidently assert how their input and subsequent changes will steer LLM output is an area that remains essential as we seek systems that enhance user agency and control.

Inline prompts should have awareness of their context and location in a document. Participants developed expectations about the context and information to which a prompt would have access. Inline prompts were commonly used to generate new, independent pieces of content. These operations need nothing more than the system’s style and context. However, sometimes participants wished to insert content in the middle of a document or build on existing content (e.g., “Explain how Riley dealt with the negative impact on his internship” or “Finish the story by having the characters run out of the house”). These requests require particular information within the current document that we did not include in the system’s prompt templates. This limitation stemmed from our intention to develop a functional enough design probe, which resulted in inline prompts occasionally generating misaligned results that frustrated users, e.g., **P02** when trying to continue the Seattle blog post with a prompt: “This is almost like a beginning again. It’s rewriting all of it.” **P02** also saw another inline prompt producing content “unrelated to the previous thought” instead of continuing “from the previous sentence” or “expand[ing] on” it. In production cases, inline AI writing assistance should incorporate the right **contextual dependencies**.

Provide opportunities for reflection about what the system knows. Through **DP4** ([section 3](#)), we aimed to promote user **reflection and discovery** while interacting with GhostWriter. To create opportunities for these activities, we include features such as the main style page, likes & dislikes page, and style history where users can examine the system’s learned style in greater depth. The main **style** page was frequently viewed during study sessions ([Figure 11](#)), and participants were generally satisfied with its content: “That is a very good job of analyzing it” (**P07**). They also tweaked the system’s style by requesting automatic analyses, making manual edits, or adding likes & dislikes. However, the **style history** and **likes & dislikes** page were less used and visited, perhaps due to the potential cost of information processing. **P15** said, “This thing is not telling me much, or at least I didn’t have the patience to go through all of it,” implicitly underscoring the benefits of more succinct representations. **P14** hinted at information usefulness: “It would be cool if you could toggle to see the style, history, or likes in the right nav so you didn’t have to go between the home

page and other options,” suggesting opportunities to further reduce fragmentation and divided attention while writing. Our observations highlight design opportunities to encourage reflection about the state of AI-assisted writing systems and improve understanding on how to act on their responses.

Consider the relationship between writers and AI. Our study provided an opportunity to gain insights into the evolving relationship between writers and AI. After both sessions were completed, we asked participants how they view the role of AI in a system like GhostWriter. 8 participants viewed AI as a **tool** due to reliance on human-provided prompts and a lack of collaboration in generating outputs: “It’s very powerful in generating well written text. But it [needed] instructions to do that” (**P11**). **P08** added, “it would become more of a collaborator” if they “spent more time adding my own text and getting to the point where we were working together.” **P16** explained that “collaborator feels too strong. It feels more like a sounding board. But I don’t think it actually understands my ideas,” suggesting more work is necessary to address human-AI communication gaps. Conversely, 6 participants viewed AI as a **collaborator**, especially for idea generation: “It helps you get started and then you can react to it” (**P12**). Similarly, **P07** said systems like GhostWriter could help “boost creative thought if you have writer’s block.” **P15** viewed AI as a “second piece of eyes, just like you run a presentation by your colleague” but would see it as more of a “potential collaborator” if there was more back and forth between the writer and system. 4 participants saw AI as **both** a tool and collaborator, and 3 thought AI could take an **advisor** role as well, e.g., if it gave “feedback or ideas [to] improve” or pointed out “plot holes or inconsistencies” (**P04**). This aligns with **P13**’s use of GhostWriter to ask for critiques and suggestions, such as “Is this a well-written sentence?” or “What would be a good picture to include here?”

7 LIMITATIONS AND FUTURE WORK

As a design and experience inquiry instrument, GhostWriter had functional limitations. Our system relied on access to OpenAI APIs that were subject to quotas and throttling. This affected some participants’ experiences when using its AI-powered features. However, users were aware of GhostWriter’s prototype nature and understanding when it occasionally produced unexpected results or made them wait.

