

Narrix: Remixing Narrative Strategies from Examples for Story Writing

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

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

Shunan Guo
Adobe Research
San Jose, California, USA
sguo@adobe.com

Eunye Koh
Adobe Research
San Jose, California, USA
eunye@adobe.com

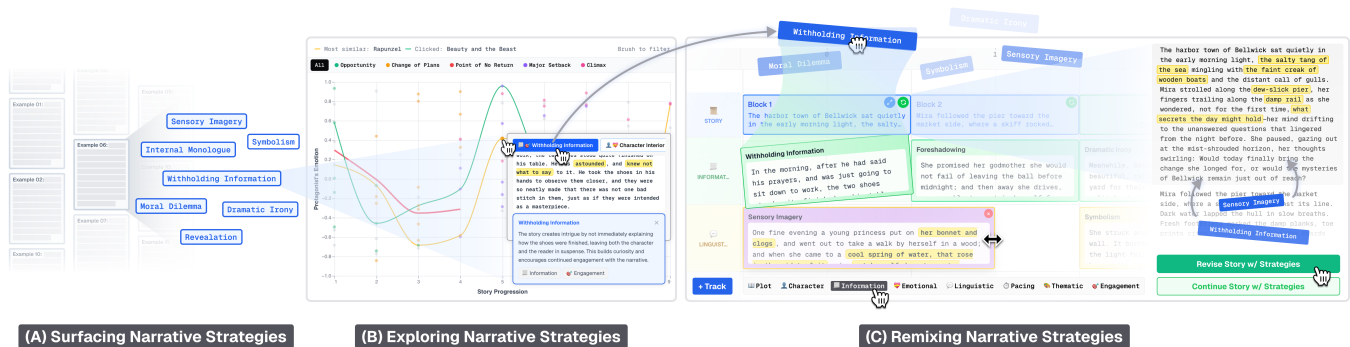


Figure 1: Three main features of NARRIX: (A) Surfacing Narrative Strategies: NARRIX automatically extracts narrative strategies from example stories, highlighting relevant lexical cues in color and explaining their function in the story. **(B) Exploring Narrative Strategies:** Users can retrieve example story content and their strategies based on evolving higher-level storytelling intents (e.g., emotional shifts, turning points) via an interactive story arc. **(C) Remixing Narrative Strategies:** Users can drag and drop selected strategies onto multi-dimensional tracks (e.g., character, plot, linguistic) to steer the AI in revising or continuing their story by integrating those strategies.

Abstract

Experienced storytellers decompose stories into local narrative strategies and how these strategies shape higher-level arcs. This decomposition helps writers recognize patterns in others' work and adapt those patterns to tell new stories. Novices, however, struggle to identify these strategies or to reuse them effectively. We present NARRIX, a novel writing tool that helps novice writers recognize narrative strategies in example stories and repurpose these strategies in their own writing. NARRIX analyzes strategies in example stories, highlights them with color-coded lexical cues and explanations, and situates them on an interactive story arc for exploration by emotional shifts and turning points. Writers then drag strategies

onto multi-dimensional tracks and apply block-scoped edits to revise or continue their drafts through controlled generation steered by specified strategies. Through a within-subjects study ($N = 12$), NARRIX showed improved participants' retention, confidence, and creative adaptation of narrative strategies compared to a baseline chat-based writing interface.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; *Natural language interfaces*; *Information visualization*.

Keywords

Story Writing; Example-Based Learning; Narrative Strategies; Cognitive Apprenticeship; Intelligent Writing Tools; Creativity Support Tools; Human-AI Interaction

ACM Reference Format:

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 (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 24 pages. <https://doi.org/10.1145/3772318.3790813>

*Work done during an internship at Adobe.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/26/04

<https://doi.org/10.1145/3772318.3790813>

1 Introduction

The writing of a good story often hinges on the execution of key narrative moments—an unexpected twist, a moment of emotional tension, or a sudden shift in tone—the impact of each paragraph depends on not just what the author says, but also how they choose to say it. For example, a suspenseful scene might build tension through sensory imagery and delayed revelation, or build toward a dramatic twist with misdirection and a sudden shift in point of view. However, novice writers often lack the creative repertoire needed to shape such moments in compelling ways. To build this repertoire, both cognitive apprenticeship [17, 18] and writing pedagogy [29, 30, 64] emphasize the value of learning from examples—not by simply lifting surface-level content, but by dissecting, internalizing, and remixing the *strategies* that underlie impactful storytelling. This principle is echoed across writing communities. As one writer in r/writing¹ puts it:

“Take their ideas, how they write dialogue, develop characters, conduct prose, etc. and turn it into your beautiful Frankenstein monster... amalgamate the things you love from your favorite stories and combine them into something that is truly yours, that entertains you, and inspires you, and that you are proud of.”

However, narrative strategies are often tacit, subtly woven into the fabric of the text. Novices may sense the impact of a compelling scene, but struggle to articulate how it works or to recreate its effect in their own writing. Exposure to too many examples can further compound the challenge, overwhelming rather than guiding writers as they attempt to locate, compare, and adapt useful techniques. Prior systems such as IntroAssist [37, 38] and CorpusStudio [19] surface genre norms and conventional structures in domains like email and academic writing. Yet, recognizing and reusing narrative strategies for story writing remains comparatively underexplored. Addressing this challenge requires more than surfacing patterns: it calls for scaffolds that helps novice writers interpret, experiment with, and internalize strategies within their own drafts.

To this end, we present NARRIX (Fig. 1), an AI-assisted tool for narrative writing. Grounded in the theory of cognitive apprenticeship [17, 18], NARRIX helps users recognize narrative strategies in example stories and repurpose these strategies in their own writing. To do so, it starts by decomposing example stories into local pieces—which we call *blocks*—and the narrative strategies those pieces use. It then helps the users recognize and understand narrative strategies by naming and labeling them in the interface, highlighting relevant lexical cues with color-coded, in-place annotations, and explaining the function of each strategy within the story. An interactive story-arc view supports users in exploring example blocks and their strategies according to evolving, higher-level storytelling intents (e.g., emotional shifts, turning points). Drawing on the metaphor of a digital audio workstation (DAW)—where musicians remix a song by arranging sound clips on a multi-track timeline and layering audio filters—NARRIX enables users to drag and drop strategies onto multi-dimensional tracks representing distinct story elements,

such as character, plot, and linguistic, to guide the AI in revising or continuing their story while applying those strategies as “filters.”

We evaluated NARRIX in a within-subjects study ($N = 12$) against a chat-based writing interface. Participants using NARRIX recalled and understood more narrative strategies, applied and remixed a greater number of strategies in their own writing, and reported higher confidence, satisfaction, and perceived creativity support than with a baseline chat-based writing interface. Their interactions reflected a cyclical exploration-learning-remix workflow, with remix functioning as the backbone and deliberate learning interwoven throughout. Strategy use followed both goal-driven and exploratory paths, supported by track-based layout and block-scoped editing that facilitated controlled, strategy-steered revisions. Overall, our work demonstrates how reifying tacit narrative strategies into visible, manipulatable units and supporting their adaptation can transform AI-assisted writing: from passive content generation into a process of deliberate learning and creative control.

2 Related Work

2.1 Examples in Creative Support Tools

CSTs in HCI help users discover, understand, and reuse examples across domains like design, programming, and writing. In this section, we review how CSTs (including writing tools) support creative work through three core practices: (1) finding relevant examples, (2) learning from examples, and (3) reusing examples in new contexts.

Examples can inspire and guide creativity, but locating the *right* example at the *right* moment remains challenging [66]. HCI systems address this with a range of retrieval methods [45, 50, 76, 95]. In design, tools like d.tour [76] support stylistic querying via descriptors such as “colorful” or “image-heavy”, while IdeateRelate [95] provides more abstract, concept-level matching. In writing, CorpusStudio [19] retrieves examples via textual (sentence-level) and structural (section-title) similarity; TaleStream [12] suggests tropes (i.e., storytelling conventions) during early ideation; and ScriptViz [70] automatically retrieves movie screenshots that satisfy user-defined SQL constraints to aid scriptwriting. However, these systems seldom model a creator’s evolving *intent* during the process. In storytelling, authors often need examples keyed to current narrative purpose (e.g., executing a turning point or shifting emotional trajectory) rather than only topic, convention, or lexical similarity. Our system supports such intent-driven retrieval by externalizing the writer’s goals through an evolving story arc and aligning examples to narrative position and affect.

The pedagogical value of examples is well established in HCI and language education [10, 22, 82, 83]. In design, novices benefit disproportionately from galleries because multiple examples help them infer underlying principles [50]. In writing, tools improve lexical and syntactic fluency by surfacing patterns from corpora: WriteAhead mines academic collocations [9]; Corpus of Contemporary American English (COCA)-based systems support ESL learners’ usage [58]; Langsmith suggests fluent academic sentences [41]; and Lettersmith scaffolds professional email writing with annotated checklists and aligned exemplars [37]. Yet these systems primarily focus on linguistic patterns or genre convention, while deeper creative strategies still remain implicit. We address this gap by

¹The subreddit r/writing (<https://www.reddit.com/r/writing/>), with over 3.2 million members as of 2025, is one of the largest and most active online communities for writers, offering a representative hub for sharing advice, experiences, and resources across genres and skill levels.

helping users explicitly discover, interpret, and repurpose narrative strategies instantiated in examples, moving beyond linguistic conventions toward strategy-level learning.

