

Narrative Keyframing for Generative Creative Writing

Chao Zhang
Cornell University
Ithaca, New York, USA
cz468@cornell.edu

Abe Davis
Cornell University
Ithaca, New York, USA
abedavis@cornell.edu

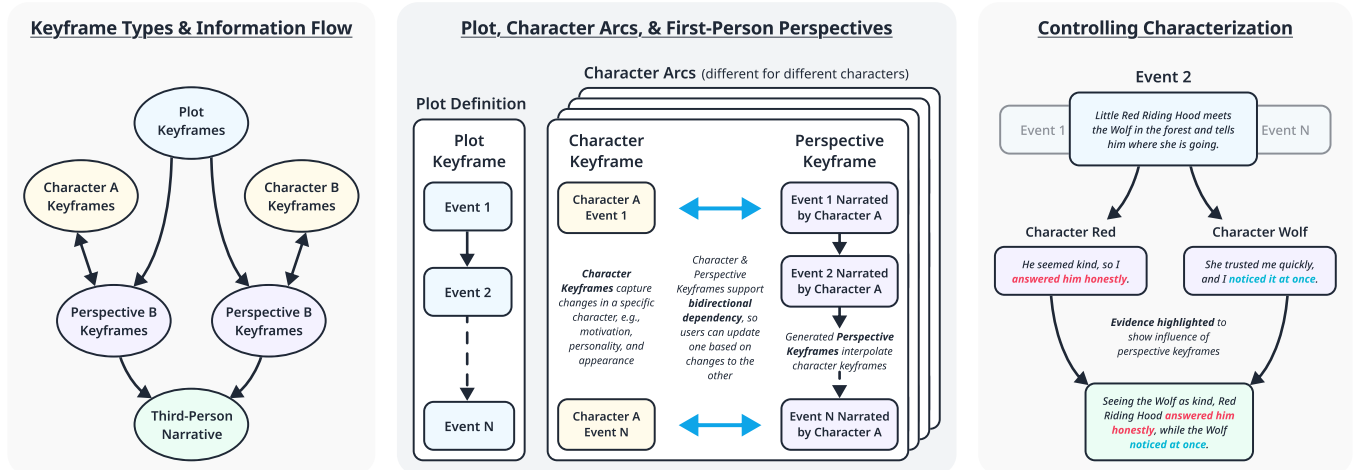


Figure 1: Narrative Keyframing for Generative Creative Writing. Our approach introduces *narrative keyframes* as high-level intermediate representations that connect story planning to narrative generation through three linked forms: plot keyframes, character keyframes, and perspective keyframes (Left). Writers define plot keyframes across events, specify character keyframes to capture how each character changes, and generate first-person perspective keyframes that can be iteratively refined in relation to character states (Center). During story generation, selected evidence from each character’s perspective keyframes provides traceability for how characterization decisions shape the resulting third-person narrative (Right).

Abstract

We introduce *narrative keyframing*, an interaction technique for AI-assisted creative writing that lets writers specify different types of narrative constraints at selected moments in a story, then use AI to generate intervening prose. Inspired by the use of keyframing in animation, narrative keyframing offers a flexible way to connect story planning with adaptive control over generated text. We explore three types of keyframes: *plot keyframes* define significant events in a story, *character keyframes* represent how individual characters change over the narrative, and *perspective keyframes* capture how individual characters experience different events through first-person narratives. Plot and character keyframes offer a flexible way to adapt the type of high-level conditioning explored in previous AI writing tools to more customizable, iterative, and fine-scale control, while perspective keyframes add a new way to control characterization and focalization by using first-person narratives as an intermediary. Through a user study, we show that narrative keyframing supports a more controllable, transparent, and engaging way to use generative AI in creative writing.



This work is licensed under a Creative Commons Attribution 4.0 International License. [UIST '26, Detroit, MI, USA](https://creativecommons.org/licenses/by/4.0/)

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2856-3/26/11
<https://doi.org/10.1145/3830398.3830586>

CCS Concepts

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

Keywords

Creative Writing; Human-AI Interaction

ACM Reference Format:

Chao Zhang and Abe Davis. 2026. Narrative Keyframing for Generative Creative Writing. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26)*, November 02–05, 2026, Detroit, MI, USA. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3830398.3830586>

1 Introduction

In animation, the practice of keyframing offers an incredibly flexible and efficient way to balance automation with creative control. The basic idea is simple: most of the important details in an animation can be derived from constraints on a sparse set of key moments, from which the rest of the animation can be interpolated [29]. By controlling the type and distribution of such constraints across a timeline, users can focus creative effort where it is most necessary and leverage automation where it is most appropriate.

Analogously, creative writing often follows a similar workflow, with writers planning out properties of key moments in a story before connecting those moments with actual prose. In literary theory,

this can be described as developing *plot*, which encompasses the core events of a story, before writing *narrative*, which encompasses how those events are ultimately presented to the reader [2, 15, 18].

Recent AI writing systems have leveraged high-level story descriptions, character specifications, or plot outlines to condition the generation of narrative prose (e.g., [35, 38, 42]). However, these systems typically treat such conditioning as a set of static prompts or global conditions, limiting fine-grained control over individual story elements and how they change throughout the generated narrative.

Inspired by animation keyframing (Fig. 2), we introduce *narrative keyframing*, a new interaction technique for AI-assisted writing. Just as keyframes in animation let artists specify key changes to different animation properties across a timeline, we introduce *narrative keyframes* as a way for authors to specify key changes to different narrative properties across a story. We explore three types of narrative keyframes: *plot keyframes*, *character keyframes*, and *perspective keyframes*. Plot keyframes represent events that take place in a story, while character keyframes represent how individual characters change across the narrative. Together, these two types of keyframes generalize the global plot and character conditioning explored in prior work by providing more adaptive and fine-grained event-level control over generated text. Our third type of keyframes, perspective keyframes, uses written or generated first-person prose to capture how different characters experience particular events. Perspective keyframes introduce first-person narratives as a novel intermediate representation for controlling characterization and focalization in subsequently generated prose. Authors can select and recombine elements from different perspective keyframes that represent how different characters experience a common event to balance how that event is portrayed in a third-person narrative.

We evaluate our approach through a technical assessment that combines automatic metrics and expert judgments, as well as a user study with 12 writers. Results show that our system produces stories with higher overall quality and richer characterization than a standard LLM baseline, and supports a more controllable, transparent, and engaging writing experience.

2 Narratological Foundations

Narrative keyframing could, in principle, be applied to many narrative properties. In our design, plot events provide the temporal structure on which writers keyframe evolving character states and character-specific perspectives. We therefore draw on theories of character arcs, characterization, and focalization to motivate these representations and the relationships among them.

Character Arcs: Character development unfolds throughout a story as various aspects of a character are selectively revealed, reinforced, and transformed at different stages. Forster’s distinction between flat and round characters suggests that flat characters remain relatively stable, while round characters tend to become more complex and develop over the course of the narrative [15]. This pattern of development for one character across the narrative is often described in creative writing practice as a *character arc* [54]. For example, a character may appear self-doubting at the beginning, conflicted in the middle, and confident by the end. Shaping a character arc requires deciding when particular traits become visible, how they are reinforced across scenes, and how they later change. This

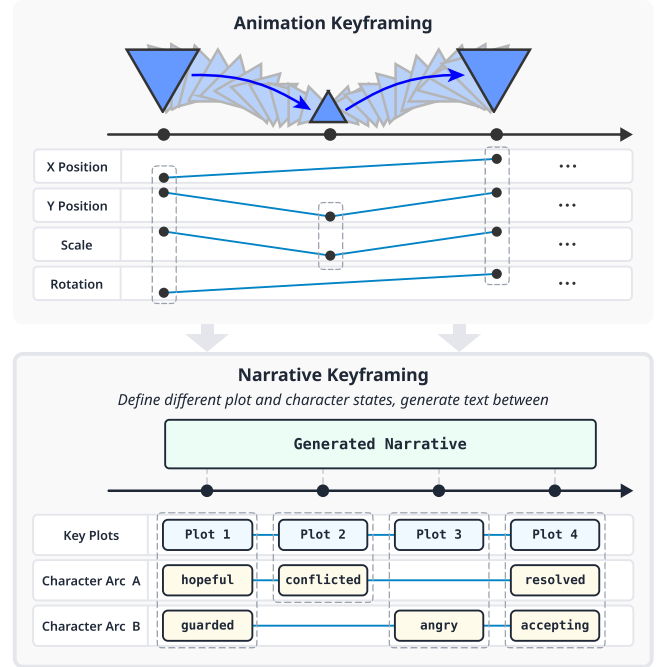


Figure 2: From animation keyframing to narrative keyframing. Animation keyframing specifies the values of different animation properties at selected points in time and interpolates the states between them. narrative keyframing applies the same interaction principle to generative creative writing: users define key plot events and character states at important points in a story, which jointly condition the generation of narrative text between those points.

informs our design choice to help writers develop character arcs by defining salient character keyframes at key moments in the story.

Characterization: Writing theory distinguishes between two complementary processes in characterization: *conceptualization*, in which authors form an understanding of a character’s attributes, and *exposition*, in which those attributes are expressed through the narrative medium [51]. In conceptualization, Egri’s character “bone structure” [12] frames characters as multidimensional constructions spanning *physiology* (e.g., age and appearance), *psychology* (e.g., personality, values, and motivations), and *sociology* (e.g., profession, status, and relationships). However, rich conceptualization alone does not ensure effective characterization. Story quality also depends on how character attributes are conveyed in specific plots [17, 21, 51]. Rather than being stated explicitly, traits are often conveyed indirectly through a range of narrative “expositors,” including physical appearance, actions, thoughts, dialogue, setting, and symbolic elements [41]. These theories inform our design choice to support both the development of rich character concepts and their translation into concrete narrative details.

Focalization: Characterization is shaped by the narrative perspective through which character attributes are made available to readers [2, 18]. Narratology describes this in terms of *focalization* [1]:

Table 1: Example of perspective-based recombination. From a single story event, writers generate first-person perspectives for different characters, then selectively recombine evidence from those perspectives to produce third-person narratives with different characterization emphases. *Red text* indicates phrases drawn from Red’s perspective, and *blue text* indicates phrases drawn from the Wolf’s perspective.

Stage	Narrative Content	Characterization Focus
Source Plot	Little Red Riding Hood meets the Wolf in the forest and tells him where she is going.	—
First-Person Perspectives		
First-Person – Red	He <i>seemed kind</i> , so <i>I answered him honestly</i> . <i>I was curious</i> and <i>did not sense any danger</i> .	Red’s perspective
First-Person – Wolf	She <i>trusted me quickly</i> , and <i>I noticed it at once</i> . <i>I spoke softly</i> so she would <i>keep talking</i> .	Wolf’s perspective
Third-Person Narratives		
Third-Person – A	Seeing the Wolf as <i>kind</i> , Red Riding Hood <i>answered him honestly</i> , while the Wolf <i>noticed at once</i> that she <i>trusted him quickly</i> .	Red’s <i>trust</i> ; Wolf’s <i>opportunism</i>
Third-Person – B	<i>Curious</i> and unable to <i>sense any danger</i> , Red Riding Hood kept speaking with the Wolf, while the Wolf <i>spoke softly</i> so that <i>she would keep talking</i> .	Red’s <i>curiosity</i> ; Wolf’s <i>manipulation</i>
Third-Person – C	Because the Wolf <i>seemed kind</i> and she <i>did not sense any danger</i> , Red Riding Hood lowered her guard, while the Wolf <i>noticed it at once</i> and <i>spoke softly</i> to cultivate her trust.	Red’s <i>misjudgment</i> ; Wolf’s <i>calculation</i>

narrative discourse may be organized around what a particular character perceives, knows, and feels, or it may shift across multiple characters within a broader narrative frame. Within this framing, *first-person narration* provides access to a character’s subjective experience, making it useful for exploring expository details such as perception, interpretation, emotion, and self-understanding. *Third-person narration*, by contrast, provides a broader frame for organizing focalization across multiple characters and scenes. When writing a third-person story, first-person narration can serve as an intermediate step for externalizing character-specific expository material. This is similar to point-of-view writing exercises [26] (also describe as “method writing” [20]) used by practitioners, in which writers explore a scene from a particular character’s perspective before integrating those insights into the broader narrative. This informs our design choice to let writers explore each character’s arc through perspective-specific narration, then compose a final story that integrates those viewpoints.