GhostWriter successfully aided us in testing our ideas for personalizing and enhancing user agency in LLM-mediated writing experiences, fulfilling its role as a design probe. However, our results should be taken in the right context. We could not capture long-term interactions and outcomes with GhostWriter in just two study sessions, e.g., how participants would use such a system over a longer period to perform real-world tasks. Additionally, we evaluated GhostWriter on two task types, a subset of all possible writing activities. As with all think-aloud and interview protocols, we cannot necessarily conclude whether participant quotes accurately reflected their internal states.

Since all participants were employees at a large technology company, they may have more exposure with LLMs and prompting than the average person. We also targeted professions where writing is likely to play a significant role, so our participants are not representative of the general public. In particular, our participant population was unbalanced in terms of gender, possibly reflecting biases in certain industry roles. Irrespective of role and self-reported background, there were no noticeable differences in participants’ accessing and using of GhostWriter’s features.

Our observations and participant feedback reveal that GhostWriter was useful and empowering in creating personalized content based on editable style and context specifications. Our findings also indicate several avenues for future work, e.g., exploring how *user agency can be preserved at different levels* by allowing local (in addition to global) adjustments to style and context, or adding *more fine-grained controls* for adjusting LLM outputs. Research on AI-assisted writing should study ways to incorporate *multiple customizable styles and contexts*, as well as ensure that LLMs have

appropriate levels of contextual awareness when generating text. *Different ways of presenting style descriptions*, or natural language user “profiles,” is another area to investigate for personalized LLM-infused systems.

For all collaborative human-AI writing endeavors, it is crucial to *study user mental models* and design with these expectations in mind. Given participant perspectives on relationship between writers and AI, future work could look at *shaping AI-mediated experiences to help AI fulfill different roles*. For instance, how can we transform LLM-powered writing systems to allow AI to serve more as collaborators or advisors rather than mere tools? Some participants suggested offering critiques or supporting iterative generation processes as a first step. Designing AI systems as collaborators may also help promote reflection and exploration opportunities. Alternatively, perhaps for most end-users, AI functioning as a “tool” already yields satisfactory outcomes, and they would not need AI to take on more collaborative roles. These are open questions that warrant further investigation and could supplement existing work on designing productivity/creativity support tools in the age of AI (e.g., [8, 31]).

It is important to carefully consider and navigate *data privacy trade-offs* when designing personalized systems. While personalization offers numerous benefits, e.g., the ability to learn and apply custom writing styles to documents, there are also risks. After engaging with GhostWriter, P14 asked, “If people put sensitive data in here, is it safe?” Many users may not even be aware of risks associated with entering personal information into data-driven systems. To mitigate these challenges, we advocate for a design approach rooted in transparency and user consent. By providing users with clear communication about how their information is used and stored, as well as global controls to turn on/off data collection [2, 24], we can protect user privacy and offer personalized experiences. These privacy concerns are not new, but they are vital to consider as the popularity of AI-powered personalization systems continues to rise. With thoughtful and ethical design practices, we can turn these challenges into opportunities to build more trustworthy, human-centered applications.

8 POTENTIAL IMPACT

Tools such as GhostWriter bear immense potential benefits and significant challenges. Benefits include democratization of writing and education opportunities. By providing style and context-based assistance, tools like GhostWriter could lower barriers in creating high-quality writing and make it more accessible to a broader range of people, regardless of background or writing skill. This could lead to more voices being heard and represented in various fields. Our observations also insinuate GhostWriter’s potential of being a tool for literacy development around the subject of style. The ability to develop a vocabulary to talk about one’s writing could help students improve their skills and understand different writing styles. This capacity could improve the overall quality of education and encourage deeper engagement with written communications.