Many CSTs support direct reuse of examples. In visual domains, style-transfer and component-level remix support rapid adaptation [26, 50, 56]. For example, Misty [56] helps users blend specific aspects (e.g., color, layout, content) from one UI example into their work-in-progress designs. In writing, systems surface example text for composition and reuse: prior work supports recycling past replies in email [65], checklists with expert-annotated examples for help-seeking [38], and corpus-driven retrieval and highlights for academic writing [19]. However, these tools largely emphasize content-level borrowing (e.g., snippets, structural cues, stylistic conventions) rather than identifying and repurposing underlying creative strategies. Our approach builds on this practice but shifts the focus to strategy-level reuse, enabling writers to insert, adapt, and combine named strategies in their own drafts.

2.2 Cognitive Apprenticeship Support

Cognitive apprenticeship supports learning complex skills by making expert strategies visible and guiding learners through *modeling, coaching, scaffolding, reflection, and exploration* [17, 18]. Modeling, coaching, and scaffolding help students acquire integrated skills through observation and guided practice; reflection help students focus their observations of expert problem solving and gain conscious access to (and control of) their own problem-solving strategies; and exploration promotes autonomy by encouraging the application of strategies in new, self-defined contexts.

Several prior tools provide partial support for this pedagogy. For example, IntroAssist [38] and Lettersmith [37] surface expert-annotated texts and prompt self-reflection, thereby supporting modeling, coaching, and reflection. However, they rely heavily on manual annotations and still require users to perform much of the interpretive work themselves (particularly challenging for novices facing complex narrative strategies) and offer limited scaffolding for deeper reasoning or creative experimentation. Schemex [90], though designed for text analysis rather than writing, introduced an AI-powered workflow that enables users to extract patterns from examples through clustering, abstraction, and refinement using contrasting examples. More relatedly, CorpusStudio [19] models corpus-level norms (e.g., section headings, common sentence patterns) and supports reflection, yet leaves underlying narrative strategies relatively underexplored.

In contrast, NARRIX extends cognitive apprenticeship support by adding active scaffolding and exploration. It leverages AI to automatically surface tacit narrative strategies from examples, situate them by narrative position and affect, and guide writers as they experiment with and adapt these strategies within their own drafts. This approach helps novices grasp not only *what* makes a narrative moment compelling, but also *how* to repurpose strategies in context.

2.3 Intelligent Writing Tools

The HCI community has a long-standing interest in intelligent writing tools [51], supporting writers across various stages, such as brainstorming ideas [12, 27, 77], planning outlines [74, 87], drafting content [20, 33, 43, 47, 48], and refining text [1, 42, 52, 72, 85]. These

tools span diverse genres, including argumentative writing [98, 102], story writing [13, 34, 96], and scientific writing [79, 80].

In story writing, early systems emphasized autonomous generation via computational planning with rules [49, 61, 75], character-based simulation [7], case-based reasoning [28, 68, 74, 86], and, increasingly, language models for sentence infilling or interpolation [2, 35, 39, 89]. Nowadays, large language models (LLMs) enable writers to steer generation text through prompting [23, 24, 40, 81, 94, 100]. While prompt-based interactions offer flexibility, they also introduce several challenges for our target user and scenario. First, prior research has shown that users without AI expertise may struggle to design effective prompts [97]. This challenge is amplified when our target users are also novices in writing: they may lack narratological knowledge (e.g., narrative strategies), which hinders their ability to craft prompts that analyze example stories, extract strategies, and transfer them into their own writing. Second, conversational interfaces typically constrain users to a linear flow. Writers have limited visibility into global story progression, find it hard to compare their drafts with example stories at the level of story arcs, and struggle to coordinate strategies across broader story structure. Third, prompt-based interactions offers limited pedagogical transparency. LLM outputs rarely reveal *which* narrative techniques are employed, *why* they work, and *how* they can transfer to new context, making it harder for novice users to learn story-writing skills.

To address these limitations, recent work has explored interfaces that enable direct manipulation or embed structured interactions into human-AI co-writing process. For example, TaleBrush allows authors to sketch story arcs to shape narratives [14]; PatchView supports worldbuilding through dust-and-magnet metaphors [15]; VISAR [102] and Polymind [87] provides node-based visual programming for interactive co-writing; Dramatron [62] offers hierarchical story generation, enabling users to control screenplay generation across different narrative elements (e.g., characters, plots, locations, and dialogue); and Friction [98] and Synthia [99] integrates a feedback-based revision workflow into AI-assisted writing to help users reflect on comments and improve their drafts. While these tools provide richer structural guidance and afford more control than prompt-only interfaces, their assistance tend to focus on shaping story structure or organizing content, and the narrative techniques invoked by the model remain implicit within generated text. In contrast, NARRIX surfaces narrative strategies extracted from exemplar stories, visualizes how they unfold across an arc, and enables users to directly manipulate and combine them. This explicit focus on strategy visibility, explanation, and transfer complements existing structural and generative approaches.

To sum up, NARRIX advances beyond prompting-based assistants and structured co-writing tools by centering both controllability and learning around example-based story writing. Our visual interfaces allow users to inspect how strategies unfold across a story arc and orchestrate them when shaping their own drafts with LLM support. Grounded in cognitive apprenticeship principles, this workflow encourages learning by exposing narrative strategies from examples, explaining when and why a strategy works, and scaffolding their application in new contexts.

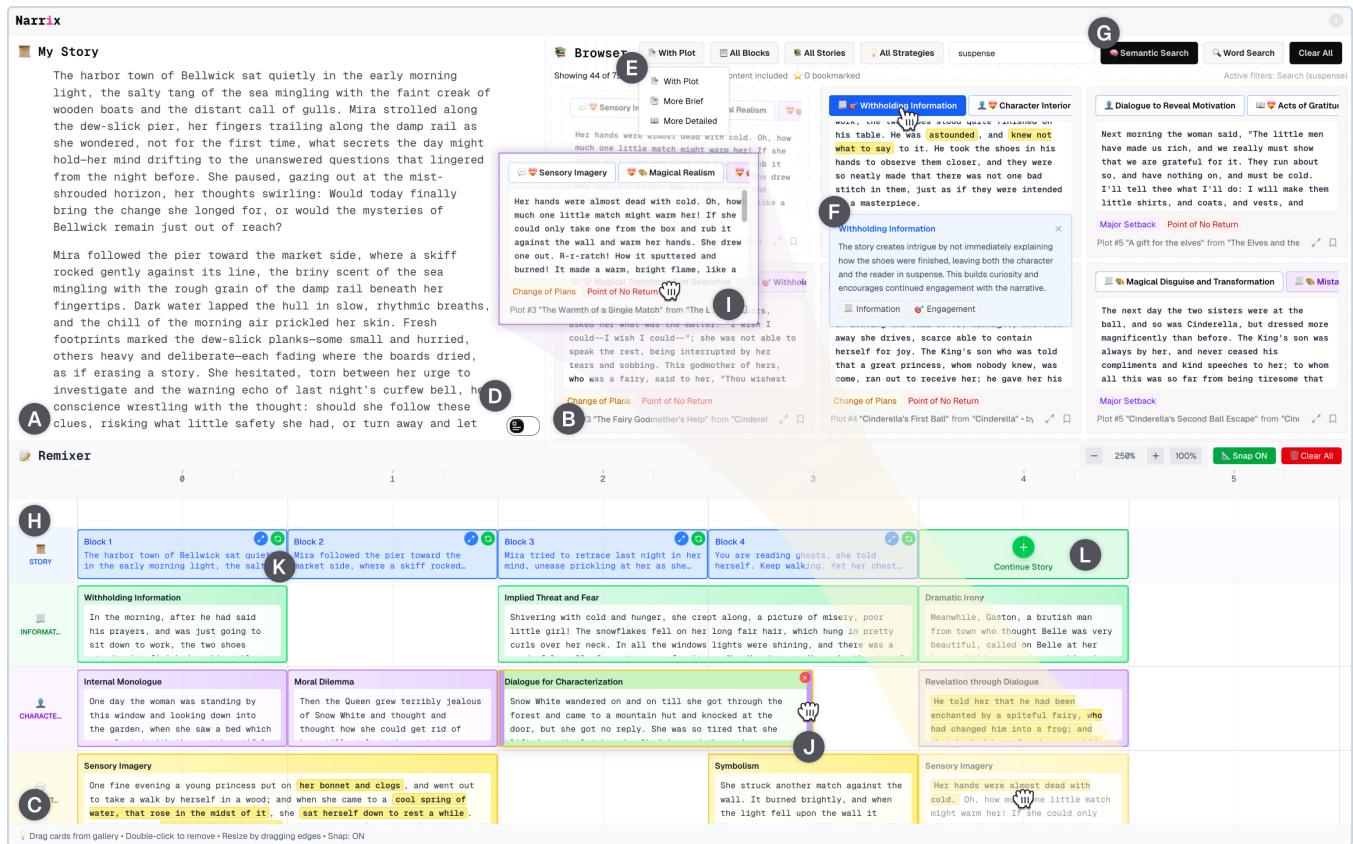


Figure 2: NARRIX’s user interface, organized into three views: (A) Markdown Editor for drafting the user’s story, with a mode switch (D) to toggle to the Story-Arc Inspector (see Fig. 3); (B) Browser for exploring example content blocks and their narrative strategies; and (C) Remix Tracks for layering strategies across creative dimensions. In the Browser, users can choose different card views (E), open strategy explanation panels (F), and perform semantic or word searches for strategies (G). Users can drag strategies from example cards onto relevant creative-dimension tracks (I) below the story track (H). Strategy tiles in the Remix Tracks can be resized to span multiple blocks (J). Users can steer AI to revise blocks (K) or continue the story (L) using the selected strategies.