3 Related Work

3.1 Keyframing in Animation

The practice of animation keyframing dates back long before the invention of computers. The earliest version was used as a pre-visualization strategy in which artists would draw keyframes to plan the flow of an animation before committing effort to drawing all the remaining frames. When animators started working in teams, this became a way to divide labor: a lead artist would draw detailed keyframes, which assistants or trainees would then interpolate [29]. With the advent of computers, the practice of keyframing evolved into a general technique for interpolating between artist-specified constraints across an animation timeline. Modern tools let users create different types of keyframes to control different properties of an animation (e.g., position, rotation, color), and users can adjust the density of keyframes over a timeline to adaptively balance interpolation with finer-grained creative control.

Abstractly, we can think of keyframing as a flexible and efficient way to balance automation with creative control across a timeline. It requires only two things: a way to localize creative constraints on the timeline and a way to interpolate between those constraints. The key insight of our work is that recent advances in generative language models make an analogous form of interpolation possible

for narrative. Language models can generate coherent narrative developments and prose between sparse constraints specified at important points in a story. This allows writers to anchor major plot events and character states while delegating intermediate developments to AI generation. As in animation, writers can vary both the type and density of keyframes to determine where direct authorial control is most important and where greater automation is appropriate.

3.2 Intelligent Writing Interfaces

The HCI community has long been interested in intelligent writing tools [30] that support writers across a range of writing tasks, including brainstorming ideas [7, 19, 42, 63], planning outlines [40, 52], drafting content [11, 22, 25, 27, 28, 61, 62], and refining text [24, 32, 39, 50, 60]. These tools span diverse genres, including argumentative writing [59, 64], story writing [8, 23, 57], and scientific writing [44, 48]. Within this broader design space, recent work has explored interaction metaphors drawn from established creative and design practices to develop visual interfaces for human-AI co-writing. Rather than relying solely on natural language prompts, these systems externalize aspects of the writing process into visual representations that allow writers to control AI generation. Different metaphors emphasize different forms of authorial control, ranging from high-level planning [9, 10, 58] to local text revision [33, 45].

For example, TaleBrush [9] supports control over generated stories by editing 2D curves that represent attributes such as surprise, while CharacterChat [42] and CharacterMeet [38] help writers construct global character personas via role-play. These methods operate at a more global level, while ours allows for specification of detailed character and plot conditions at specific plot points. Another line of research focuses on local text-level control. For instance, Texterial [45] conceptualizes text as clay, allowing users to refine generated content through gestural sculpting, while Textoshop [33] borrows interactions from drawing software to support editing operations such as shortening and reordering text. These systems provide useful mechanisms for revising surface-level text but do not explicitly support control over the narrative dimensions of a story.

A group of systems more closely related to our work explores how writing can be represented as the manipulation of visual structures composed of discrete writing elements. For example,

VISAR [64] and Polymind [52] draw inspiration from node-based visual programming, enabling users to control text generation through interconnected nodes. However, VISAR focuses on supporting logical structure in argumentative writing, while Polymind emphasizes constructing AI workflows from microtasks such as “summarize” and “brainstorm,” rather than controlling the progression of a narrative. Dramatron is closer to our work in its staged decomposition of story generation into narrative elements, including characters, plot, locations, and dialogue. However, this decomposition is controlled through sequential user prompts in a Colab environment, limiting users to comparatively rigid interactions within a fixed hierarchical decomposition of the story.

Uniquely, our keyframing interaction allows authors to control different types of narrative constraints (e.g., plots and characters) at selected moments in a story while leaving the progression between them to AI generation. This approach provides fine-grained yet flexible control over character arcs and narrative development throughout the writing process.

3.3 Interactive Characterization Tools

Recent HCI systems have explored characterization as an interactive process, often by enabling writers to build characters through simulation and dialogue with LLM-powered personas. For example, CharacterChat [42] supports character creation through conversational roleplay and progressive manifestation, while CharacterMeet [38] extends this idea to support writers throughout the broader process of story character construction via chatbot avatars. Related systems similarly use persona-driven or multi-agent character simulation to help writers explore character traits, backstory, and possible narrative developments [4, 16, 37, 53]. These systems primarily help users define global, static character sheets, but offer limited support for understanding and controlling how an established character profile is expressed in a story or how character traits evolve across plot events. To address this gap, our work uses first-person perspectives as an intermediate representation for conditioning narrative generation. First-person perspectives instantiate abstract character definitions into concrete narrative text, allowing users preview how a character is portrayed before selecting what to emphasize in the final prose. First-person perspectives also fit naturally within our keyframing interaction by enabling users to shape the progression of a character arc at key plot events. To our knowledge, ours is the first work to explore first-person character narratives as an intermediate representation for controlling AI-assisted creative writing.

4 Design Goals

Drawing on the narratological theories and related HCI work discussed above, we derive three design goals for our instantiation of narrative keyframing. We use characterization as a concrete design context for exploring how narrative keyframes can support control across story planning, perspective exploration, and narrative generation. The resulting goals concern how writers shape character development across a story, translate abstract character ideas into narrative form, and understand how those decisions are reflected in generated text.

- **[DG1] *Shaping character development across plot points.*** Characterization unfolds over narrative progression rather than appearing all at once [15]. Writers should be able to represent how a character changes at key moments in the story, inspect that progression, and revise it as the narrative develops. This goal follows from theories of character arcs [54] and motivates support for planning characterization across plot points rather than specifying characters only once at the beginning.
- **[DG2] *Manifesting abstract character traits into concrete narrative evidence.*** Characterization depends not only on defining who a character is, but also on expressing those qualities through narrative details [17, 21, 51]. Writers should be able to explore how abstract traits may be realized in language through individual characters’ perspectives, then inspect concrete evidence such as actions, thoughts, dialogue, appearance, and setting. This goal follows from theories of conceptualization, exposition [51], and focalization [1], and motivates support for using perspective-specific narration as an intermediate step between character planning and story generation.
- **[DG3] *Tracing how characterization decisions propagate into generated story text.*** Because characterization may be developed through multiple intermediate steps (e.g., conceptualization of characters, first-person perspectives), writers should be able to inspect how earlier decisions influence later generations. This includes tracing how character traits and perspective-specific details are carried into the third-person narrative. This goal follows from theories of focalization and perspective [1, 2, 18], and motivates interfaces that make the relationship between characterization inputs and generated story text visible.

5 System Design

Informed by these design goals, we instantiate narrative keyframing in an interactive system that connects story planning, character development, perspective exploration, and narrative generation. The system is organized around three linked keyframe types—plot, character, and perspective—that allow writers to specify narrative constraints at key plot points and use them to guide the generation of the final story. In this section, we describe the overall workflow of this system, the three types of keyframes, and the implementation details.

5.1 Workflows and Views

Our system supports a flexible and iterative workflow that moves from story planning to character and perspective exploration, and finally to narrative generation. Users (1) create a story outline organized by plot keyframes; (2) define character keyframes to represent how each character develops across the story; (3) generate perspective keyframes to explore how those traits may be expressed in narrative form; (4) select the character traits and textual evidence they want to emphasize; and (5) generate third-person narratives conditioned on those keyframes and selections. The following subsections describe the features that support each stage of this workflow.

To support this workflow, the interface consists of three coordinated views:

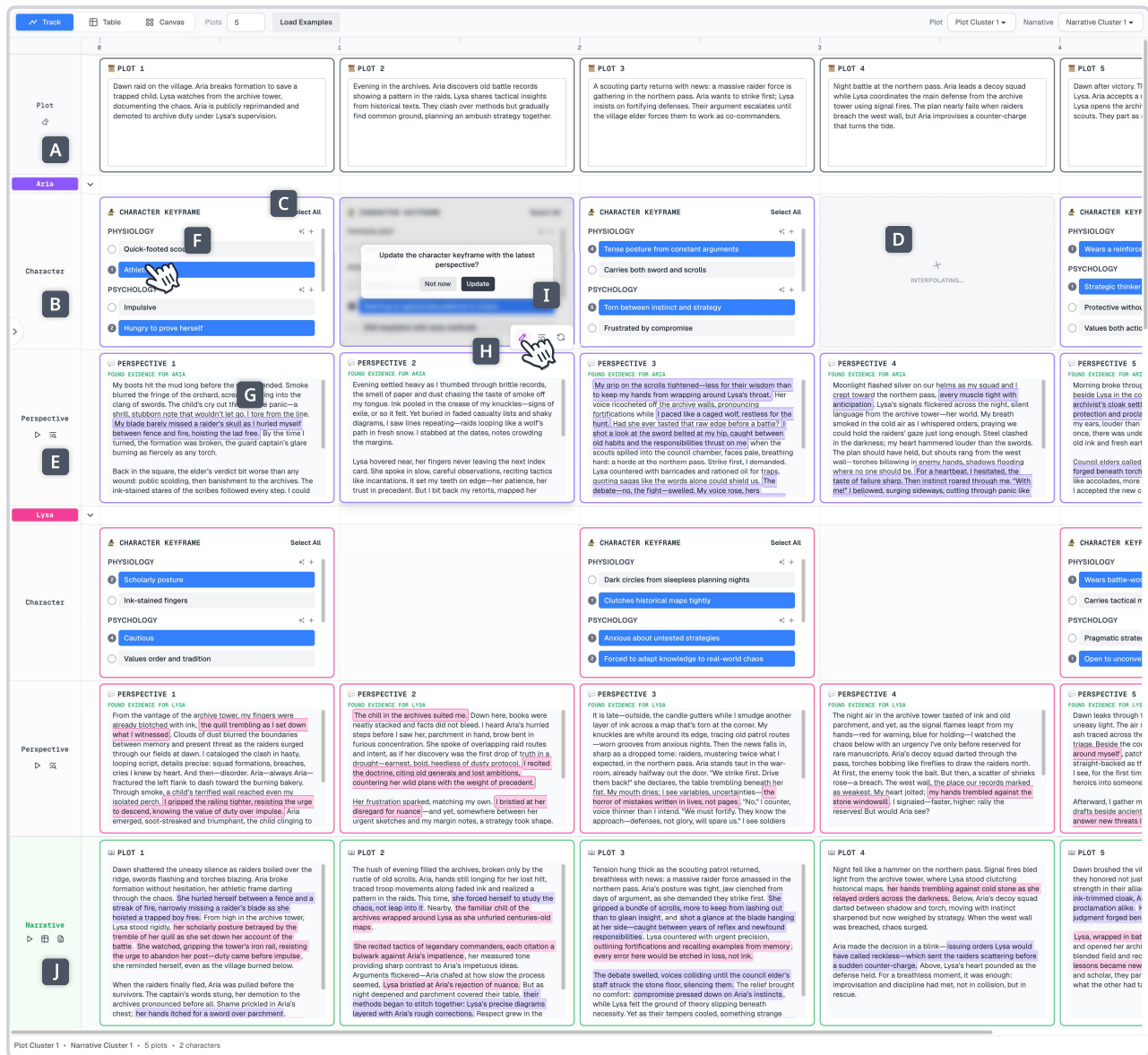


Figure 3: Track View. The Track view supports the main end-to-end workflow from story planning to generation. Writers define (A) plot keyframes, which represent the major events in a story outline. Then, they (B) specify character keyframes for certain plots, and can trigger the AI to (C) suggest new traits or (D) automatically interpolate character development between existing character keyframes. The system generates (E) first-person perspective generates based on plot and character keyframes. Selecting a specific trait (F) highlights its corresponding textual evidence (G) within the perspective keyframe. The system supports bidirectional editing; manually editing a perspective keyframe (H) prompts the system to update the corresponding character keyframe (I). Finally, writers generate a (J) third-person narrative that synthesizes the selected textual evidence from perspective keyframes, with text color-coded by character.