Our work also reveals challenges. There are ethical and privacy implications around how tools like GhostWriter use and host data. Besides bias, key issues around non-consensual style appropriation, intellectual property, and authorship require careful consideration, design thinking, and possible regulation. Another challenge centers on the environmental impact of large AI models. An increase in computational power demands could have implications for energy use and associated carbon emissions. In addition to leveraging datacenters that use renewable green energy, we encourage efforts to develop smaller, on-device models that are as capable, yet more energy efficient.

9 CONCLUSION

In this work, we explore how we can design AI-assisted writing systems that allow control over *personalization* and champion user *agency*. Based on a set of research questions and design principles, we designed and implemented

GhostWriter, which we use as a design probe to study the potential of LLMs in crafting personalized writing experiences through style and context.

Our evaluation revealed that GhostWriter is effective in preserving user agency when interacting with a system with AI-fueled writing assistance. Participant feedback also illustrated the value in offering both implicit and explicit mechanisms to teach the system about one's writing preferences.

Guided by these findings, we present a series of design lessons to shape the future of collaborative human-AI writing. As a whole, we hope our work can inspire others looking to leverage LLMs to augment and complement human capabilities, providing a reference for exploring the new challenges and opportunities that arise when designing and using these emerging technologies.

ACKNOWLEDGMENTS

REFERENCES

- [1] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *Ai Magazine* 35, 4 (2014), 105–120.
- [2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
- [3] Soulef Benhamdi, Abdesselam Babouri, and Raja Chiky. 2017. Personalized recommender system for e-Learning environment. *Education and Information Technologies* 22 (2017), 1455–1477.
- [4] Karim Benharrak, Tim Zindulkia, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2023. Writer-Defined AI Personas for On-Demand Feedback Generation. *arXiv preprint arXiv:2309.10433* (2023).
- [5] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1077–1086.
- [6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [7] Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, Zhifeng Chen, et al. 2019. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2287–2295.
- [8] John Joon Young Chung, Shiqing He, and Eytan Adar. 2022. Artist support networks: Implications for future creativity support tools. In *Designing Interactive Systems Conference*. 232–246.
- [9] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [10] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces*. 329–340.
- [11] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond text generation: Supporting writers with continuous automatic text summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [12] Nouha Dziri, Ehsan Kamalloo, Kory W Mathewson, and Osmar Zaiane. 2018. Augmenting neural response generation with context-aware topical attention. *arXiv preprint arXiv:1811.01063* (2018).
- [13] Xiaoxuan Fang, Davy Tsz Kit Ng, Jac Ka Lok Leung, and Samuel Kai Wah Chu. 2023. A systematic review of artificial intelligence technologies used for story writing. *Education and Information Technologies* (2023), 1–37.
- [14] Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. Style transfer in text: Exploration and evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [15] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.
- [16] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social dynamics of AI support in creative writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [17] Connor Graham and Mark Rouncefield. 2008. Probes and participation. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*. 194–197.
- [18] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M Drucker. 2019. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
- [19] Md Naimul Hoque, Bhavya Ghai, Kari Kraus, and Niklas Elmquist. 2023. Portrayal: Leveraging NLP and Visualization for Analyzing Fictional Characters. (2023).