3 System Design

We present NARRIX, an interactive system that supports novice writers in discovering, interpreting, and remixing narrative strategies from examples into their own stories. In this section, we first introduce the five design goals, grounded in the principles of cognitive apprenticeship from our related work §2.2, that guided the development of NARRIX. Then, we present the main interfaces and features of NARRIX and illustrate their use through a usage scenario featuring Zoey, a novice writer, leverages NARRIX in daily writing practice. Lastly, we provide the technical details on how narrative examples are processed and modeled to enable these interactions, along with implementation details of the system architecture.

3.1 Design Goals

Ward [91] defines example-based creativity as a form of “structured imagination”—modifying an existing solution and applying it to a new context. We aim to support novice writers by making tacit narrative strategies visible and guiding them through applying,

interpreting, and experimenting with those strategies in their own writing. To this end, our system design is grounded in the five components of cognitive apprenticeship [18], operationalized through the following design goals. First, NARRIX aims to **model (DG-M)** narrative techniques by surfacing them in example texts, highlighting how they are realized in compelling storytelling moments. Second, it **coaches (DG-C)** users by directing their attention to specific narrative moves and explaining their contextual effectiveness. Third, it **scaffolds (DG-S)** the interpretation and application of these strategies by bridging the gap between abstract understanding and concrete writing actions. Fourth, it encourages **reflection (DG-R)** by helping users compare their own writing with example texts to identify similarities, differences, and areas for improvement. Finally, it enables **exploration (DG-E)** by assisting users to experiment with combining different narrative strategies and observe how these choices shape storytelling in creative and varied ways.

3.2 Interface & Features

Informed by the design goals, we design and develop NARRIX. The interface of NARRIX consists of three coordinated views (Fig. 2), each supporting a key aspect of example-based story writing: a Markdown Editor (Fig. 2A) / Story-Arc Inspector (Fig. 3) (with a mode switch for drafting or visualizing story arcs, Fig. 2D), a Browser (Fig. 2B) for exploring example story blocks and their narrative strategies, and Remixer (Fig. 2C) for remixing and layering strategies across different creative dimensions. In this section, we will introduce the key features in these three views that help users (1) surface, (2) explore, and (3) remix narrative strategies.

3.2.1 Surfacing Narrative Strategies. In Browser (Fig. 2B), each uploaded story is segmented into coherent blocks, each comprising several sentences that represent a distinct narrative beat. To make abstract strategies visible and understandable to writers, NARRIX surfaces strategies, explanations, and concrete textual evidence directly within each block.

Strategy Browser: Every block appears as a card. Users can switch between different card views (Fig. 2E): with-plot (a detailed card with the story content of the block, as shown in Fig. 2F), more-brief (a brief summary card without any story content), or more-detailed (a context-rich card showing more story context). Users can expand \llcorner cards to view the full story or bookmark \blacksquare them for easy access.

Strategy Annotation: For each card, the system infers one or more narrative strategies, such as `Sensory Imagery`, `Internal Monologue`, `Personification` [DG-M]. To help writers understand how abstract strategies function within authentic story contexts, NARRIX incorporates an explanation panel for each narrative strategy (Fig. 2F) which activates upon user click. This panel (i) describes how the strategy operates in the story, (ii) lists its dimension tags (e.g., character, information, linguistic), and (iii) highlights lexical cues (e.g., words, phrases) such as `glimmering`, `whispered to herself`, or `as if the trees could talk` that signal the strategy in context [DG-C]. To further support exploration, the system also provides semantic search (Fig. 2G) that allows users to find instances of specific strategies across all uploaded examples based on their names, explanations, and lexical features.

3.2.2 Exploring Narrative Strategies. As the segmented story blocks and their strategies accumulate, even a modest set of example stories can quickly balloons into a large and complex information space. Dozens of narratives may expand into hundreds of blocks, each associated with multiple candidate strategies. While this richness offers learning potential, it can also overwhelm writers, especially those seeking targeted inspiration or tactical guidance. To help writers find strategies most relevant to their current creative goals, NARRIX provides interactive visualizations that organize the example space by storytelling intent, both structurally through narrative arcs and functionally through turning points and strategy roles.

Story-Arc Visualization: A story arc charts the protagonist's emotional journey of a narrative and is widely used to convey storytelling intent [84]. NARRIX juxtaposes the user's evolving arc with those derived from the example corpus, displayed within a shared coordinate space: story block order appears on the x -axis and affective valence on the y -axis. The user draft is rendered as



Figure 3: NARRIX's Story-Arc Inspector, which juxtaposes the user's evolving story arc (red line) with arcs from example stories. Each point represents a story block, positioned by story progression (x -axis) and the protagonist's emotional valence (y -axis). Hovering over a point (B) reveals its details and highlights the corresponding card in the Browser; brushing the points (C) selects a range of story progression and emotion. The system automatically overlays the most similar example arc in yellow (D) and optionally additional arcs in green (E) toggled by user click for comparison. Users can filter story blocks by turning-point type (F) to focus on strategies serving specific narrative purposes.

a red line (Fig. 3A), while each example block is shown as a scatter-plot point \bullet [DG-M]. Selecting or hovering over a point (Fig. 3A) highlights the corresponding card in the Browser (Fig. 3B), while brushing the points (Fig. 3C) allows users to select ranges of story progression and emotion valence [DG-S].

To help writers discover examples with comparable emotional trajectories and draw attention to structurally relevant strategies, NARRIX automatically identifies and overlays the most similar example arc in yellow (Fig. 3D) (§3.4.3). Optionally, users can overlay additional arcs in green by clicking other points, enabling side-by-side comparison of multiple narrative trajectories [DG-R].

Turning-Point Filtering: Turning points (`Opportunity`, `Change of Plans`, `Point of No Return`, `Major Setback`, and `Climax`) represent the underlying goal or function of individual story blocks and offer a complementary lens for exploring narrative strategies [67, 84]. Each story block in the example corpus is automatically categorized by its turning point type, if any (§3.4.2) [DG-M]. Users can filter the scatterplot points and Browser cards simultaneously by turning points (Fig. 3F), enabling focused exploration of strategies that serve specific narrative purposes within their story.

3.2.3 Remixing Narrative Strategies. Once they identify preferred strategies, users can incorporate and experiment with them in their own drafts using Remixer interface. This view, designed based on the metaphor of a digital audio workstation (DAW), enables writers to apply narrative strategies with fine-grained, track-based control.

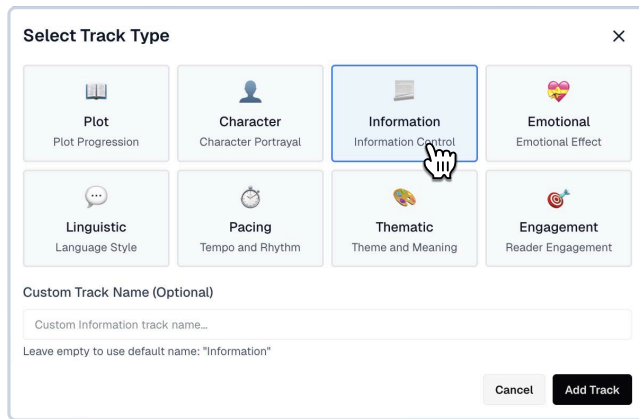


Figure 4: NARRIX’s Track Selection panel for adding tracks of different creative dimensions to the Remix workspace. Users can choose from eight predefined dimensions, e.g., Plot, Character, Information, Emotional, Linguistic, Pacing, Thematic, and Engagement.

In a DAW, musicians remix a song by arranging sound clips on a multi-track timeline and layering audio filters to achieve the desired effect. Analogously, the Remixer interface in NARRIX arranges story blocks (cf. sound clips) in draft order and allows writers to layer narrative strategies (cf. audio filters) across multiple creative tracks, such as characterization, linguistic style, and information delivery. This metaphor provides a familiar mental model of “drag-and-drop, layer, tweak, and audition” interactions, enabling users to engage with narrative constructions through what Ward [71, 91] describes as “structured imagination.”