- The **Track** view (Fig. 3) supports the main end-to-end workflow, allowing writers to move from outline creation to character keyframes, first-person perspective exploration, and final third-person story generation.
- The **Table** view (Fig. 4) presents the outline, each character's first-person perspective, and the generated third-person narrative side by side for each plot, helping writers compare them

in parallel and trace how outline content is developed into the final narrative, as well as how details from perspectives are transformed and incorporated into the third-person narrative.

- The **Canvas** view (Fig. 5) supports broader exploration by allowing writers to create multiple character arcs for a single character and combine different arcs across characters to generate alternative versions of the final narrative.

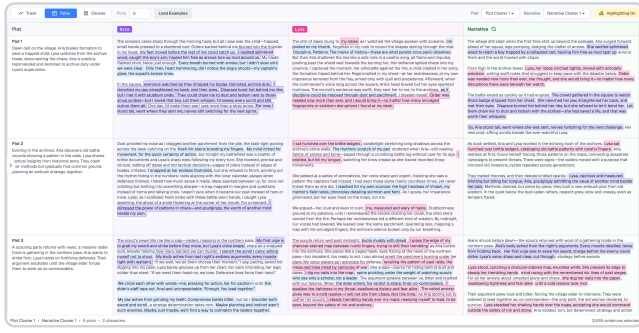


Figure 4: Table View. The Table View aligns plot keyframes, perspective keyframes, and the generated narrative on a row-by-row basis. Color-coded highlights indicate evidence selected from each character’s perspective and show where it is reflected in the final third-person narrative, supporting comparison and traceability across representations.

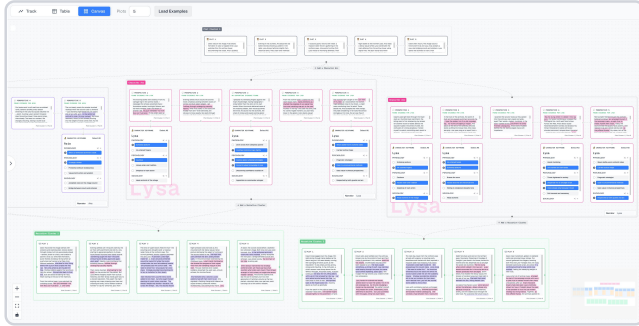


Figure 5: Canvas View. The Canvas View presents story materials as a node-link graph, making branching character arcs and their recombination explicit. The visualization helps writers compare alternative narrative paths and inspect how different character arc combinations lead to different generated stories.

5.2 Plot Keyframes

Our system begins with *plot keyframes* (Fig. 3A), which represent the major events in a story outline. Each plot keyframe acts as a structural anchor for the corresponding character keyframes, perspective keyframes, and generated narrative. This event-based representation helps writers break a story into manageable units, plan character development across events (DG1), and maintain alignment between high-level plot structure and later generated text.

5.3 Character Keyframes

To help writers control character development, our system lets them define *character keyframes* (Fig. 3B) at key plots in the story outline. Each keyframe captures the character’s state at a particular moment in the narrative and serves as a building block for shaping that character’s arc over the course of the story (DG1).

Defining Character Traits: Based on Egri’s character “bone structure” [12] and prior work [42], each keyframe organizes traits into three dimensions: *physiology* (e.g., age and appearance), *psychology*

(e.g., personality, values, and ambitions), and *sociology* (e.g., profession, status, and relationships). For each dimension, users can add their own traits via the plus icon + or ask the AI via the sparkles icon ✨ to suggest three additional traits based on the existing traits and story outline (Fig. 3C).

Interpolating Character Development: Because writers may not want to manually specify every intermediate character state, the system can automatically interpolate character keyframes (Fig. 3D) between existing ones to suggest how a character may transition over story progression (DG1). These generated keyframes remain fully editable, allowing users to revise, add, or remove traits as needed.

5.4 Perspective Keyframes

To help writers explore how a character’s traits may be expressed in narrative form, our system generates first-person *perspective keyframes* (Fig. 3E) from character keyframes. These perspectives externalize character attributes into textual evidences that can later be used to guide the generation of third-person narratives (DG2).

Generating First-Person Perspectives: After creating character keyframes for a character, users can click the play icon ▶ to generate first-person perspectives for that character (Fig. 3F). To guide generation, we incorporate Rimmon-Kenan’s typology [41] of textual indicators of character traits, including *direct definition*, *actions*, *speech*, *appearance*, and *environment*, into the prompt to guide the model to manifest the user-defined character attributes for the corresponding perspective keyframe (DG2). For example, if a user specifies the trait “self-doubting,” the generated perspective may express it through evidence such as hesitation in action (“*I paused before reaching for the door*”), self-questioning in thought (“*What if I get this wrong again?*”), or uncertainty in speech (“*I’m not sure this is a good idea*”).

Inspecting and Selecting Textual Evidence: Given a generated first-person perspective, users can click the search icon 🔍 (Fig. 3G) to ask the AI to identify textual evidence for the character traits they defined (DG2), based on Rimmon-Kenan’s typology [41]. Users can then click individual character traits (Fig. 3F) to highlight or hide the corresponding evidence in the perspective keyframes. Highlighted evidence (e.g., Fig. 3G) is selected by default for use in generating the final third-person narrative; and users can click to manually deselect any highlighted passage.

Bidirectional Editing Between Character Keyframes and Perspective Keyframes: Each character keyframe and its corresponding perspective keyframes are bidirectionally linked within an plot. Users can click the pencil icon ✎ (Fig. 3H) to edit a perspective keyframe, which triggers an update to the corresponding character keyframe (Fig. 3I). Conversely, editing a character keyframe triggers regeneration of the associated perspective keyframe.

5.5 Third-Person Narratives

After exploring characters through perspective keyframes and selecting the traits and textual evidence they want to emphasize, users can click the play icon ▶ to generate a third-person narrative (Fig. 3J). Before generation, the system opens a panel for users to review all selected textual evidence from each character’s perspective in

each plot. Once users confirm their selections, the system generates a third-person narrative based on the plots and enriches it with the selected textual evidences of character traits. In the generated narrative, passages derived from different characters’ perspectives are highlighted in different colors. This color coding helps users trace how character traits are expressed in first-person perspectives and how those materials are later synthesized into the final third-person narrative (DG3). Lastly, users can click the file icon  to populate the generated third-person narrative into a Markdown text editor, where they can further review and refine it.

5.6 Implementation Notes

The system is built with the Next.js framework, which supports server-side rendering for API calls, including calls to the OpenAI API for prompting pre-trained GPT models, and to the Firebase APIs for logging user events. We use React Flow to build the node-based canvas and Slate.js to build the text editor. We instruct GPT-4.1 to suggest character traits, interpolate character keyframes, generate first-person perspectives, extract evidence from perspectives, and generate narratives based on selected traits and evidences. Sample prompts are provided in Appendix A. The source code will be open-sourced upon publication.

6 System Evaluation

Our system instantiates narrative keyframing through plot, character, and perspective keyframes, with character development serving as the primary narrative property under writers’ control. Accordingly, our evaluation focuses on the system’s support for characterization. To this end, we evaluated the system through two studies: a technical evaluation (Study 1) of story outputs and a user study (Study 2) of writers’ experiences. The goal of Study 1 is to validate that our choice of narratologically-motivated conditioning (traits, first-person perspectives, and evidence-guided recombination) could lead to richer characterization without hurting quality relative to a common outline-to-narrative baseline. Study 2 served as our primary evaluation of user support, investigating how the system supports characterization in writing practice. Together, these studies assess both output quality and the interaction benefits of the system.

6.1 Study 1: Technical Evaluation

Before conducting the user study, we performed a technical evaluation to examine whether our approach could produce higher-quality stories than a vanilla LLM baseline.

6.1.1 Method. Here, we describe the method of our technical evaluation, including the materials and metrics.

Materials: To generate stories, we used the ten writing prompts from the CoAuthor dataset [31], supplemented by another ten prompts randomly sampled from the WritingPrompts dataset [13]. The full list of prompts is provided in Appendix B.1. For each prompt, we first instructed GPT-4.1 to generate five diverse outlines featuring two main characters and following a three-act structure [14] (setup, confrontation, and resolution). We then provided each outline as input to both our system and a vanilla prompt-based LLM baseline, asking each condition to expand the outline into a

complete story. Both conditions used the same underlying model, GPT-4.1. For our system, we enabled the automatic character trait suggestion feature to generate three traits in each category for each character at each plot. We then used these traits to generate first-person perspectives and randomly selected two pieces of evidence per character per plot to generate the final third-person story.

Metrics: We evaluated the quality of the generated stories from the two conditions using the WQRM-PRE model from Chakrabarty et al. [5], which was trained on expert preference data and has been shown to align with expert judgments. This model produces a scalar quality score for each story. We used these scores for both pairwise comparisons (between the two stories generated from the same prompt) and a paired t-test over the corpus.

To complement the automatic evaluation, we also conducted a human validation on a randomly sampled subset of 30 story pairs from the two conditions. Three independent raters with creative-writing experience recruited from Prolific compared the paired stories in randomized order, with condition identities removed. All three human raters self-reported that they were experienced professional writers. Two had 4–6 years of experience, and one had 7–10 years, including work in screenwriting and editorial evaluation. For each pair, raters indicated (1) which story they preferred overall and (2) which story exhibited richer characterization. These two dimensions were chosen to validate both the general quality signal captured by the automatic evaluator and the characterization-focused contribution of our system. Each rater was compensated with \$20.

6.1.2 Results. The results suggest that our system can produce story outputs that are rated more favorably than those from the baseline in terms of both characterization and overall story quality.

Quantitative Results: In the pairwise comparison by the model, stories generated by our system were preferred in 72 out of 100 cases. The paired t-test on the model scores also showed a significant difference in favor of our system ($M = 6.24$ vs. 5.74 , $t = 6.52$, $p < .001^{***}$). The human validation showed a similar pattern. Using majority voting across the three raters, our system’s stories were preferred for richer characterization in 96.7% of pairs (29 out of 30) and for overall quality in 83.3% of pairs (25 out of 30), with no pairs favoring the baseline under majority vote. When pooling all individual judgments ($N = 90$), raters preferred our system’s stories for characterization in 84.4% of cases and for overall quality in 73.3% of cases.