- [20] Zhiqiang Hu, Roy Ka-Wei Lee, Charu C Aggarwal, and Aston Zhang. 2022. Text style transfer: A review and experimental evaluation. *ACM SIGKDD Explorations Newsletter* 24, 1 (2022), 14–45.
- [21] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Drui, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 17–24.
- [22] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative writing with an ai-powered writing assistant: Perspectives from professional writers. *arXiv preprint arXiv:2211.05030* (2022).
- [23] Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, and Rada Mihalcea. 2022. Deep learning for text style transfer: A survey. *Computational Linguistics* 48, 1 (2022), 155–205.
- [24] Matthew Jörke, Yasaman S Sefidgar, Talie Massachi, Jina Suh, and Gonzalo Ramos. 2023. Pearl: A Technology Probe for Machine-Assisted Reflection on Personal Data. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 902–918.
- [25] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 115–135.
- [26] Tae Soo Kim, Yoonjoo Lee, Jamin Shin, Young-Ho Kim, and Juho Kim. 2023. EvalLM: Interactive Evaluation of Large Language Model Prompts on User-Defined Criteria. *arXiv preprint arXiv:2309.13633* (2023).
- [27] Tae Soo Kim, Arghya Sarkar, Yoonjoo Lee, Minsuk Chang, and Juho Kim. 2023. LMCanvas: Object-Oriented Interaction to Personalize Large Language Model-Powered Writing Environments. *arXiv preprint arXiv:2303.15125* (2023).
- [28] Tae Wook Kim and Quan Tan. 2023. Repurposing Text-Generating AI into a Thought-Provoking Writing Tutor. *arXiv preprint arXiv:2304.10543* (2023).
- [29] Vivek Kulkarni and Vibul Raheja. 2023. Writing Assistants Should Model Social Factors of Language. *arXiv preprint arXiv:2303.16275* (2023).
- [30] Moritz Langner, Peyman Toreini, and Alexander Maedche. [n. d.]. A Situation Awareness Perspective on Intelligent Writing Assistants: Augmenting Human-AI Interaction with Eye Tracking Technology. ([n. d.]).
- [31] Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. 2023. When is a Tool a Tool? User Perceptions of System Agency in Human-AI Co-Creative Drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1978–1996.
- [32] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.
- [33] Cheng Li, Mingyang Zhang, Qiaozhu Mei, Yaqing Wang, Spurthi Amba Hombaiah, Yi Liang, and Michael Bendersky. 2023. Teach LLMs to Personalize—An Approach inspired by Writing Education. *arXiv preprint arXiv:2308.07968* (2023).
- [34] Jinming Li, Wentao Zhang, Tian Wang, Guanglei Xiong, Alan Lu, and Gerard Medioni. 2023. GPT4Rec: A generative framework for personalized recommendation and user interests interpretation. *arXiv preprint arXiv:2304.03879* (2023).
- [35] Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. 2021. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. *arXiv preprint arXiv:2104.08786* (2021).
- [36] Cerstin Mahlow. 2023. Writing Tools: Looking Back to Look Ahead. *arXiv preprint arXiv:2303.17894* (2023).
- [37] Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 107–110.
- [38] 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*. 1–34.
- [39] Sheshera Mysore, Mahmood Jasim, Andrew McCallum, and Hamed Zamani. 2023. Editable User Profiles for Controllable Text Recommendation. *arXiv preprint arXiv:2304.04250* (2023).
- [40] Felicia Y. Ng, Jina Suh, and Gonzalo A. Ramos. 2020. Understanding and Supporting Knowledge Decomposition for Machine Teaching. *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (2020). <https://api.semanticscholar.org/CorpusID:219487290>
- [41] Jakob Nielsen. 2005. Ten usability heuristics. (2005).
- [42] Haekyu Park, Gonzalo Ramos, Jina Suh, Christopher Meek, Rachel Ng, and Mary Czerwinski. 2023. FoundWright: A System to Help People Re-find Pages from Their Web-history. *arXiv preprint arXiv:2305.07930* (2023).
- [43] Ajay Patel, Delip Rao, and Chris Callison-Burch. 2023. Learning Interpretable Style Embeddings via Prompting LLMs. *arXiv preprint arXiv:2305.12696* (2023).
- [44] Napol Rachatasumrit, Gonzalo Ramos, Jina Suh, Rachel Ng, and Christopher Meek. 2021. ForSense: Accelerating online research through sensemaking integration and machine research support. In *26th International Conference on Intelligent User Interfaces*. 608–618.
- [45] Filip Radlinski, Krisztian Balog, Fernando Diaz, Lucas Dixon, and Ben Wedin. 2022. On Natural Language User Profiles for Transparent and Scrutable Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2863–2874.
- [46] Vibul Raheja, Dhruv Kumar, Ryan Koo, and Dongyeop Kang. 2023. CoEdIT: Text Editing by Task-Specific Instruction Tuning. *arXiv preprint arXiv:2305.09857* (2023).
- [47] Gonzalo Ramos, Christopher Meek, Patrice Simard, Jina Suh, and Soroush Ghorashi. 2020. Interactive machine teaching: a human-centered approach to building machine-learned models. *Human–Computer Interaction* 35, 5-6 (2020), 413–451.
- [48] Gonzalo A. Ramos, Felicia Y. Ng, Nicole Sultanum, Christopher Meek, Jina Suh, and Soroush Ghorashi. 2019. Do Machine Teachers Dream of Algorithms. <https://api.semanticscholar.org/CorpusID:213596802>