Track Setup: The top track (Fig. 2H) is the story track, displaying user-authored blocks in sequence. The system creates new block whenever the user presses Enter in Editor. Below the story track, users can add tracks that represent different dimensions of storytelling by clicking **+Track** and selecting from (Fig. 4): **Plot**, **Character**, **Information**, **Emotional**, **Linguistic**, **Pacing**, **Thematic**, and **Engagement**. These dimensions are grounded in foundational theories from narratology and writing studies [3, 46, 60, 69], capturing key creative aspects emphasized in story writing. Appendix A.2 provides detailed definitions of each dimension.

Multiple tracks (of the same or different dimensions) can be added as needed [DG-E]. As users interact with the Remixer, the Browser remains synchronized with the currently focused track dimension, automatically updating the displayed information in each card with the relevant strategies.

Drag-and-Drop Remixing: Users can drag cards from the Browser onto any track (Fig. 2I). If multiple strategies in a card are associated with the target dimension, the system prompts the user to select which to apply, based on its explanations and relevant lexical features highlighted in the example. Selected strategies are displayed as tiles beneath the corresponding block, and users can resize these tiles to span multiple blocks by dragging handles on either side (Fig. 2J). This drag-and-drop interaction enables users

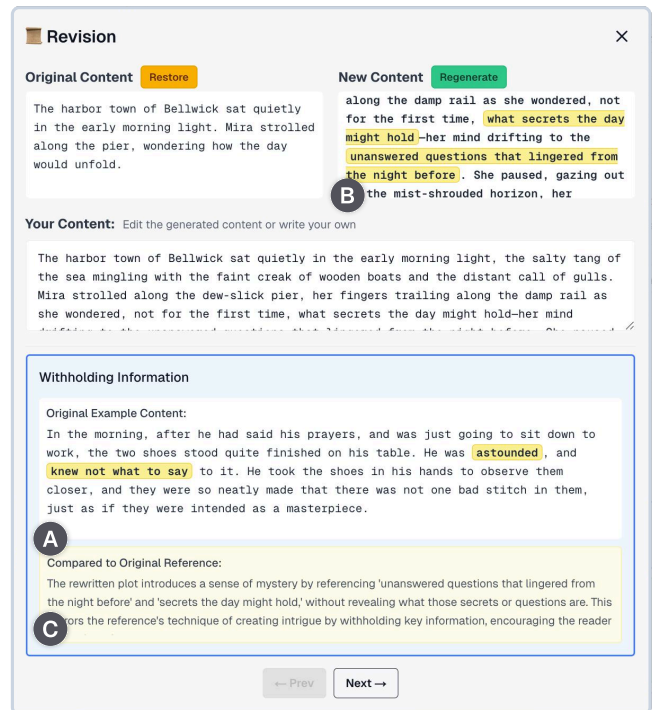


Figure 5: NARRIX’s Reflective Comparison view, which contrasts: (A) the original example content and its highlighted lexical cues for the selected strategy, (B) the revised user content with corresponding cues, and (C) a generated comparison explaining how the strategy is realized in both texts and what differences or similarities exist.

to experiment with different combinations of narrative strategies, supporting planning and creative experimentation [DG-E].

Revision and Continuation: When strategy tiles are present under a story block tile, users can click a revise **C** icon (Fig. 2K) to have the system revise that block by applying the selected strategies, while maintaining narrative coherence [DG-S]. Similarly, if strategies are present beneath the last placeholder tile in the story track, users can click **Continue Story** (Fig. 2L), optionally provide a short description of the desired next story block, and have the AI continue the story by applying the specified strategies. Users can further revise, regenerate, or add the generated content to their story. An expand **⌵** icon (Fig. 2K) allows users to review the generation history and restore previous versions as needed.

Reflective Comparison: After each revision or continuation, NARRIX generates a side-by-side comparison view (Fig. 5) to support learning through contrast and reflection [DG-R]. The view shows: (i) how the strategy operates in the original example, including its lexical features (Fig. 5A); (ii) how it is realized in the revised user content, with corresponding lexical features (Fig. 5B); and (iii) the key differences and similarities between them (Fig. 5C). By explicitly connecting the strategies intent with textual outcome, this view helps writers internalize how strategies operate and facilitates the transfer of learned strategies to new contexts.

3.3 Example Usage Scenario

Zoey, a novice fiction writer, is drafting a 1,000-word micro story for a community challenge. After completing an opening scene, Zoey finds the scene lacking in tension and vivid imagery. To strengthen the draft, they upload ten classic fairy-tale retellings (e.g., *Cinderella*, *The Frog Prince*, *Rapunzel*) to NARRIX to surface narrative strategies and might be repurposed.

3.3.1 Browsing and Searching for Narrative Strategies. Zoey begins by exploring the Browser, which displays block cards segmented from the uploaded stories, each annotated with candidate strategies. Seeking to add suspense to the opening, Zoey types “suspense” into the search bar. Cards tagged with strategies such as **Withholding Information** rise to the top (Fig. 2F). Clicking on this strategy opens a panel (Fig. 2F) to read a detailed explanation panel where Zoey read how it works and study highlighted lexical cues such as **astounded** and **knew not what to say**. Before moving on, Zoey skims several other cards and bookmarks the ones that seem most promising.

3.3.2 Remixing Strategies with Creative Tracks. To weave these tactics into the opening, Zoey clicks **+Track** to first add a **Information** track, then a **Character** track, and finally a **Linguistic** track (Fig. 2C). Zoey drags **Withholding Information** onto the **Information** track to heighten narrative stakes; **Sensory Imagery** from *The Frog Prince* goes onto the **Linguistic** track, enriching descriptive detail; and **Internal Monologue** from *Rapunzel* drops onto the **Character** track, revealing the protagonist’s unease. NARRIX rewrites the opening paragraphs with these layered strategies after Zoey clicks the revise button (Fig. 2K). A comparative panel (Fig. 5) contrasts their use in the exemplars with the revised draft, spotlighting phrases such as **secrets the day might hold** and **lingered from the night before**.

3.3.3 Exploring Strategies via Interactive Story Arcs. As the draft develops, Zoey switches to Story-Arc Inspector to reflect on the story’s evolving emotional structure. The protagonist’s emotional trajectory appears as a red arc that drops sharply and then plateaus (Fig. 3A). NARRIX automatically highlights the most similar example arc—in this case, from *Rapunzel*—in yellow (Fig. 3D). Zoey clicks the **Climax** filter to isolate example story blocks that serve as climactic moments (Fig. 3F); Browser cards and scatter points update to show only the narrative endgame. Next, Zoey brushes a band along the *y*-axis, focusing on points with high emotional intensity, to surface strategies that resolve conflict through dramatic payoff and to refine the set of candidate endings.

3.3.4 Guided Continuation and Iterative Refinement. To complete the story, Zoey drags **Dramatic Irony** onto the **Information** track beneath the last placeholder tile and clicks **Continue Story**. After typing **Let readers grasp the hidden truth while the protagonist remains unaware.**, NARRIX generates an ending applying the selected strategy. Satisfied with this ending, Zoey revisits the middle paragraphs. By layering additional strategies—**Symbolism** onto the **Linguistic** track and **Moral Dilemma** and **Dialogue for Characterization** onto the **Character** track—Zoey deepens tension, voice, and thematic texture, leaving the draft poised for further iterative polish with NARRIX’s strategy-driven support.

3.4 Technical Details of Processing Narrative Examples

Our system employs a modular processing pipeline that transforms raw narrative examples into structured, analyzable components that support storytelling. In this section, we describe the technical pipelines (Fig 6) and evaluations for (1) inferring narrative strategies, (2) detecting key turning points from narrative examples, and (3) modeling story arcs.

3.4.1 Inferring Narrative Strategies. This component is responsible for identifying and explaining abstract narrative strategies within story context. The pipeline begins by prompting GPT-4.1 to segment each example story into story blocks, each comprising several sentences that represent distinct narrative beats (Fig 6A). We then prompt GPT-4.1 to infer narrative strategies (as short labels) from each block, along with explanations and relevant lexical cues (i.e., words or phrases that signal the strategy) (Fig. 6B). Given known LLM failure modes, such as hallucination and bias [44, 101], we describe our iterative prompt design process, discuss risks and mitigations, and report human evaluation of the final prompts, including representative failure cases.

Prompting Techniques: Following established prompt design patterns [92], we used a two-part prompt comprising a structured system instruction and a specific context. The system prompt positioned the model as an “expert literary analyst,” provided clear definitions and scope for “narrative strategies,” and specified the required outputs: (1) a concise name, (2) a brief explanation, and (3) verbatim lexical features from the example. The context prompt embedded the story block and reinforced evidence extraction requirements. A fixed JSON schema ensured consistent formatting and grounded each strategy in explicit textual evidence.