Expert Comments: In addition to evaluating the stories, experts shared the reasons for their judgments. Ratets consistently noted that stories from our approach provided “*specific, physical characterization*” through concrete behavioral details—characters whose “*fingers often trembling as she fidgeted with her necklace*,” whose “*knees itched where he’d knelt too long*,” or who arrived “*still wearing his coffee shop apron*.” Our stories also “*embodied*” emotional changes rather than “*summarizing*” them, allowing readers to feel “*how it felt*” rather than simply being “*told what they did*.” These details were seen as transforming characters from “*functional role-players*” into “*vividly specific people*,” giving them more “*texture*.” Most of these details appeared to originate from the intermediate first-person perspectives in our pipeline. However, in the minority of cases where raters preferred the baseline for overall quality,

they cited its advantages in conciseness and narrative flow. Raters described baseline stories as having “*cleaner sentence structure*” and “*better pacing that holds the tension*,” suggesting that the added characterization detail could occasionally come at the cost of readability, with our stories sometimes feeling “*bogged down by overwriting*.”

6.2 Study 2: User Evaluation

Study 1 provided preliminary evidence that our pipeline can produce promising story outputs compared with a vanilla LLM baseline. However, output quality alone does not explain how writers experience the system or whether its interaction design supports characterization during writing. We therefore conducted a within-subjects study with 12 writers of different levels of creative writing expertise to investigate how narrative keyframing supports characterization during generative creative writing. Specifically, we examined how narrative keyframing shapes writers’ experiences of **controlling**, **manifesting**, and **tracing** characterization.

6.2.1 Method. Here, we describe the method of the study, including the baseline, participants, procedure, and analysis.

Baseline: Recent HCI systems for characterization in story writing, such as CharacterChat [42] and CharacterMeet [38], use chatbots to role-play story characters in support of character construction. Similarly, our baseline included character sheets and character chatbots for defining and interacting with characters (similar to CharacterChat and CharacterMeet). In addition, to reflect common chatbot-based writing tools such as ChatGPT, our baseline also provided a story outline for plot conditioning and a story chatbot for generation and ideation. Example screenshots of the baseline are shown in Fig. 7.

Participants: We recruited 12 participants (8 female and 4 male), aged 25–65 ($M = 38.92$, $SD = 15.13$), through crowdsourcing platforms, social networks, and word of mouth. All participants reported proficiency in reading and writing in English. We recruited participants with a range of creative writing expertise: 3 identified as professional writers with published work, 4 as advanced writers (2 of whom had also published work), 1 as an intermediate writer, and 4 as beginner writers. We indicate participants’ self-reported writing expertise when quoting them in the qualitative results. Detailed information about each participant’s prior writing experience, including relevant roles, genres, projects, and publications, is provided in Appendix B.4. In addition, all participants reported prior experience using AI tools for writing. Their self-reported familiarity with using AI tools to support writing, measured on a 5-point scale (1 = none, 5 = extensive), was 3.92 ($SD = 1.08$). We complemented these self-reports with participants’ textual descriptions of their experience using AI tools for writing, including the tasks and purposes for which they had used them; these descriptions are also provided in Appendix B.4. We compensated each participant with \$20.

Procedure: The study began with informed consent¹ and a demographics questionnaire. Participants then completed two 30-minute writing sessions, each based on a different writing prompt (Appendix B.2), one with our system and one with the baseline. The order of systems and prompts was counterbalanced across participants. Each

¹The study received approval from our institution’s IRB.

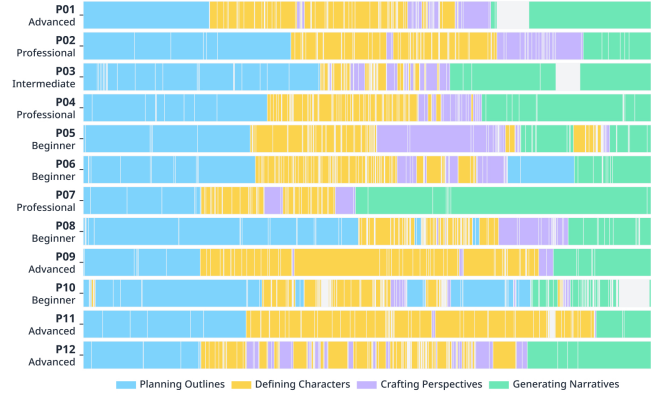


Figure 6: Overview of participants’ time distribution. Each row represents a single participant, labeled by their creative writing expertise. The x-axis tracks normalized time as a percentage of task progress. Colored segments denote specific activity categories: *Planning Outlines* (drafting the story outline), *Defining Characters* (creating character keyframes), *Crafting Perspectives* (generating perspective keyframes and selecting evidence), and *Generating Narratives* (generating, reviewing, and revising narratives). The width of each segment reflects the relative time spent on that activity. Participants completed the three-act story writing task with our system in 25.92 mins on average.

session began with a 3–5-minute tutorial covering the key features of the assigned system. Participants were then asked to create a three-act [14] outline based on the prompt, then using the assigned system to turn the outline into a story they are satisfied with. Participants were encouraged to focus on characterization. After each writing session, participants completed standardized post-condition measures, including the Creativity Support Index [6] and the AI System Experience survey [55], in 7-point Likert scale. Following both conditions, participants completed a comparative questionnaire assessing which system better supported characterization. This included questions regarding controllability over character development, manifestation of character traits in the story, and traceability between characterization work and story text. The study concluded with a 15-minute semi-structured interview to gather qualitative reflections on participants’ experiences across the two conditions (questions are listed in Appendix B.5). The entire study lasted approximately 90 minutes per participant.

Analysis: For quantitative measures in the post-condition standardized surveys, we employed the Wilcoxon signed-rank test to account for the small sample size and the non-normal distribution of the data. To analyze the exit comparative questionnaire, we conducted a one-sample Wilcoxon signed-rank test using the neutral rating (4) as the population mean following prior work [56]. For the qualitative analysis of interview transcripts, we followed established thematic analysis protocols [3, 43] to identify emerging topics. The entire research team collectively reviewed the coding outcomes to refine the high-level themes.

Table 2: Survey results of perceived experience on AI systems [55] and Creativity Support Index (CSI) [6] under two conditions. Wilcoxon signed-rank paired t-test W-values and p-values (*: $p < .05$, **: $p < .01$, ***: $p < .001$) are reported. Like previous work [34, 47], we omitted the Collaboration factor to avoid confusion, as the tasks did not involve human collaboration.

Scales	Ours		Baseline		Statistics	
	M	SD	M	SD	W	p
AI System Experience						
Match Goal	6.58	1.17	5.75	1.66	23.00	.072
Think Through	6.92	0.29	4.83	2.08	45.00	.004**
Transparent	6.42	0.79	4.00	2.17	55.00	.003**
Controllable	6.42	1.44	5.33	1.44	37.00	.047*
Creativity Support Index						
Enjoyment	6.92	0.29	5.75	1.36	28.00	.010**
Immersion	6.25	0.97	4.67	1.78	28.00	.011*
Worth Effort	6.75	0.45	6.08	1.31	15.00	.027*
Exploration	6.58	0.52	4.58	2.02	45.00	.004**
Expressiveness	6.58	0.90	5.08	1.83	45.00	.004**

6.2.2 Results. We begin with an overview of user interaction patterns derived from logged events. This is followed by quantitative results from post-condition standardized surveys. Finally, we present quantitative results from the exit comparative survey regarding the control, traceability, and manifestation of characterizations, accompanied by qualitative user comments on each aspect.

Interaction Patterns: As illustrated in Fig. 6, participants showed a consistent workflow across four primary activity phases when using our system: planning the outline, defining character arcs, generating first-person perspectives and selecting evidence, and finally generating and editing the third-person narrative. This aligns with our system’s designed workflow. We then examined how participants used our system and the baseline differently as both systems supported staged writing from outline to narrative. We found that participants followed a similar workflow (from outlines to characters to narratives) in both conditions, but differed in how they controlled characterization: participants in baseline mainly revised global character personas, while participants using our system repeatedly edited keyframes at specific plot points.

In addition, time allocation when using our system across the phases in Fig. 6 varied depending on the user’s writing expertise. Beginners dedicated the largest proportion of their time to the initial planning phase (42.94%, compared to 29.90% for Professionals and 23.04% for Advanced users). Advanced writers, in contrast, invested heavily in designing character snapshots (49.26%, compared to 27.55% for Beginners and 28.03% for Professionals). Professionals adopted a more balanced approach to early setup and spent the highest proportion of their time refining the final third-person narrative (31.29%, compared to 17.82% for Advanced users and 14.56% for Beginners).

Standardized Surveys: Table 2 shows the quantitative results from the AI System Experience and CSI surveys. Overall, our approach provides a significantly more controllable, transparent, and engaging creative experience than the baseline.

AI System Experience. Both systems were generally capable of meeting the task objectives ($M = 6.58$ vs. 5.75 , $W = 23.00$, $p = .072$). However, participants rated narrative keyframing significantly higher than the baseline in terms of system transparency and cognitive support. Specifically, the system helped users better *think through* the task ($M = 6.92$ vs. 4.83 , $W = 45.00$, $p = .004^{**}$) and was perceived as significantly more *transparent* regarding its generative processes ($M = 6.42$ vs. 4.00 , $W = 55.00$, $p = .003^{**}$). Furthermore, users felt they had significantly more *control* over the generation process when using narrative keyframing ($M = 6.42$ vs. 5.33 , $W = 37.00$, $p = .047^{*}$). These benefits of transparency and controllability are further supported by our subsequent analysis of user ratings and comments regarding the affordances of our systems over controlling, manifesting, and tracing characterization.

Creativity Support Index. The results from the CSI indicate that narrative keyframing provided better support for generative creative writing workflows compared to the baseline. Users reported significantly higher levels of *enjoyment* ($M = 6.92$ vs. 5.75 , $W = 28.00$, $p = .010^{**}$) and felt more *immersed* in the activity ($M = 6.25$ vs. 4.67 , $W = 28.00$, $p = .011^{*}$). The system was also perceived as more *worth the effort* required ($M = 6.75$ vs. 6.08 , $W = 15.00$, $p = .027^{*}$). Crucially for creative tasks, our system scored significantly higher in *exploration* ($M = 6.58$ vs. 4.58 , $W = 45.00$, $p = .004^{**}$) and *expressiveness* ($M = 6.58$ vs. 5.08 , $W = 45.00$, $p = .004^{**}$), suggesting that the our system allowed users to better explore the design space of characterization and express their creative intent in storytelling.

Controlling Characterization: We found that participants perceived narrative keyframing as offering stronger support for controlling characterization than the baseline. In the comparative survey, participants reported that it better supported their control over how each character’s perspective was reflected in the final story ($M = 6.00$, $SD = 1.41$, $V = 73.50$, $p = .003^{**}$), as well as their ability to deliberately shape each character’s portrayal at different points in the story ($M = 6.42$, $SD = 0.79$, $V = 78.00$, $p < .001^{***}$).