- [49] Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2023. LaMP: When Large Language Models Meet Personalization. *arXiv preprint arXiv:2304.11406* (2023).
- [50] Patrice Y Simard, Saleema Amershi, David M Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, et al. 2017. Machine teaching: A new paradigm for building machine learning systems. *arXiv preprint arXiv:1707.06742* (2017).
- [51] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L Glassman. 2022. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction* (2022).
- [52] Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. *arXiv preprint arXiv:1506.06714* (2015).
- [53] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.
- [54] Zijie J Wang, Aishwarya Chakravarthy, David Munechika, and Duen Horng Chau. 2024. Wordflow: Social Prompt Engineering for Large Language Models. *arXiv preprint arXiv:2401.14447* (2024).
- [55] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, Vol. 10. CEUR-WS Team.
- [56] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *27th International Conference on Intelligent User Interfaces*. 841–852.
- [57] Haopeng Zhang, Xiao Liu, and Jiawei Zhang. 2023. SummIt: Iterative Text Summarization via ChatGPT. *arXiv preprint arXiv:2305.14835* (2023).
- [58] Tianyi Zhang, Faisal Ladha, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B Hashimoto. 2023. Benchmarking large language models for news summarization. *arXiv preprint arXiv:2301.13848* (2023).
- [59] Xin Zhao. 2022. Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC Journal* (2022), 00336882221094089.
- [60] Hanxun Zhong, Zhicheng Dou, Yutao Zhu, Hongjin Qian, and Ji-Rong Wen. 2022. Less is more: Learning to refine dialogue history for personalized dialogue generation. *arXiv preprint arXiv:2204.08128* (2022).
- [61] Wenxuan Zhou, Sheng Zhang, Hoifung Poon, and Muhaoo Chen. 2023. Context-faithful prompting for large language models. *arXiv preprint arXiv:2303.11315* (2023).

A APPENDIX: LLM PROMPTS & OPERATIONS

We list the LLM operations and prompts used in GhostWriter’s backend, which uses OpenAI’s API. We include a summary of all backend operations in [Table 2](#). We use LangChain’s chat model implementations to allow separate SystemMessage and HumanMessage prompts, setting `max_tokens = 512` in both cases. Corresponding prompt templates for each operation are included below, with HumanMessage prompts in *blue* and SystemMessage prompts in *gray*.

A.1 Prompts by Operation

get_style. This function is the first step when an automatic or manual **style update** is requested by the user. The HumanMessage specifies the *text* to extract style from, as well as the user’s list of *likes* and *dislikes*, if available.

HumanMessage

```
Text: {text}
Likes: {likes}
Dislikes: {dislikes}
```

Then, the SystemMessage uses this information to generate a description of the text’s writing style.

SystemMessage

```
You are a literary critic who is analyzing the writing style of a writing sample.  
Describe each of these characteristics in order: tone, voice, word choice, sentence structure, and paragraph structure.
```

Table 2. LLM operations used in GhostWriter's backend.