We iterated the prompts through short, repeated cycles of diagnostic review, in which we performed systematic inspection of sampled model failures to identify recurring error patterns. We addressed the error patterns identified in each iteration by adding constraints in the prompt: (a) **evidence grounding** via mandatory verbatim lexical cues that ties each claim to observable text; (b) a **constrained schema** reduces free-form generation and simplifies downstream verification; and (c) **definition priming** that narrows scope to functional narratology rather than plot summary. These constraints specifically target common failure modes (e.g., over-generalization, fabricated evidence, and hallucinations) and are supported by NLP findings that extractive, source-attributed justifications improve factual faithfulness and auditability [53], while verification-style prompting has been shown to reduce hallucinations by forcing consistency checks against evidence [21]. However, we acknowledge that constraints in a prompt encourage grounding but do not technically guarantee or “tie” it in a deterministic sense. Although such techniques do not eliminate the risk of ungrounded output, they can increase transparency and make failures easier to detect and adjudicate. The final prompts are provided in Appendix B for replication.

Expert Evaluation: Because inferring narrative strategies from text is inherently subjective—with no single correct answer or established ground truth—we conducted a human-centered evaluation

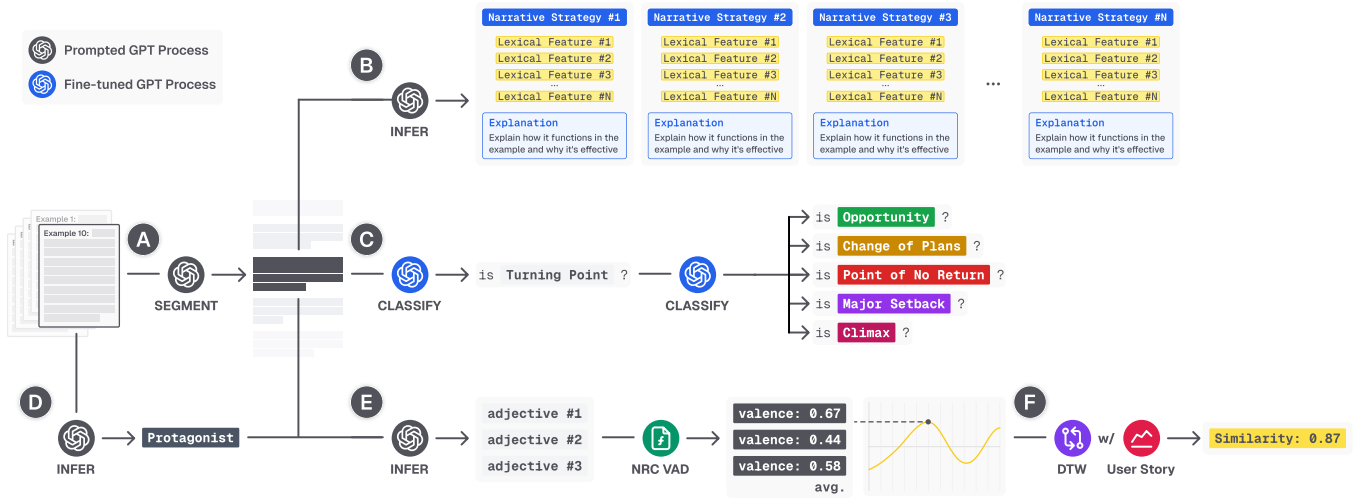


Figure 6: NARRIX’s technical pipeline for processing narrative examples. (A) Segment each example story into content blocks. (B) Infer narrative strategies for each block, including a concise label, explanatory description, and highlighted lexical features. (C) Classify each block for the presence of turning points using fine-tuned models. (D) Identify the story’s protagonist. (E) Infer the protagonist’s emotions per block, map them to valence scores using the NRC VAD Lexicon, and sketch a story arc. (F) Compare the user’s arc to example arcs using DTW to retrieve the most similar narrative trajectory.

with ten creative writing experts, each with over ten years of writing experience. We processed the dataset from Tian et al. [84], and then sampled 100 candidates from output, each consisting of a story block, an inferred strategy, its explanation, and highlighted lexical features. Each expert rated 20 unique examples, with two experts per example to support inter-rater reliability analysis.

We asked experts to assess two aspects: (1) whether the inferred strategy accurately captured the functional narratology of the story block and whether the accompanying explanation was coherent and insightful; (2) whether the lexical cues reflected the identified strategy. Ratings were recorded on a 3-point scale (1 = incorrect, 2 = correct but partially informative, 3 = correct and fully informative). The intra-class correlation coefficient (ICC) was 0.62, indicating good agreement among raters [16].

The average score for strategy quality was 2.54 ($SD = 0.62$), with 89% of examples received a score of 2 or above from both experts, suggesting that hallucinations or major interpretive failures were rare. Among the failure cases, raters noted several recurring patterns: plausible labels but were not sufficiently supported from the local block (e.g., naming a MacGuffin without showing how it drives decisions), explanations that summarized a “chain reaction” instead of isolating the causal hinge, or references to a line of dialogue or a setting detail without clarifying their role in the scene. Lexical cues received an average rating of 2.48 ($SD = 0.58$), with only 8% of examples rated as incorrect. Errors reported by raters mainly involved generic or misaligned cues, such as selecting long spans instead of pinpointing the words that realize the claimed strategy.

Overall, the results suggest our methods lead to model’s outputs that were perceived as largely accurate and interpretable. Nonetheless, LLM errors cannot be fully eliminated. For users, these errors may surface as overconfident labels, incomplete explanations, or

overgeneral cues. NARRIX buffers these risks by (i) surfacing evidence highlights, (ii) giving users control over strategy selection, and (iii) requiring confirmation for every AI change to the text. We provide a detailed error analysis in Appendix C, and encourage future work to extend our methods.

3.4.2 Detecting Turning Points. To detect turning points in narrative flow, we modeled the task as five independent binary classification problems, one per turning point type $T = \{t_1, t_2, t_3, t_4, t_5\}$, as each story block b_i may contain zero or more turning points ($|T_i| \geq 0$), where $T_i \subseteq T$ (Fig. 6C). For each turning point type $t_k \in T$, we constructed a dataset comprising 300 positive story blocks $P_k = \{b_i \mid b_i \text{ contains } t_k\}$ and 300 negative story blocks $N_k = \{b_j \mid b_j \text{ does not contain } t_k\}$ based on the labels in the dataset of Tian et al. [84]. Each dataset was split into training (60%), validation (20%), and test (20%) sets. We then fine-tuned a GPT-4o-based binary classifier f_k for each type of turning point using the corresponding training and validation sets., where $f_k(b) = 1$ if the story block b contains turning point type t_k , and $f_k(b) = 0$ otherwise. The fine-tuned models achieved consistently strong performance across all types on the test set, with average accuracy (0.88), precision (0.87), recall (0.92), and F1 scores (0.89), outperforming a prompting-only GPT-4.1 baseline in terms of accuracy: Opportunity (*ours*: 0.91 vs. *baseline*: 0.78), Change of Plans (*ours*: 0.88 vs. *baseline*: 0.61), Point of No Return (*ours*: 0.85 vs. *baseline*: 0.69), Major Setback (*ours*: 0.84 vs. *baseline*: 0.66), and Climax (*ours*: 0.91 vs. *baseline*: 0.87).

3.4.3 Modeling Story Arcs. This component constructs quantitative representations of narrative structures and enables comparison between stories based on their affective trajectories. Here, we first describe how we sketch story arcs and then outline our approach for measuring the similarity between story arcs.

Sketching Story Arcs: We followed the method from Tian et al. [84] to derive emotional arcs from stories. We first instruct GPT-4o to identify the main character of the story (Fig. 6D). Then, for each story block b_i in a narrative, we ask the same LLM to infer three adjectives, $W_i = \{w_{i1}, w_{i2}, w_{i3}\}$, that describe the protagonist’s emotions as the story progresses (e.g., *amused, relaxed, anxious*) (Fig. 6E). To quantify these emotions, we utilize the NRC VAD Lexicon [63] to obtain the valence scores of w_{ij} , ranging from 0 to 1, where $j \in \{1, 2, 3\}$. NRC VAD Lexicon provides human ratings of valence for more than 55,000 English words and phrases. For each story block, we use the average scores of w_{ij} to represent the valence of b_i obtaining $V(b_i)$, creating a quantitative emotional trajectory for the entire narrative.

Measuring Similarity of Story Arcs: We compare emotional trajectories (valence curves) using Dynamic Time Warping (DTW) [5] (Fig. 6F). Let $A = [a_1, \dots, a_m]$ and $B = [b_1, \dots, b_n]$ denote two valence sequences (bounded in $[-1, 1]$). The dynamic programming (DP) recurrence is

$$D_{i,j} = |a_i - b_j| + \min\{D_{i-1,j}, D_{i,j-1}, D_{i-1,j-1}\}, \quad (1)$$

$$1 \leq i \leq m, 1 \leq j \leq n.$$

with boundary conditions enforcing a shared start: $D_{0,0} = 0$, $D_{i,0} = +\infty$ for $1 \leq i \leq m$, and $D_{0,j} = +\infty$ for $1 \leq j \leq n$. Given the user story can be incomplete, we allow the alignment to end anywhere in the reference, so the DTW distance is $D^* = \min_{1 \leq j \leq n} D_{m,j}$. Lastly, we convert D^* to a similarity score $S \in [0, 1]$ using an upper-bound normalization:

$$S = \max\left(0, 1 - \frac{D^*}{2 \cdot \max(m, n)}\right). \quad (2)$$

We retrieve the most similar arc by ranking stories in the corpus by S and visualize it alongside the user’s arc.