The Keyframed, Staged Workflow Enhances Controllability. Participants said that narrative keyframing gave them a stronger sense of control by breaking characterization into a keyframed, staged workflow. Rather than relying on a single conversational thread, they could define character keyframes, inspect first-person narratives, and then select what should carry forward into the final story. This made the relationship between their inputs and the generated output feel more direct and predictable. P01 (Advanced) noted that “*your inputs had a direct bearing on the outcome*” and appreciated being able to “*emphasize what you wanted... at each step of the process*” and “*tweak it exactly how you want it to.*” Participants contrasted this with the baseline, where control depended more on prompting and revising through chat. As P06 (Beginner) put it, “*Because everything is clearly separated and structured, ... the control of them is more direct.*”

Character Keyframes Make Character Arcs Explicit. A major source of perceived control was the ability to define character states separately across plots. Participants said that the keyframing structure made character arcs explicit and editable, allowing them to shape not only who a character was, but how that character changed

over plots. P07 (Professional) described using the system to create a stronger arc: *“I was able to see... how Monifa’s character... she was, like, this passive character, but then in the third act, she got her backbone.”* P02 (Professional) similarly appreciated being able to change a character internally, *“like, changing from selfish to selfless.”* In contrast, participants noted that the baseline largely maintained a single persona unless they manually re-specified it. As P10 (Beginner) explained, *“you only have one version of the persona.”* More broadly, participants valued being able to see and design *“the journey of the characters’ traits through the acts”* (P04, Professional). Some also found the interpolation feature useful for scaffolding transitions between character states. For example, P10 (Beginner) said that when they knew where a character should begin and end, the interpolated middle state *“really speeds up the scaffolding.”*

Providing Fine-Grained Control Through Selection and Emphasis. Participants also experienced control through the system’s selection mechanisms. After generating first-person perspectives, they could choose which traits and evidence to emphasize in the final third-person story. This let them move beyond simply accepting or rejecting generated text, and instead curate what aspects of characterization should carry forward. P02 (Professional) emphasized the flexibility of keeping, discarding, and highlighting AI-generated character features: *“Because I can keep it but not use it, I can discard it so it’s gone entirely, and then I can highlight one or two for each one.”* For several participants, this emphasis mechanism functioned as a concrete lever of agency. As P08 (Beginner) explained, *“I can select which evidence or characteristics, I don’t want to emphasize, or I want to emphasize. So... that give me more of control and agency.”* P04 (Professional) summarized this difference succinctly: *“The huge difference is having the ability to choose which aspects of the character to focus on in the story.”* In this sense, narrative keyframing supported control both by helping participants define characters, and by helping them decide what should matter most in the final narrative.

Manifesting Characterization: We found that participants preferred narrative keyframing over the baseline for manifesting characterization in the story. In the comparative survey, participants rated narrative keyframing significantly higher both for helping them incorporate character details that enriched characterization in the story ($M = 5.50, SD = 1.93, V = 66.00, p = .016^*$) and for helping them translate abstract character ideas into concrete story details in the narrative ($M = 5.67, SD = 1.16, V = 55.00, p = .003^{**}$).

Perspective Keyframes Help Concretize Characterization. Participants valued first-person perspectives as an intermediate representation that made abstract character ideas more concrete before final story generation. By inspecting how the AI rendered a character’s personality, motivations, and emotions through that character’s own voice, they could assess whether the intended characterization had been realized and refine mismatches when needed. As P06 (Beginner) explained, the first-person perspective *“gives me a good window about how the AI understands my description of their personality. So if I see some mismatch, I can then refine my character properties. It’s a good iteration process to both help me refine my thoughts and help me refine AI’s thoughts.”* Participants also appreciated that these perspectives manifested traits through concrete narrative forms such as description, dialogue, and emotion.

P02 (Professional) said, *“I really like seeing how a character trait manifested itself in narrative, or dialogue, or expressed emotions of a character in any scene,”* while P07 (Professional) noted that the system did *“a really good job”* rendering characterization through sensory details such as *“her heels clacking.”* Together, these first-person narratives made characterization feel less like a list of trait words and more like lived, narratively grounded behavior.

Perspective Keyframes Provide Reusable Materials for Story Generation. Participants also described first-person perspectives as a rich pool of material that they could draw from when composing the final story. Instead of asking the AI to directly generate a third-person narrative from sparse character descriptions, they could first generate a fuller first-person account and then selectively carry forward the parts they wanted to emphasize. P05 (Beginner) explained, *“I think it’s better to first write a complete and exhaustive first-person perspective, such that you can first ensure that what you are putting down is actually reflecting what you are planning with the character.”* The same participant later described these outputs as *“kind of a cast of assets I can use in the final story,”* from which they could choose passages to reflect what they truly wanted to emphasize. P01 (Advanced) similarly noted that these first-person perspectives were especially useful as a starting point for third-person omniscient writing, because they provided a base that participants could adapt and build on with the voices of different characters.

Tracing Characterization: We found that participants preferred narrative keyframing over the baseline for tracing characterization in the story. In the comparative survey, participants rated it significantly higher both for supporting their ability to trace how specific character traits were reflected in the final story ($M = 6.25, SD = 1.06, V = 66.00, p = .001^{**}$) and for helping them identify which parts of the final story expressed the character attributes they intended ($M = 6.17, SD = 1.40, V = 64.50, p = .002^{**}$).

Visible Mappings Help Writers Trace Traits into Text. Participants valued narrative keyframing’s explicit mappings between character traits, first-person perspectives, and final story text. Instead of inferring whether a trait had been reflected in the generated writing, they could directly inspect the connection through highlights and visual links. P02 (Professional) described this as a *“one-to-one correspondence,”* explaining, *“Whatever I had selected from the AI auto output, or whatever I typed in, it connected those two directly.”* This visibility also reduced the burden of manually searching for evidence in the text. As P06 (Beginner) noted, the system made it easier to verify that intended traits were actually reflected in the generated paragraphs, rather than having to find that evidence independently. Participants also appreciated visual encodings such as color, which made traceability easier to perceive at a glance. P12 (Advanced) remarked, *“I think humans respond well to color... it helps with the transparency.”* Together, these features made characterization traceable both conceptually and perceptually through the interface itself.

Traceability Supports Reflection on Writing Decisions. Traceability also helped participants reflect on how their characterization decisions shaped the generated story. Participants described gaining a clearer understanding of how the system transformed their inputs into narrative language, which in turn made the writing process feel more meaningful and intentional. P02 (Professional) explained,

“I’m able to understand, oh, there’s a one-to-one relationship between the specific things I put for the character... And then that is specifically output in the story, and the software allows me to visually see those two together, and that allows my brain to make the connection between my chosen description and the language model’s manipulation of that into a narrative that matches all the stuff that I typed.” This visibility appeared to support a more reflective and deliberate mode of authorship. P02 contrasted the experience with less structured forms of writing, noting, *“I think writers who just start writing and just do stream of consciousness don’t ever understand why they’re putting anything down. Whereas this allowed me to see that.”* The same participant later connected this traceability to a stronger sense of human contribution: *“I did so much high-level thinking, making decisions, which is what us humans are supposed to do.”* This visibility supported a more reflective and deliberate mode of authorship, helping participants see themselves as actively shaping characterization rather than merely reacting to generated text.

7 Discussion

7.1 Design Implications

7.1.1 Characterization as an Evolving Creative Process. Prior work often treats characterization as something specified upfront, for example through character descriptions, personas, or profiles that are then used to guide later generation stages [35, 36, 38, 42]. Our findings suggest that characterization is better understood as an evolving creative process that develops in at least two directions: horizontally across story plots, as characters change over events, and vertically across perspectives, as those changes are explored through different points of view. Rather than collapsing characterization into a single prompt or profile, AI writing systems can better support this process by exposing intermediate representations that help writers plan, inspect, and refine character development before generating final prose.

7.1.2 Perspective as a Design Material for Writing Tools. Prior work has used role-play to let AI enact user-defined characters and help writers refine character profiles before writing [16, 38, 42, 49, 53]. Our findings suggest a broader role for perspective in AI writing tools. In our system, first-person perspectives served as an intermediate representation that helped writers concretize abstract character ideas, inspect how traits might be expressed in language, and reason across multiple characters’ inner lives. This suggests that perspective is not only useful for conversational exploration, but can also function as a reusable design material that bridges character planning and story generation.

7.1.3 Traceability Matters for Generative Creative Writing. Our findings suggest that traceability is an important property of AI writing tools. By making visible how selected traits and evidence were reflected in generated text, our system helped writers inspect whether their intentions were carried into the final story and understand how character-related inputs were transformed across stages of the workflow. This was valuable for both verification and reflection: participants used these mappings to check what had been emphasized, compare generated text with their intentions, and refine characterization decisions accordingly. This aligns with prior work showing the value of provenance for transparent AI-assisted

writing [22, 46]. More broadly, this suggests that AI writing tools should make connections between writer-authored materials and generated outputs explicit, especially when supporting creative work that unfolds through multiple stages.

7.2 Limitations and Future Work

Our system has several limitations. First, our system currently supports a relatively structured workflow. It may feel restrictive for writers who prefer to develop stories in a more fluid or improvisational way. Future work could explore more flexible interactions that allow writers to move between planning, character exploration, and drafting in a less linear manner. Additionally, although keyframing and first-person perspective keyframes can extend to longer stories, our current design does not yet cleanly support highly non-linear plots, such as second-act flashbacks, which are interesting directions for future work. Lastly, our current instantiation of narrative keyframing focuses primarily on characterization through character and perspective keyframes. Future work could explore keyframing other evolving narrative properties, such as tone, pacing, or scenes, as well as interactions for coordinating multiple property tracks within the same story.

Our study also has several limitations that future work could address. First, the study included 12 participants, which is a relatively small sample size. Although we conducted statistical tests, we do not treat these results as conclusive; instead, they should be interpreted as promising but preliminary. Second, due to time constraints, the writing task asked participants to produce a short story using a three-act structure. Future work should examine how our system supports longer-form story writing. Third, our evaluation focused on a controlled writing task with two prompts and a comparison against one chatbot-based baseline. Future work could test the system with a broader range of writing goals and baseline conditions.

8 Conclusion

In this paper, we introduce narrative keyframing, a new interaction design for generative creative writing that uses high-level representations of plot events, character arcs, and narrative perspective to help writers plan and guide story generation. We instantiate this design in an interactive system that connects high-level story planning to final narrative generation through plot, character, and perspective keyframes, supporting writers in shaping character arcs, exploring first-person expression, and guiding third-person storytelling. Through a technical evaluation and a user study, we find that our approach produces stories with higher overall quality and richer characterization, while also supporting a more controllable, transparent, and engaging writing experience than a baseline system that reflects the current design of generative creative writing tools. More broadly, this work shows how keyframing can serve as an interaction paradigm for human-AI co-writing by helping writers balance automation with creative control.

References

- [1] Mieke Bal. 2004. Narration and Focalization. *Narrative theory: Critical concepts in literary and cultural studies* 1 (2004), 263–296.
- [2] Mieke Bal. 2004. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, Scholarly Publishing Division, Toronto.