Operation	Inputs	Outputs
<code>get_style</code> : extract style from a single writing sample, incorporating likes & dislikes (if any)	text, likes (<i>optional</i>), dislikes (<i>optional</i>)	new style, style summary
<code>compare_styles</code> : compare two writing styles and generate similarity rating	style1, style2	comparison, comparison summary, similarity rating
<code>continue_text</code> : generate new text to continue a piece of writing in a given style	text, style, context	new text
<code>rewrite_text</code> : rewrite text in a given writing style	text, style	rewritten text
<code>generate_text</code> : generate text to address an <i>in-line</i> prompt, following a given writing style and context	prompt, style, context	new text
<code>ask_prompt</code> : generate text to address a <i>contextual</i> prompt for the specified document, following a given writing style	document, prompt, style	new text
<code>summarize_text</code> : generate summary of a style description/comparison or list of likes/dislikes	text, summary type	summary
<code>format_text</code> : format text as an HTML string	text	formatted HTML string

For voice, state whether the writer uses passive or active voice.

Include examples from the text as evidence for tone, voice, and word choice only.

For sentence structure and paragraph structure, also comment on how varied they are.

Please incorporate information about the writer's likes and dislikes into the writing style.

DO NOT include any specific context about the text, such as the topic, theme, character names, or genre.

Describe one aggregate writing style for the entire writing sample. **DO NOT separate your response based on different texts.**

Your answer should be 3-5 sentences long maximum.

Please DO NOT format your answer as a numbered or bulleted list.

compare_styles. This function is also called when a **style update** is requested. The HumanMessage provides two writing styles, `style1` and `style2`, to compare.

HumanMessage

Here is the previous style: {`style1`}.

Here is the new style: {`style2`}.

Then, the SystemMessage produces a qualitative comparison of these two styles, as well as a numerical rating to indicate how similar they are. When passed to the frontend, this "similarity" rating is inverted to produce the "difference" rating shown to participants, as we thought users would be more interested in finding contrasts between writing styles.

SystemMessage

Imagine you are a literary critic who is comparing two writing styles.

First, state how similar the styles are on a scale from 1 (not at all similar) to 10 (practically identical).

For example, "I would rate the similarity of these styles as 6 out of 10." **Replace 6 with your rating.**

Then, describe the how the new style differs from the previous, using examples from the text as evidence.

Your answer should be 3-5 sentences long maximum.

Please DO NOT format your answer as a numbered or bulleted list.

continue_text. This function is called when the user invokes the corresponding **continue** feature in GhostWriter. The HumanMessage simply provides the *text* to be continued.

HumanMessage

{text}

Given this text, the *SystemMessage* is prompted to generate an appropriate continuation using the system's current *style* and *context*.

SystemMessage

You are a ghostwriter who must continue a piece of writing using this writing style: {style}.

First, determine what kind of text this is (e.g., short story, poem, song, news article, etc.), but DO NOT mention this in your response. Then, continue the piece of writing using the same style.

You do not need to finish the piece, but **you should write at least one paragraph/verse/stanza/etc, and no more than 3** to continue it.

Use new lines to match the sentence and paragraph structure of the original text.

For example, add a new line after each paragraph in a story or article, and after each line in a poem.

Follow formatting instructions if provided too (e.g., write in bullet points).

Make sure to follow all of these rules too:

DO NOT mention ghostwriting or the original author/writer.

DO NOT include, refer to, or summarize any parts of the style description in your response.

For example, if a character named Mary was mentioned in the style description, **DO NOT** mention Mary in your response unless prompted to.

DO NOT repeat text or generate similar content to what is already in the original piece of writing.

Please use the same topic, genre, point of view, and theme of the original text.

For example, if the original text is a poem about a dog, you should continue the poem about the same dog.

DO NOT start a new piece of writing.

If relevant, keep this context in mind when continuing the text: {context}.

rewrite_text. This function is called when the user invokes the corresponding **rewrite** feature in GhostWriter. The HumanMessage simply provides the *text* to be rewritten.

HumanMessage

{text}

Then, the SystemMessage is prompted to rewrite the given text using the system's current *style*.

SystemMessage

You are a ghostwriter who must rewrite the given text concisely using this writing style: {style}.