3.5 Implementation Notes

NARRIX is built with the Next.js² framework, which supports server-side rendering for API calls, including those to the OpenAI APIs³ for instructing pre-trained GPT models, and the Firebase⁴ APIs for logging user events. We use Vega-Lite⁵ to render the interactive story arcs and scatter plots, and BlockNote⁶ to build the Markdown editor. Sample prompts and outputs for inferring narrative strategies, generating story content, identifying the protagonist, and extracting emotional adjectives are provided in Appendix B.

4 User Study

We conducted a preliminary within-subjects study with 12 novice story writers to evaluate the efficacy of NARRIX. The study aims to answer the following questions:

RQ1 *How do novice writers perceive the usability, usefulness, creativity support, and overall experience of NARRIX?*

RQ2 *In what ways does NARRIX help novice writers learn narrative strategies from examples during story writing?*

²<https://nextjs.org/>

³<https://openai.com/api/>

⁴<https://firebase.google.com/>

⁵<https://vega.github.io/vega-lite/>

⁶<https://www.blocknotejs.org/>

RQ3 *In what ways does NARRIX help novice writers remix narrative strategies from examples into their own stories?*

4.1 Baseline

Given that no existing solutions are directly comparable to NARRIX in terms of supporting interaction with narrative strategies in examples, we implemented a baseline system that closely resembled NARRIX in UI but excluded the key features unique to our approach.

Both systems shared the same Markdown text editor, ensuring a consistent writing environment. However, in the baseline, we removed the interactive story arc visualization. In the Browser panel, example stories were shown as story cards; each card could be clicked to reveal the full text of the example story, but without highlighting or extracting narrative strategies. The Remix panel in NARRIX was replaced by an AI chat assistant. This assistant, powered by the same underlying LLM but designed to mimic mainstream chat-based writing interfaces (such as ChatGPT Canvas): users could interact via free-form chat prompts, and the system would respond conversationally in a dedicated output panel. This baseline followed design conventions used in prior work [59, 73, 98], which also implemented chat-based AI writing assistants as comparison conditions when evaluating novel writing tools.

In summary, this baseline design (1) retains basic non-contributory features of NARRIX (e.g., the Markdown editor) to minimize interface confounds, (2) uses the same underlying model to isolate interaction design effects, (3) mirrors real-world practices where users have access to general-purpose AI tools like ChatGPT, and (3) integrates story editing, example browsing, and AI assistance within a unified workspace to avoid unnecessary window switching. A screenshot of the baseline interface is provided in Appendix D.1.

4.2 Participants

We recruited 12 participants (8 female, 4 male), ages 23–28 ($M = 26.27$, $SD = 1.68$), from a large software organization in the United States via internal communication channels and word of mouth. We sought novice writers and, following prior work that recruited ESL (English as a Second Language) writers as novices for writing tasks [36, 98], targeted non-native English speakers during recruitment. All participants reported regularly engaging in writing and wanting to improve their creative writing skills; each had prior creative writing experience and self-rated their expertise on a 7-point scale ($M = 2.33$, $SD = 1.07$; 1 = beginner, 7 = professional).

On a 5-point scale, participants reported actively seeking and referring to examples in writing ($M = 4.08$, $SD = 0.90$; 1 = never, 5 = always) and regularly using generative AI tools (e.g., ChatGPT) in their writing activities ($M = 4.67$, $SD = 0.50$; 1 = never, 5 = daily). Appendix D.2 provides detailed participant information (gender, age, creative writing proficiency, example-usage frequency, and AI-tool usage and frequency). We compensated each participant with a \$30 digital gift card.

4.3 Tasks and Materials

We designed two themed micro-story writing tasks. Each task prompted participants to write a short story in response to a writing prompt. This format draw on established microfiction practices,

such as flash fiction or five-sentence fiction, which challenge writers to convey a complete story in just a few lines. The two writing prompts were adapted from the online writing community *r/WritingPrompts*⁷: (1) write a story that begins with a feeling of joy and ends with a sense of surprise or uncertainty; and (2) write a story that begins with a feeling of confidence and ends with a sense of doubt or realization. These two prompts were selected because they exemplify widely used creative writing exercises that encourage the practice of core narrative skills and are commonly featured in both online writing communities and traditional writing workshops.

We collected twenty widely recognized literary stories (e.g., *Cinderella*, *Little Red Riding Hood*, and *The Little Match Girl*) as examples. By providing universally familiar and shared cultural stories as materials, we aimed to reduce extensive reading time required to comprehend unfamiliar plots and help participants focus on the writing task itself. In addition, the choice of these stories addresses practical considerations, as they are all in the public domain and free from copyright restrictions. The twenty stories were randomly split into two sets of ten, one for each prompt, controlling the length (1,448.8 vs. 1,461.6 words) to ensure consistency across tasks. We also piloted both writing prompts to ensure they were comparable in difficulty and creative potential.

4.4 Study Procedure

The study began with informed consent and a demographics questionnaire, followed by a brief orientation introducing the concept of learning from examples, with guidelines adapted from cognitive apprenticeship theory [17, 18] and community writing resources. The orientation emphasized actively analyzing example stories to surface underlying strategies, rather than copying surface features, and introduced example strategies. Participants were instructed to focus on identifying and interpreting deeper techniques from examples, and to consider how they might adapt them in their own writing.

Participants then completed two 30-minute writing sessions, one with NARRIX and one with the baseline; the order of systems and task materials was counterbalanced across participants. Each session began with a 3–5-minute tutorial covering key features of the assigned system. Participants were then encouraged to explore the provided example stories and compose a micro-story, using system features to revisit examples and experiment as needed.

After each story, participants were asked to perform a brief recall exercise, listing narrative strategies learned from the examples and providing a short definition for each, followed by a post-task survey (see §4.5). During the recall tasks, participants were not allowed to check the examples or the system. Both task sessions were completed in a single sitting, with a short break (approximately 5 minutes) between conditions. The study concluded with a 15-minute semi-structured interview to gather qualitative reflections on their experiences across the two conditions. The entire study lasted approximately 90 minutes per participant.

4.5 Measures

We aligned our measures directly with the research questions, combining standardized survey instruments, performance in recall tasks, quantitative usage logs, and qualitative coding from interviews.

⁷<https://www.reddit.com/r/WritingPrompts/>

To address **RQ1**, we assessed participants' subjective experiences via post-task surveys: the short-form Usability Metric for User Experience (UMUX-LITE) [54], NASA Task Load Index (NASA-TLX) [31], Creativity Support Index (CSI) [11], and five items adapted from Wu et al. [93] to capture experiences using an AI system. All items used a 7-point Likert scale.

To address **RQ2**, we measured *information retention*—the immediate internalization and recall of narrative strategies—following prior work [25]. We assessed the quantity of information retention by counting the number of distinct strategies each participant recalled and the quality of information retention by coding their written descriptions as vague (0), partial understanding (0.5), or full understanding (1). We also included a survey item on self-reported confidence in understanding these strategies, along with two items assessing whether the system helped participants identify narrative strategies in examples and understand strategies in context. In interviews, we further probed whether and how the system supported learning narrative strategies and participants' perceptions of its impact on longer-term writing skill development.

To address **RQ3**, we evaluated how participants remixed strategies into their own story writing. Two survey items captured self-reported confidence and satisfaction in adapting strategies with each system, supplemented by three items assessing whether the system helped them apply strategies in their stories, reflect on their usage, and experiment with combining strategies for creative exploration. We analyzed usage logs to characterize remix behaviors (e.g., frequency of revisions/continuations linked to strategies, number of tracks used across creative dimensions, number of distinct strategies applied). Finally, interview themes revealed participants' strategies on selecting and remixing narrative strategies, their practices for steering AI generation, and evolving attitudes toward example-based writing across conditions.

4.6 Analysis

For quantitative measures (e.g., Likert-scale responses, number of narrative strategies recalled, and feature-usage logs), we used the Wilcoxon signed-rank test due to the small sample size and the non-normal distribution of the data. For qualitative analysis of interview transcripts, we followed established open-coding protocols [6, 78]. Two researchers independently coded the transcripts, then discussed, reached a consensus, and created a consolidated codebook. This codebook was then used for thematic analysis to identify emerging topics from the interviews. The entire research team collectively reviewed the coding outcomes to refine high-level themes.

5 Findings

In the following sections, we will investigate our research questions in depth and present the corresponding findings.