- [3] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- [4] Marc Cavazza, Fred Charles, and Steven J. Mead. 2001. Characters in Search of an Author: AI-Based Virtual Storytelling. In *Virtual Storytelling Using Virtual Reality Technologies for Storytelling*, Olivier Balet, Gérard Subsol, and Patrice Torguet (Eds.). Springer, Berlin, Heidelberg, 145–154. https://doi.org/10.1007/3-540-45420-9_16
- [5] Tuhin Chakrabarty, Philippe Laban, and Chien-Sheng Wu. 2025. AI-Slop to AI-Polish? Aligning Language Models through Edit-Based Writing Rewards and Test-Time Computation. <https://doi.org/10.48550/arXiv.2504.07532> [cs]
- [6] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4 (Aug. 2014), 1–25. <https://doi.org/10.1145/2617588>
- [7] Jean-Peic Chou, Alexa Fay Siu, Nedim Lipka, Ryan Rossi, Franck Dernoncourt, and Maneesh Agrawala. 2023. TaleStream: Supporting Story Ideation with Trope Knowledge. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. ACM, San Francisco CA USA, 1–12. <https://doi.org/10.1145/3586183.3606807>
- [8] 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 (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. <https://doi.org/10.1145/3491102.3501819>
- [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 (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. <https://doi.org/10.1145/3491102.3501819>
- [10] John Joon Young Chung and Max Kreminski. 2024. Patchview: LLM-Powered Worldbuilding with Generative Dust and Magnet Visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. <https://doi.org/10.1145/3654777.3676352>
- [11] Paramveer S. Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel P. Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-Writing with Language Models. [arXiv:2402.11723](https://arxiv.org/abs/2402.11723) [cs]
- [12] Lajos Egri. 1995. *The Art Of Dramatic Writing: Its Basis in the Creative Interpretation of Human Motives*. Touchstone, New York, NY (u.a.).
- [13] Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical Neural Story Generation. <https://doi.org/10.48550/arXiv.1805.04833> [arXiv:1805.04833](https://arxiv.org/abs/1805.04833) [cs]
- [14] Syd Field. 2005. *Screenplay: The Foundations of Screenwriting*. Delta, New York.
- [15] Edward Morgan Forster. 1927. *Aspects of the Novel*. Harcourt, Brace, New York.
- [16] Jiaying Fu, Xiruo Wang, Kate Vi, Zhouyi Li, Chuyan Xu, and Yuqian Sun. 2025. "I Like Your Story!": A Co-Creative Story-Crafting Game with a Persona-Driven Character Based on Generative AI. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–5. <https://doi.org/10.1145/3706599.3721163>
- [17] James Garvey. 1978. Characterization in Narrative. *Poetics* 7, 1 (1978), 63–78.
- [18] Gerard Genette and Jonathan Culler. 1990. *Narrative Discourse: An Essay in Method*. Cornell University Press, Ithaca.
- [19] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing Using Language Models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1002–1019. <https://doi.org/10.1145/3532106.3533533>
- [20] Jack Grapes. 2017. *Method Writing: The First Four Concepts*. Bombshelter Press, Los Angeles, CA.
- [21] William John Harvey. 1968. *Character and the Novel*. Cornell University Press, Ithaca.
- [22] Md Naimul Hoque, Tasfia Mashiat, Bhavya Ghai, Cecilia D. Shelton, Fanny Chevalier, Kari Kraus, and Niklas Elmquist. 2024. The HaLLMark Effect: Supporting Provenance and Transparent Use of Large Language Models in Writing with Interactive Visualization. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3613904.3641895>
- [23] Chieh-Yang Huang, Shih-Hong Huang, and Ting-Hao Kenneth Huang. 2020. Heteroglossia: In-Situ Story Ideation with the Crowd. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376715>
- [24] Takumi Ito, Naomi Yamashita, Tatsuki Kuribayashi, Masatoshi Hidaka, Jun Suzuki, Ge Gao, Jack Jamieson, and Kentaro Inui. 2023. Use of an AI-Powered Rewriting Support Software in Context with Other Tools: A Study of Non-Native English Speakers. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3586183.3606810>
- [25] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. <https://doi.org/10.1145/3544548.3581196>
- [26] Jodileghorn. 2010. Writing Exercise: Switching Points of View.
- [27] Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S. Bernstein. 2017. Mechanical Novel: Crowdsourcing Complex Work through Reflection and Revision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 233–245. <https://doi.org/10.1145/2998181.2998196>
- [28] Taewook Kim, Hyomin Han, Eytan Adar, Matthew Kay, and John Joon Young Chung. 2024. Authors' Values and Attitudes Towards AI-Bridged Scalable Personalization of Creative Language Arts. <https://doi.org/10.1145/3613904.3642529> [arXiv:2403.00439](https://arxiv.org/abs/2403.00439) [cs]
- [29] John Lasseter. 1987. Principles of traditional animation applied to 3D computer animation. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '87)*. Association for Computing Machinery, New York, NY, USA, 35–44. <https://doi.org/10.1145/37401.37407>
- [30] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsgans, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L. C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Agnia Sergeyuk, Antonette Shibani, Disha Shrivastava, Lila Shroff, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia H. Rho, Shannon Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. <https://doi.org/10.1145/3613904.3642697> [arXiv:2403.14117](https://arxiv.org/abs/2403.14117) [cs]
- [31] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. <https://doi.org/10.1145/3491102.3502030>
- [32] Yoonjoo Lee, Tae Soo Kim, Minsuk Chang, and Juho Kim. 2022. Interactive Children's Story Rewriting Through Parent-Children Interaction. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. Association for Computational Linguistics, Dublin, Ireland, 62–71. <https://doi.org/10.18653/v1/2022.in2writing-1.9>
- [33] Damien Masson, Young-Ho Kim, and Fanny Chevalier. 2025. Textoshop: Interactions Inspired by Drawing Software to Facilitate Text Editing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. <https://doi.org/10.1145/3706598.3713862>
- [34] Damien Masson, Zixin Zhao, and Fanny Chevalier. 2025. Visual Story-Writing: Writing by Manipulating Visual Representations of Stories. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3746059.3747758>
- [35] 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 (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–34. <https://doi.org/10.1145/3544548.3581225>
- [36] Kyeongman Park, Minbeom Kim, and Kyomin Jung. 2025. A Character-Centric Creative Story Generation via Imagination. In *Findings of the Association for Computational Linguistics: ACL 2025*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 1598–1645. <https://doi.org/10.18653/v1/2025.findings-acl.82>
- [37] Syemin Park, Soobin Park, and Youn-kyung Lim. 2025. Constella: Supporting Storywriters' Interconnected Character Creation through LLM-Based Multi-Agents. <https://doi.org/10.48550/arXiv.2507.05820> [arXiv:2507.05820](https://arxiv.org/abs/2507.05820) [cs]
- [38] Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–19. <https://doi.org/10.1145/3613904.3642105>
- [39] Mohi Reza, Nathan Laundry, Ilya Musabirov, Peter Dushniku, Zhi Yuan "Michael" Yu, Kashish Mittal, Tovi Grossman, Michael Liut, Anastasia Kuzminykh, and Joseph Jay Williams. 2023. ABScribe: Rapid Exploration of Multiple Writing Variations in Human-AI Co-Writing Tasks Using Large Language Models. <https://doi.org/10.48550/arXiv.2310.00117> [arXiv:2310.00117](https://arxiv.org/abs/2310.00117) [cs]
- [40] Mark O. Riedl. 2008. Vignette-Based Story Planning: Creativity through Exploration and Retrieval. In *Proceedings of the 5th International Joint Workshop on Computational Creativity*. Association for Computational Creativity, Madrid, Spain, 41–50.
- [41] Shlomith Rimmon-Kenan. 2003. *Narrative Fiction: Contemporary Poetics*. Routledge, London.
- [42] Oliver Schmitt and Daniel Buschek. 2021. CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation

- with a Chatbot. In *Creativity and Cognition*. ACM, Virtual Event Italy, 1–10. <https://doi.org/10.1145/3450741.3465253>
- [43] Raymond Scupin. 1997. The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology. *Human Organization* 56, 2 (1997), 233–237. <https://doi.org/10.1145/3584931.3607492>
- [44] Hua Shen, Chieh-Yang Huang, Tongshuang Wu, and Ting-Hao Kenneth Huang. 2023. ConvXAI : Delivering Heterogeneous AI Explanations via Conversations to Support Human-AI Scientific Writing. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 384–387. <https://doi.org/10.1145/3584931.3607492>
- [45] Jocelyn Shen, Nicolai Marquardt, Hugo Romat, Ken Hinckley, Nathalie Riche, and Fanny Chevalier. 2026. Textual: A Text-as-Material Interaction Paradigm for LLM-Mediated Writing. <https://doi.org/10.1145/3772318.3790330> arXiv:2603.00452 [cs]
- [46] Momin N. Siddiqui, Nikki Nasser, Adam Coscia, Roy Pea, and Hari Subramonyam. 2025. DraftMarks: Enhancing Transparency in Human-AI Co-Writing Through Interactive Skeuomorphic Process Traces. <https://doi.org/10.48550/arXiv.2509.23505> arXiv:2509.23505 [cs]
- [47] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–26. <https://doi.org/10.1145/3613904.3642400>
- [48] Lu Sun, Stone Tao, Junjie Hu, and Steven P. Dow. 2024. MetaWriter: Exploring the Potential and Perils of AI Writing Support in Scientific Peer Review. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 94:1–94:32. <https://doi.org/10.1145/3637371>
- [49] Yuqian Sun, Xingyu Li, Shunyu Yao, Noura Howell, Tristan Braud, Chang Hee Lee, and Ali Asadipour. 2025. ORIBA: Exploring LLM-Driven Role-Play Chatbot as a Creativity Support Tool for Original Character Artists. <https://doi.org/10.48550/arXiv.2512.12630> arXiv:2512.12630 [cs]
- [50] Selen Türkay, Daniel Seaton, and Andrew M. Ang. 2018. Itero: A Revision History Analytics Tool for Exploring Writing Behavior and Reflection. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188474>
- [51] Lina Varotsi. 2019. *Conceptualisation and Exposition: A Theory of Character Construction*. Routledge, New York. <https://doi.org/10.4324/9780429060762>
- [52] Qian Wan, Jiannan Li, Huanchen Wang, and Zhicong Lu. 2025. Polymind: Parallel Visual Diagramming with Large Language Models to Support Prewriting Through Microtasks. <https://doi.org/10.48550/arXiv.2502.09577> arXiv:2502.09577 [cs]
- [53] Yi Wang, Qian Zhou, and David Ledo. 2024. StoryVerse: Towards Co-Authoring Dynamic Plot with LLM-Based Character Simulation via Narrative Planning. In *Proceedings of the 19th International Conference on the Foundations of Digital Games (FDG '24)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3649921.3656987>
- [54] K. M. Weiland. 2016. *Creating Character Arcs: The Masterful Author's Guide to Uniting Story Structure*. PenForASword, London.
- [55] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–22. <https://doi.org/10.1145/3491102.3517582>
- [56] Yu-Chun Grace Yen, Jane L. E. Hyoungwook Jin, Mingyi Li, Grace Lin, Isabelle Yan Pan, and Steven P. Dow. 2024. ProcessGallery: Contrasting Early and Late Iterations for Design Principle Learning. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 112 (April 2024), 35 pages. <https://doi.org/10.1145/3637389>
- [57] 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 (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 841–852. <https://doi.org/10.1145/3490099.3511105>
- [58] Chao Zhang, Shunan Guo, Abe Davis, and Eunye Koh. 2026. Narrix: Remixing Narrative Strategies from Examples for Story Writing. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM, Barcelona Spain, 1–24. <https://doi.org/10.1145/3772318.3790813>
- [59] Chao Zhang, Kexin Ju, Peter Bidoshi, Yu-Chun Grace Yen, and Jeffrey M. Rzeszotarski. 2025. Friction: Deciphering Writing Feedback into Writing Revisions through LLM-Assisted Reflection. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–27. <https://doi.org/10.1145/3706598.3714316>
- [60] Chao Zhang, Kexin Ju, Zhuolun Han, Yu-Chun Grace Yen, and Jeffrey M. Rzeszotarski. 2025. Synthia: Visually Interpreting and Synthesizing Feedback for Writing Revision. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3746059.3747703>
- [61] Chao Zhang, Xuechen Liu, Katherine Ziska, Soobin Jeon, Chi-Lin Yu, and Ying Xu. 2024. Mathemyths: Leveraging Large Language Models to Teach Mathematical Language through Child-AI Co-Creative Storytelling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–23. <https://doi.org/10.1145/3613904.3642647>
- [62] Chao Zhang, Yiren Liu, Lunyu Nie, Jeffrey M. Rzeszotarski, Yun Huang, and Tal August. 2026. From Words to Widgets for Controllable LLM Generation. <https://doi.org/10.48550/arXiv.2604.10925> arXiv:2604.10925 [cs.HC]
- [63] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. 2022. StoryDrawer: A Child-AI Collaborative Drawing System to Support Children's Creative Visual Storytelling. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3491102.3501914>
- [64] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–30. <https://doi.org/10.1145/3586183.3606800>

A Implementation Details

In this section, we present the prompts used to instruct GPTs to suggest character traits, interpolate character keyframes, generate first-person perspectives, extract evidence from perspectives, and generate narratives based on selected traits and evidences.