First, determine what kind of text this is (e.g., short story, poem, song, news article, etc.), but DO NOT mention this in your response. Then, rewrite the text using the given style.

Use new lines to match the length, sentence structure, and paragraph structure of the original text.

For example, add a new line after each paragraph in a story or article, and after each line in a poem.

DO NOT mention ghostwriting or the original author/writer.

DO NOT include, refer to, or summarize any parts of the style description in your response.

Please use the same topic, genre, point of view, and theme of the original text.

generate_text. This function is called the **inline prompt** feature is invoked. The HumanMessage simply provides the *prompt* written by the user.

HumanMessage

{prompt}

Then, the SystemMessage is prompted to generate text given this prompt, while taking into account the system's current *style* and *context*.

SystemMessage

You are a ghostwriter who must generate new text using this writing style: {style}.

If relevant, keep this context in mind when generating text: {context}.

DO NOT mention ghostwriting or the original author/writer.

DO NOT include, refer to, or summarize any parts of the style description in your response.

For example, if a character named Mary was mentioned in the style description, **DO NOT** mention Mary in your response unless prompted to.

Also, **DO NOT** start a new piece of writing (e.g., start a new story) unless prompted to.

ask_prompt. This function is called the **contextual prompt** feature is invoked. The HumanMessage provides the *document* (i.e., text selected by user), the *prompt* to apply to this document, and the system's current *style* to use when generating text. No SystemMessage is used in this case.

HumanMessage

You are given a document and a prompt. The prompt is a question or a task that you need to answer or complete.

The document is a piece of text that you can use to answer the question or complete the task. If no document is provided, you can use your own knowledge to answer the question or complete the task, and do not mention the absence of a document. The document and prompt are specified in the following data structure:

`{'document': {document}, 'prompt': {prompt}}`

You are also given a style description of the writing style that you must use to answer the question or complete the task. **DO NOT** include, refer to, or summarize any parts of the style description in your response. The style description is specified in the following data structure:

`{"style": {style}}`

Given this document, prompt, and style answer the question or complete the task.

summarize_text. This function is used to generate summaries for **style descriptions** and **comparisons**, as well as **likes & dislikes**. We provide different HumanMessages to generate the corresponding summary, but each takes a single input: the *text* to be summarized. No SystemMessages are used in this case.

- Writing Style:

HumanMessage

Summarize the key points of this writing style description using three short, comma-separated phrases:
`{text}`.

- Style Comparison:

HumanMessage

Given this comparison of two writing styles: `{text}`, summarize how the new writing style changed from the previous using one, short sentence.

For example, “Style has changed to be more lighthearted, with simpler words and straightforward structure”

Replace this summary with your own answer.

- Likes:

HumanMessage

In 1 sentence, summarize what these passages that the writer likes has in common: `{text}`.

If justifications are provided, incorporate them into your summary.

Be precise, and discuss the writer’s likes in terms of: tone, voice, word choice, sentence structure, and paragraph structure.

Start your response with “The writer likes...”.

- Dislikes:

HumanMessage

In 1 sentence, summarize what these passages that the writer dislikes has in common: `{text}`.

If justifications are provided, incorporate them into your summary.

Be precise, and discuss the writer’s dislikes in terms of: tone, voice, word choice, sentence structure, and paragraph structure.

Start your response with “The writer dislikes...”.

format_text. This function is used to format **style descriptions** and **comparisons** as HTML objects to display in the frontend. The HumanMessage simply provides the *text* to be formatted.

HumanMessage

{text}

Then, the SystemMessage is prompted to format the given text (e.g., add bolding and italics for emphasis).

SystemMessage

You are a web developer who is formatting a paragraph of text as HTML.

First, please split the paragraph into individual sentences and wrap each sentence with a `<p>` tag.

****Your answer should have at least three `<p>` tags.****

Then, inside each `<p>` tag, add one `` tag and multiple `` tags to emphasize key words and phrases.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009