5.1 RQ1: General Usage and Perception of NARRIX

5.1.1 Usage Patterns. We first examine participant usage of NARRIX from event logs. As shown in Fig. 7, we categorized events into three groups: Exploring (e.g., interacting with the story arc inspector and applying filters), Learning (e.g., checking and reading stories,

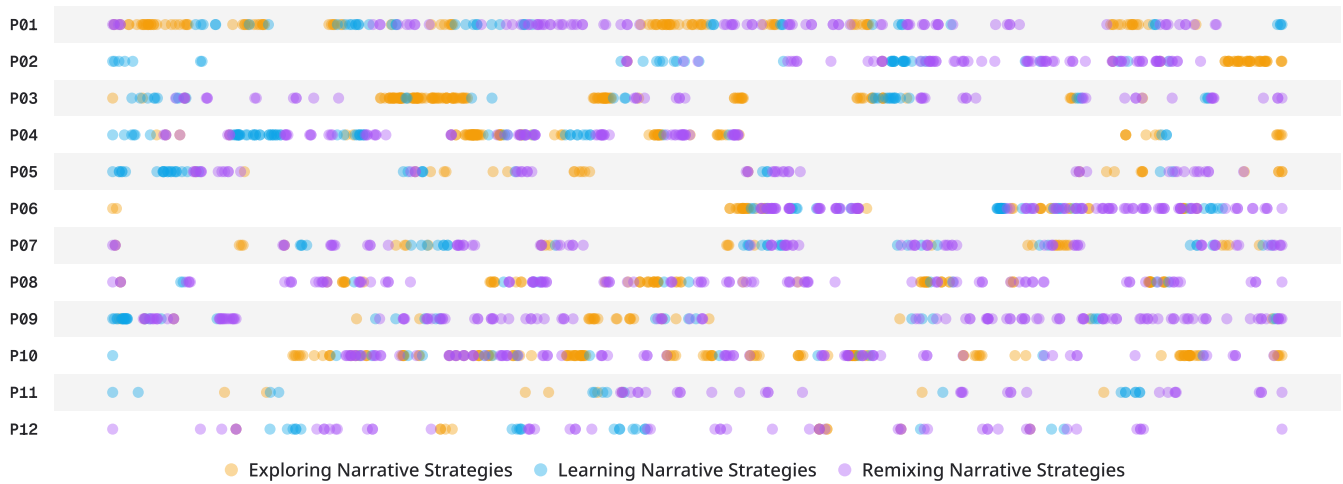


Figure 7: The distribution of events related to exploring, learning, and remixing narrative strategies across the normalized time span. Exploring Narrative Strategies (yellow circles) includes actions such as interacting with the interactive story arc and applying filters. Learning Narrative Strategies (blue circles) includes actions such as checking and reading stories, strategies, explanations, and lexical hints. Remixing Narrative Strategies (purple circles) includes actions such as manipulating strategies and tracks within the Remixer, as well as revising or continuing the story.

Table 1: Survey results of perceived experience on task workload, creativity support, and AI system experience under two conditions, where the Wilcoxon signed-rank paired t-test W -values and p -values (*: $p < .05$, **: $p < .01$, ***: $p < .001$) are reported.

Scales		NARRIX		Baseline		Statistics	
		M	SD	M	SD	W	p
NASA Task Load	Mental	4.42	1.44	4.00	2.05	23.00	.525
	Physical	4.00	1.86	3.00	1.28	34.50	.170
	Temporal	3.83	1.80	4.00	1.86	25.50	.878
	Performance	5.42	1.08	4.17	1.12	56.00	.043*
	Effort	3.67	1.23	4.33	1.16	2.00	.089
	Frustration	2.58	1.24	3.83	1.34	2.50	.019*
Creativity Support Index	Enjoyment	6.25	.87	3.58	1.78	45.00	.009**
	Exploration	6.08	1.00	3.42	1.88	52.00	.014*
	Expressiveness	6.25	.75	4.33	1.67	45.00	.009**
	Immersion	5.08	1.50	3.08	1.56	41.00	.032*
	Results Worth Effort	6.00	.95	4.17	1.53	63.00	.008**
AI Experience	Match Goal	5.92	.90	4.17	1.40	43.00	.017*
	Think Through	5.67	1.07	2.92	1.68	53.00	.010*
	Transparent	6.00	.95	2.67	1.50	78.00	.002**
	Controllable	6.00	1.13	3.33	1.56	63.00	.008**
	Collaborative	6.33	.78	3.42	1.62	55.00	.006**

strategies, explanations, and lexical hints), and Remixing (e.g., manipulating strategies and tracks within the Remixer workspace, as well as revising or continuing the story) Narrative Strategies.

Most participants (P01, P03-P05, P07-P10) frequently switched among Explore, Learn, and Remix throughout their session, **engaging in a cyclical process of trying, checking, and adjusting rather than following a linear pipeline**. On average, Remix (702.56s, 48.56%) was the most frequent and widely distributed activity, particularly in the mid-to-late stages of the session, serving

as the backbone with exploration and learning woven around it. Exploration (460.79s, 31.85%) typically appeared early or in bursts to spark new ideas, often preceding or punctuating remix phases. Learning (283.50s, 19.59%) occurred in shorter, episodic clusters between remix phases.

5.1.2 Perceived Usability. We computed System Usability Scale (SUS) scores using the UMUX-LITE. Both our system and Baseline received reasonable usability ratings, with NARRIX scoring 82.64

(typically considered “good” [4]) and Baseline scoring 59.03 (typically considered “ok”). A Wilcoxon signed-rank test showed that the SUS score for NARRIX was significantly higher than that of Baseline ($W = 57.00, p = .036^*$).

5.1.3 Perceived Cognitive Load. The overall perceived workload, calculated by averaging all six raw NASA-TLX scores (with the “Performance” measure inverted), did not differ significantly ($W = 32.50, p = .638$) between NARRIX ($M_N = 3.35, SD_N = .87$) and Baseline ($M_B = 3.67, SD_B = 1.23$). However, participants reported being significantly more satisfied with their performance ($W = 56.00, p = .043^*$) when using NARRIX ($M_N = 5.42, SD_N = 1.08$) compared to Baseline ($M_B = 4.17, SD_B = 1.12$). They also perceived significantly less frustration ($W = 2.50, p = .019^*$) with NARRIX ($M_N = 2.58, SD_N = 1.24$) than with Baseline ($M_B = 3.83, SD_B = 1.34$).

5.1.4 Perceived Creativity Support. Participants rated NARRIX significantly higher than Baseline across multiple CSI dimensions. They reported greater enjoyment ($M_N = 6.25, SD_N = .87$ vs. $M_B = 3.58, SD_B = 1.78, W = 45.00, p = .009^*$) and immersion ($M_N = 5.08, SD_N = 1.50$ vs. $M_B = 3.08, SD_B = 1.56, W = 41.00, p = .032^*$). NARRIX also supported exploration ($M_N = 6.08, SD_N = 1.00$ vs. $M_B = 3.42, SD_B = 1.88, W = 52.00, p = .014^*$) and expressiveness ($M_N = 6.25, SD_N = .75$ vs. $M_B = 4.33, SD_B = 1.67, W = 45.00, p = .009^*$) more effectively than Baseline. Participants further felt their results were more worth the effort ($M_N = 6.00, SD_N = .95$ vs. $M_B = 4.17, SD_B = 1.53, W = 63.00, p = .008^*$). Together, these results indicate that NARRIX not only facilitated story writing but also enriched the creative process by making it more engaging, exploratory, and expressive.

5.1.5 Perceived AI Experience. Participants also perceived NARRIX as a more effective and collaborative AI partner. They felt the system better matched their goals ($M_N = 5.92, SD_N = .90$ vs. $M_B = 4.17, SD_B = 1.40, W = 43.00, p = .017^*$) and encouraged deeper thinking ($M_N = 5.67, SD_N = 1.07$ vs. $M_B = 2.92, SD_B = 1.68, W = 53.00, p = .010^*$). Ratings of transparency ($M_N = 6.00, SD_N = .95$ vs. $M_B = 2.67, SD_B = 1.50, W = 78.00, p = .002^*$) and controllability ($M_N = 6.00, SD_N = 1.13$ vs. $M_B = 3.33, SD_B = 1.56, W = 63.00, p = .008^*$) were also significantly higher. Importantly, participants highlighted a stronger sense of collaboration with NARRIX ($M_N = 6.33, SD_N = .78$ vs. $M_B = 3.42, SD_B = 1.62, W = 55.00, p = .006^*$). Overall, NARRIX was experienced as a more collaborative, steerable partner, clarifying what the AI was doing and enabling users to direct assistance toward their narrative intents.

5.2 RQ2: Learning Narrative Strategies with NARRIX

5.2.1 Quantitative Findings. Participants perceived NARRIX as significantly more effective in helping them identify narrative strategies in examples [DG-M] ($M_N = 6.58, SD_N = .52$ vs. $M_B = 1.92, SD_B = 1.51; W = 78.00, p = .002^*$) and understand those strategies in context [DG-C] ($M_N = 6.08, SD_N = 1.17$ vs. $M_B = 1.67, SD_B = .99; W = 78.00, p = .002^*$).

How participants retained strategies after exposure from the tool: We examined participants’ performance in information retention (Fig. 9A). The results showed that participants recalled significantly more narrative strategies ($W = 78.00, p = .002^*$) after using NARRIX ($M_N = 4.08, SD_N = 1.51$) than after using Baseline ($M_B = 1.00, SD_B = .95$). When considering the quality of their descriptions of the recalled strategies (i.e., their understanding of the strategies), participants using NARRIX ($M_N = 2.75, SD_N = 1.47$) also achieved significantly higher scores ($W = 66.00, p = .004^*$) than with Baseline ($M_B = .83, SD_B = .91$). They further reported feeling significantly more confident in their understanding of narrative strategies when supported by NARRIX ($M_N = 5.33, SD_N = 1.16$) compared to Baseline ($M_B = 2.58, SD_B = 1.51; W = 66.00, p = .004^*$).