A.1 Suggesting Physiology Traits

Physiology:
You are a story development assistant.

Full story: `<full_outline>`
Targeting plot: `<current_plot>`
Character: `<character_name>`
Existing physiology traits: `<existing_traits>`

Brainstorm three concise physiology traits for this targeting plot in the story. These traits should sharpen characterization and help guide revisions to improve the story.

Guidelines:

- Focus on Physical appearance, clothing, body language, visible characteristics.
- Keep each trait 3-8 words
- Ground traits in the story plot and the full story
- Avoid repeating existing traits or near-duplicates
- Return exactly three traits

Return JSON that matches the provided schema.

A.2 Suggesting Psychology Traits

Psychology:
You are a story development assistant.

Full story: `<full_outline>`
Targeting plot: `<current_plot>`
Character: `<character_name>`
Existing psychology traits: `<existing_traits>`

Brainstorm three concise psychology traits for this targeting plot in the story. These traits should sharpen characterization and help guide revisions to improve the story.

Guidelines:

- Focus on Emotions, motivations, thoughts, beliefs, mental state.
- Keep each trait 3-8 words
- Ground traits in the story plot and the full story
- Avoid repeating existing traits or near-duplicates
- Return exactly three traits

Return JSON that matches the provided schema.

A.3 Suggesting Sociology Traits

Sociology:
You are a story development assistant.

Full story: `<full_outline>`
Targeting plot: `<current_plot>`
Character: `<character_name>`
Existing sociology traits: `<existing_traits>`

Brainstorm three concise sociology traits for this targeting plot in the story. These traits should sharpen characterization and help guide revisions to improve the story.

Guidelines:

- Focus on Social roles, relationships, status, interactions with others.
- Keep each trait 3-8 words
- Ground traits in the story plot and the full story
- Avoid repeating existing traits or near-duplicates
- Return exactly three traits

Return JSON that matches the provided schema.

A.4 Interpolating Character Keyframes

You are analyzing a character's narration to extract their traits at this specific plot in the story.

Narration of this plot by `<character_name>` :
`<perspective_of_this_plot>`

If nearby snapshots exist:
Character keyframes from nearby plots:
`<Previous|Following>` keyframe of `<keyframe_name>` :
Physiology: `<trait1, trait2, ...>`
Psychology: `<trait1, trait2, ...>`
Sociology: `<trait1, trait2, ...>`

If first-person perspective is available:
Full narration (context only – do NOT quote from this section):
`<full_perspective_text>`

Based on the narration text `< and nearby keyframes>`, extract the character traits for `<character_name>` at this plot.

Guidelines:

- Physiology: Physical appearance, clothing, body language, visible characteristics
- Psychology: Emotions, motivations, thoughts, beliefs, mental state
- Sociology: Social roles, relationships, status, interactions with others
- Consider the character's development trajectory from nearby keyframes (if any)
- Only include traits that are evident or strongly implied in the text
- Keep trait descriptions concise (3-8 words each)
- Return 2-5 traits per category when evident

Evidence requirements:

- For every trait you include, add one entry to traitEvidence with the traitCategory, the exact trait wording, and an evidenceText.
- Each evidenceText must be a verbatim quote from the narration above (do NOT pull from the context section).
- If you cannot find a direct quote, omit the trait entirely.

Return JSON that matches the provided schema.

A.5 Generating First-Person Perspectives

You are a story writer. Your job is to write a first-person narration for a given story plot from the perspective of a specified character.

Requirements:

- Stay faithful to the facts, chronology, and causality in the Targeting plot and the Full story.
- Do not contradict established details (names, places, outcomes, revealed secrets, injuries, timelines, motivations already shown).
- Do not add new major plot events; you may add small, plausible sensory details and moment-to-moment actions that do not change the plot's outcome.
- Keep tense and POV consistent: first-person ("I", "me", "my").
- Show emotions and thinking through actions, speech, appearance, environment, and specific observations (avoid generic statements like "I was scared" unless grounded in concrete detail).
- Match the tone and genre implied by the Full story.
- Strict length limit: Maximum 200 words.

Character voice rules:

- If character traits are supplied, demonstrate those traits through concrete choices in diction, focus, and interpretation (what they notice, what they ignore, how they justify things).
- If no traits are provided, infer a character-consistent voice from the Full story and Targeting plot.

Return each result as a JSON object that satisfies the provided schema.

Full story:

```
<plot[0]>
<plot[1]>
<...>
```

Targeting plot:

```
<targeting_plot>
```

Narrator: <character_name>

If a character keyframe with traits exist for this plot:

Character traits:

- Physiology: <trait1, trait2, ...>
- Psychology: <trait1, trait2, ...>
- Sociology: <trait1, trait2, ...>

If no character keyframe for this plot:

Character traits: (none provided)

If custom prompt is provided:

ADDITIONAL INSTRUCTIONS FROM USER:

```
<custom_prompt>
```

A.6 Extracting Evidence from Perspectives

You are an expert literary analyst. Identify direct textual evidence (i.e., verbatim phrases) that confirms the given character traits.

Full story (background only—do NOT quote from this section):

```
<group_context>
```

Current plot (ONLY source for evidence):

```
<reflection>
```

Characters and traits to verify:

```
<character_name>
  Physiology:
    - <trait_value>
    - <trait_value>
  Psychology:
    - <trait_value>
  Sociology:
    - <trait_value>
```

Evidence categories to classify each phrase:

- directDefinition: Explicit direct statements or labels about the character
- actions: Physical actions, behaviors, or body language
- speech: What the character says, how they speak, or how other characters say about them
- appearance: Visual descriptions of the character
- environment: Surroundings, context, or setting that characterizes the person

Instructions:

1. Scan the current snippet for exact short phrases that directly or indirectly demonstrate each listed trait.
2. Only report evidence that appears verbatim in the current snippet text.
3. When one phrase supports multiple traits from the same category, list all matching traits together.
4. Assign each phrase to exactly one evidence category from the list above.
5. Return characterEvidence entries in the same order as the character list above.
6. Return JSON that matches the provided schema exactly. Do not include explanations outside the schema.

A.7 Generating Third-Person Narratives

Generating Third-Person Narratives:

You are a narrative writer. Expand the provided story outline into a third-person story.

Story outlines:

```
<plot[0]>
<plot[1]>
<...>
```

Main characters:

```
<character_list>
```

Instructions:

- Write a cohesive full story that follows the outline exactly
- Use third-person narration
- Include both main characters throughout
- Maintain chronological order and clear act progression
- Return the story per act, in order, with 1-2 paragraphs per act
- Each act entry should include the act number, the act label from the outline, and the act text

Return JSON that matches the provided schema.

Enriching with Selected Traits and Evidence:

You are a narrative editor. Your job is to make the original story read **better** by seamlessly integrating the selected details.

Original story + selected details:

Plot 1: `<plot_description>`

- Character: `<character_name>`

- Traits: `<trait1, trait2, ...>`

If snippets exist for this plot:

Selected details:

1. `"<evidence_from_perspectives>"`
2. `"<evidence_from_perspectives>"`

If no snippets:

Selected details: (none)

Plot 2: `<plot_description>`

`<...>`

Requirements:

- Preserve the original plot, beat order, and third-person narration.
- Do NOT add new events, attempts, or outcomes beyond what the original story already includes.
- Integrate details naturally (avoid "laundry lists" of descriptions).
- Avoid overwriting: keep sentences clear and varied in length; do not let any one sentence run on too long.
- Maintain continuity (names, timelines, locations, and cause-and-effect must remain consistent).

Snippet usage tracking:

For each plot with selected character details, output "snippetUsages" as pairs of:

- originalSnippet: exact text from selected details (first-person)
 - verbatimInNarrative: an EXACT substring from your third-person narrative showing your transformation (≤ 25 words unless impossible)
- For events without details, snippetUsages must be an empty array.

Output:

- Return JSON matching the provided schema.

If custom prompt is provided:

ADDITIONAL INSTRUCTIONS FROM USER:

`<custom_prompt>`

B Evaluation Details

B.1 Writing Prompts in Technical Evaluation

To generate stories, we used the ten writing prompts from the CoAuthor dataset [31], supplemented by another ten prompts randomly sampled from the WritingPrompts dataset [13]. The full list of the 20 writing prompts are shown below:

- (1) Once upon a time there was an old mother pig who had one hundred little pigs and not enough food to feed them. So when they were old enough, she sent them out into the world to seek their fortunes. You know the story about the first three little pigs. This is a story about the 92nd little pig. The 92nd little pig built a house out of depleted uranium. And the wolf was like, "dude."
- (2) A woman has been dating guy after guy, but it never seems to work out. She's unaware that she's actually been dating

the same guy over and over; a shapeshifter who's fallen for her, and is certain he's going to get it right this time.

- (3) When you die, you appear in a cinema with a number of other people who look like you. You find out that they are your previous reincarnations, and soon you all begin watching your next life on the big screen.
- (4) Humans once wielded formidable magical power. But with over 7 billion of us on the planet now, mana has spread far too thinly to have any effect. When hostile aliens reduce humanity to a mere fraction, the survivors discover an old power has begun to reawaken once again.
- (5) An alien has kidnapped Matt Damon, not knowing what lengths humanity goes through to retrieve him whenever he goes missing.
- (6) You're Barack Obama. Four years into your retirement, you awake to find a letter with no return address on your bedside table. It reads, "I hope you've had a chance to relax, Barack... but pack your bags and call the number below. It's time to start the real job." Signed simply, "JFK."
- (7) Following World War III, all the nations of the world agreed to 50 years of strict isolation from one another in order to prevent additional conflicts. Fifty years later, the United States comes out of exile, only to learn that no one else went into isolation.
- (8) Your entire life, you've been told you're deathly allergic to bees. You've always had people protecting you from them, be it your mother or a hired hand. Today, one slips through and lands on your shoulder. You hear a tiny voice say, "Your Majesty, what are your orders?"
- (9) All of the "#1 Dad" mugs in the world change to show the actual ranking of dads suddenly.
- (10) When you're 28, science discovers a drug that stops all effects of aging, creating immortality. Your government decides to give the drug to all citizens under 26, but you and the rest of the "Lost Generations" are deemed too high-risk. When you're 85, the side effects are finally discovered.
- (11) A boy pretends he is an astronaut in order to help cope with concepts and situations he can't understand.
- (12) By the time humans come along, elves had invented space travel, and dwarves had split the atom. One hundred years later, the world looks like your typical fantasy setting. How did it happen?
- (13) A time traveller interviews major historical figures at three points in their lives: their 16th birthday, the day after they made their most important decision, and the day before they die.
- (14) Everyone has superpowers, but the richer you are, the weaker your powers become.
- (15) Every fifty years, the accumulated wealth of the world is randomly redistributed. Tonight is the eve of the global redistribution.
- (16) A woman comes into the same diner every morning, orders the same meal, and always leaves without eating a bite.
- (17) Due to a crossed line, a customer support worker has to deal with a hostage situation. Meanwhile a hostage negotiator has to deal with a disgruntled customer.

- (18) Construction workers are exposed to a relic of magical power while beginning work on a new building. Slowly, it begins to change them...
- (19) A retired supervillain is in the bank with his 6-year-old daughter when a new crew of supervillains comes in to rob the place.
- (20) Twin brothers with a strong telepathic connection discover the elixir of life. Only one is granted immortality, but their telepathic connection transcends the mortal brother's death, providing the first physical world/afterlife connection.

B.2 Writing Prompts in User Evaluation

Below are the two writing prompts used in the user evaluation.

- (1) Two characters with very different personalities are forced to work together toward a difficult goal. At first they clash, but over time they must decide whether to trust each other. Write a story about how their relationship develops.
- (2) Two characters who trust each other uncover a secret that could change their community. One wants to reveal it; the other wants to keep it hidden. Write a story about how their relationship and choices change as they face the consequences.

B.3 Baseline System Interface

Our baseline system includes character sheets and character chatbots for defining and interacting with characters (similar to CharacterChat [42] and CharacterMeet [38]), as well as a story outline for plot conditioning and a story chatbot for generation and ideation. Fig. 7 shows an example screenshot of the baseline interface.

B.4 Participant Information

through crowdsourcing platforms, social networks, and word of mouth. All participants reported proficiency in reading and writing in English. Participants had a range of creative writing expertise: 3 identified as professional writers with published work, 4 as advanced writers (2 of whom had also published work), 1 as an intermediate writer, and 4 as beginner writers. In addition, all participants reported prior experience using AI tools for writing. Their self-reported familiarity with using AI tools to support writing, measured on a 5-point scale (1 = none, 5 = extensive), was 3.92 ($SD = 1.08$). Detailed participant information is shown in Table 3.

B.5 Interview Questions

- (1) Looking across the two systems, how did your experience of developing and writing characters differ?
- (2) In which system did you feel more in control of how characterization was portrayed in the final story?
- (3) How did using 1st-person character perspectives as an intermediate step affect your understanding and expression of the character in the final story?
- (4) How did you decide when to use each of the three interfaces (canvas, track, and table)? For what kinds of tasks did you prefer each one, and why?
- (5) How, if at all, did either system affect your ability to deliberately plan character development throughout the story?
- (6) How did each system help or hinder you in translating abstract character ideas into concrete story details?
- (7) How easy was it, in each system, to trace aspects of the final story back to your intended character traits or earlier characterization work?
- (8) For characterization specifically, how would you use the system in your future writing practice?
- (9) What improvements would you make to the system to better support characterization in your future writing practice?

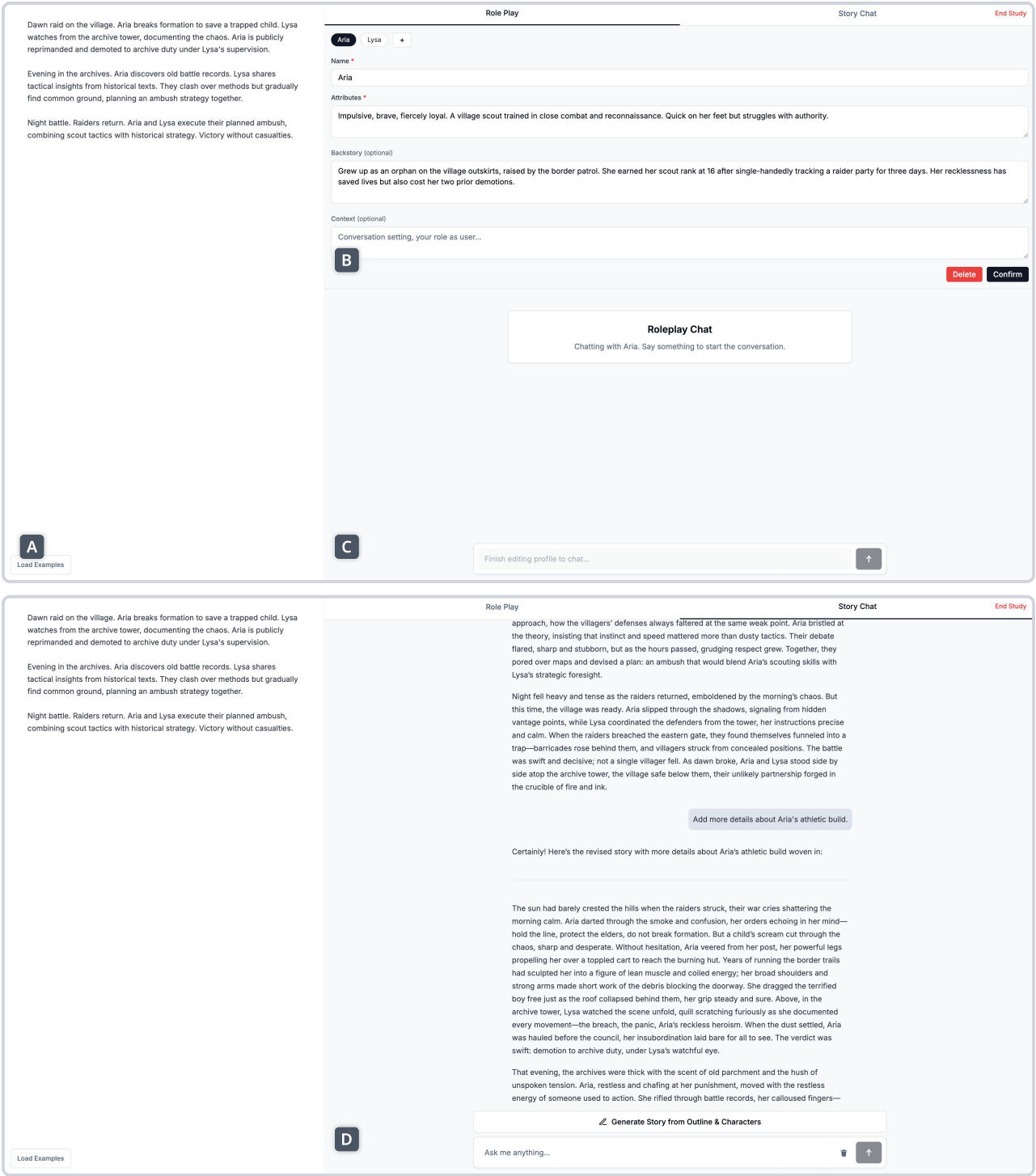


Figure 7: Screenshots of the baseline system. The baseline system includes a story outline (A) for plot conditioning and a story chatbot (D) for generation and ideation, as well as character sheets (B) and character chatbots (C) for defining and interacting with characters (similar to CharacterChat [42] and CharacterMeet [38]).

Table 3: Demographic information for participants. This table presents participants' ages, genders, creative writing experience, and use of AI tools for writing. We slightly modified their descriptions of writing experience and AI use to avoid identifiable information.

ID	Gender	Age	Creative Writing Experience	AI Usage Experience
P01	Female	49	Advanced: I published a speculative fiction young adult book recently.	High: I've used ChatGPT for fact checking and to check spelling, grammar and comprehensibility. I've used Claude Sonnet 4.6 to check my stories for developmental weaknesses (character arcs, plot).
P02	Male	60	Professional: I am working on book 30 in a series of science fiction, printing all 29 previous books at The Book Patch. I have created hundreds of educational workbooks for teachers, many full of poems, stories, descriptions, word problems, etc.	Extensive: I rely on Claude and ChatGPT to assist me before, during, and after writing. I set up the entire world within a book, using these two (and Gemini occasionally) to give me detailed explanations of the specific content each chapter will explore, how things are done (piloting a ship, digging out a gem, communicating with an alien species, dangers in space travel, etc.) and ways to allow my two main characters to experience awe, curiosity, wonder, focus, patience, etc.
P03	Female	28	Intermediate: I practiced writing stories for exams.	High: I use ChatGPT for academic writing, such as polishing my content and helping with my thoughts.
P04	Female	33	Professional: I am currently working on a screenplay set during the American Revolution, The film follows the first black poet to be published in the US.	Extensive: I use it a lot to check historical accuracies. I also use it for feedback on outlines because it gets me brainstorming.
P05	Male	25	Beginner: I enjoy the idea of writing stories, though I haven't explored it very much yet.	Limited: I use AI tools to find the best word/phrase to describe something. I also use AI to re-write sentences when I feel a sentence sounds weird or tedious.
P06	Female	28	Beginner: I'm interested in writing fictions, but I haven't tried it much yet.	Limited: I've used ChatGPT, Claude, Gemini. I usually ask AI to help me rewrite emails.
P07	Female	55	Professional: I have a master's degree in creative writing and have published several horror and sci-fi short stories in literary journals. I sold a TV movie and one of my short stories was published in a New York Times best-selling anthology.	High: I mainly use ChatGPT for help with my freelance clients, such as writing newsletters or bios. When I was interviewing for a position that involved writing verticals (micro stories), I asked ChatGPT to show me an example of a sci-fi vertical. I used Copilot to help me generate beats for a new screenplay.
P08	Female	25	Beginner: As a hobby, I wrote some short pieces, mostly short scenes. I used to write short science fictions when I was younger.	Moderate: I used AI tools for academic writing a lot, especially paraphrasing and proofreading.
P09	Female	65	Advanced: I have written two ebooks and a screenplay (131 pages long) available for sale on Amazon Kindle. Most currently, I frequently write scripts and record them for audio entertainment. I have ghost written a few books, wrote a one hour script for a podcast, and won a national writing contest as a college student for film criticism.	High: I just recently experimented with AI for writing. As a new Grandma, I asked for inspiration for a baby's storybook audio. To my surprise, about twenty minutes later, my AI had created a flawless nature themed personalized story.
P10	Male	29	Beginner: I've seldom been writing fiction and stories recently, but I do write user journeys for my products.	High: ChatGPT and Claude are mostly used to help with scientific writing. They have been mostly very helpful.
P11	Male	46	Advanced: I've been writing raps for over 25 years as well as poetry. I also have written three fictional books but haven't published anything and I'm currently writing two more.	Extensive: I have used Gemini, Chat gpt, and Perplexity to help with editing, brainstorming ideas and story pacing. I really like to play around with the AI using it for creative writing as well as role-playing scenarios.
P12	Male	24	Advanced: I write short stories, poems, and also occasionally write articles, I also have published a magazine.	Extensive: I use AI in my writing to help me critique my word choice and improve the flow of my papers. I try not to simply ask it to generate responses for me, but if I am not interested in the topic I am writing about and it is a run-of-the-mill paper, I sometimes ask AI for ideas about things to write about.