5.2.2 Qualitative Findings. Next, we report the key findings identified from our qualitative analysis of interviews regarding how NARRIX helped participants learn narrative strategies in story writing.

Making tacit strategies explicit and nameable: Participants consistently emphasized that NARRIX surfaced strategies they had previously used only “by feel,” making them visible and nameable. For example, P01 remarked, “before I relied on intuition, but it extracts the strategies and lets me examine them more rationally.” Similarly, P05 explained: “It let me learn... stories are split into small segments with strategies... I didn’t know these strategies before. As a novice, without this assistance I wouldn’t know where to start.” Naming strategies was seen as the first step toward deliberate learning and transfer.

Supporting intentional and systematic learning: Participants emphasized that explicit strategy extraction, filtering, and highlighting allowed them to study examples purposefully rather than passively. P02 reflected that “selecting strategies itself was a learning process... it extracts narrative strategies and I can consciously decide which to learn and use.” P04 praised the ability to click on a strategy “to see its description and better understand its definition,” and P10 highlighted that strategies were “clearly explained, with text passages showing where each strategy applied.” Others stressed that story arcs made structural choices concrete and helped them link strategies to narrative placement (e.g., P03, P04, P10, P11). As P04 described, “the curve was especially helpful... I could directly see which sentiment I preferred, and clicking a strategy showed the corresponding section, which deepened my understanding of how it was defined.”

Modeling and coaching strategies through structured exemplars: Several participants described NARRIX as “like a teacher” (P03) or “like a mentor” (P09) that guided them with structured exemplars. Pre-segmented examples reduced the cognitive burden of reading long texts (P07), while visualizations such as story arcs provided scaffolding by making story structure and emotional flow visible (e.g., P09, P12). This teacher-like quality reflects the principles of cognitive apprenticeship: modeling strategies explicitly and coaching learners through contextual examples. In contrast, the chat baseline was often seen as simply producing text without fostering these learning processes (P11). P12 similarly noted that Baseline “does many steps at once, writing directly to the end,” which lowered their involvement.

Fostering long-term awareness and transfer: Our findings also show preliminary evidence of longer-term benefits of naming and

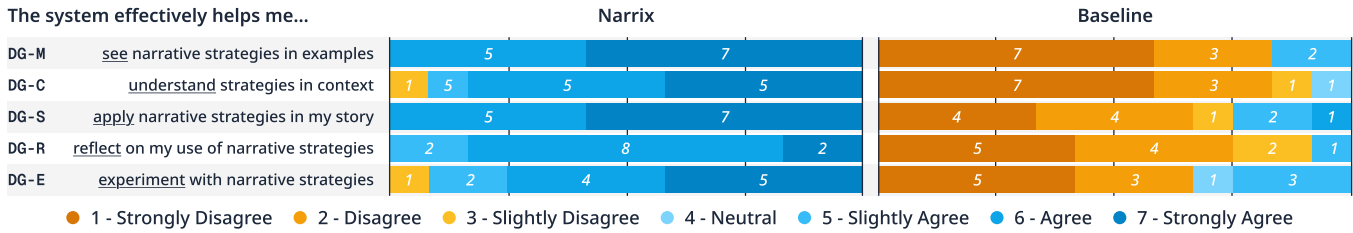


Figure 8: Participants’ responses to a 7-point Likert-scale questionnaire, assessing perceived support across five design goals (see §3.1) in both the Baseline condition and our system.

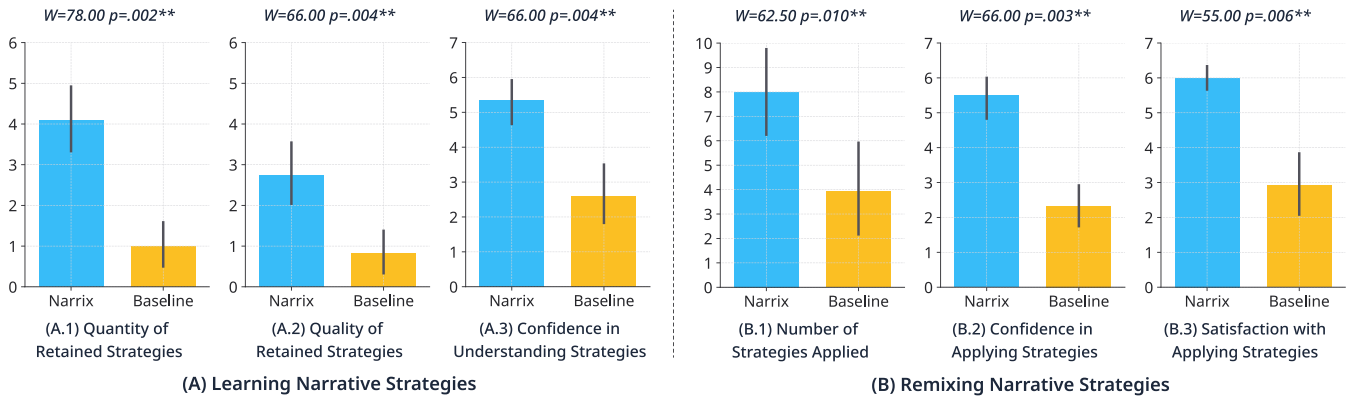


Figure 9: Bar plots illustrating the statistical metrics of participant performance in (A) learning narrative strategies and (B) remixing narrative strategies, where the t-values from the Student’s paired t-test, W-values from the Wilcoxon signed-rank paired test, and p-values (*: $p < .05$, **: $p < .01$, *: $p < .001$) are reported. Error bars represent 95% confidence intervals (CIs).**

practicing strategies. P02 explained that “*next time I’ll think to apply certain strategies again.*” Similarly, P03, P06, P07, and P10 believed that repeatedly seeing strategies classified and applied would promote durable awareness and skill transfer in their own future writing. As P03 reflected, “*With longer use my skills would improve... I’d become more aware of specific strategies I may already use but didn’t know by name.*” Echoing this, P07 remarked, “*...in the future, when I read novels, I’ll pay more attention to which strategies the author is using.*” Future long-term deployment studies could further quantitatively examine how writers retain and apply the strategies they learn from NARRIX.

5.3 RQ3: Remixing Narrative Strategies with NARRIX

5.3.1 Quantitative Findings. Participants perceived NARRIX as significantly more effective in supporting them across three dimensions of remixing (Fig. 8). It helped them apply narrative strategies in their stories [DG-S] ($M_N = 6.58, SD_N = .52$ vs. $M_B = 2.58, SD_B = 1.78$; $W = 66.00, p = .003^{***}$), reflect on their usage of strategies [DG-R] ($M_N = 6.00, SD_N = .60$ vs. $M_B = 2.00, SD_B = 1.21$; $W = 78.00, p = .004^{**}$), and experiment with different combinations of strategies [DG-E] ($M_N = 6.00, SD_N = 1.21$ vs. $M_B = 2.50, SD_B = 1.73$; $W = 66.00, p = .004^{**}$).

To objectively evaluate the writing quality of stories produced using the two systems, we employed the LAMP-P-WRITING-QUALITY-RM model [8], which was trained on expert preference data and

has been shown to align with expert judgments. We used the model to conduct pairwise comparisons between all stories written with NARRIX and those from the baseline system (12 stories in each condition, yielding 144 cross-set comparisons). NARRIX won 107 out of 144 comparisons (74.3%), and a binomial test confirmed that stories produced with NARRIX were significantly better than those written using the baseline system ($p < .001^{***}$).

How participants invoked the tool for remixing: As shown in Fig. 9, participants using NARRIX ($M_N = 8.00, SD_N = 3.25$) applied significantly more narrative strategies than those using Baseline ($M_B = 3.92, SD_B = 3.53$; $W = 62.50, p = .010^{**}$) when steering the AI to revise or continue their stories. Participants also reported feeling more confident in applying strategies with NARRIX ($M_N = 5.33, SD_N = 1.16$ vs. $M_B = 2.58, SD_B = 1.51$; $W = 66.00, p = .003^{**}$) and more satisfied with their performance in integrating strategies into their writing ($M_N = 6.00, SD_N = .60$ vs. $M_B = 2.92, SD_B = 1.83$; $W = 55.00, p = .006^{**}$). Participants typically drafted their stories themselves and used NARRIX primarily to revise their own text. On average, they invoked the revision feature 5.42 times ($SD = 3.29$) per story, versus only 1.83 uses ($SD = 2.41$) of the AI continue function. Put differently, about 75% of NARRIX interactions started from user-written text followed by AI revisions, while only 25% were direct AI continuations. This pattern indicates that participants used NARRIX mainly to explore and apply narrative strategies to improve their drafts rather than delegating content generation to the AI.

Table 2: Usage of tracks across creative dimensions in the Remixer workspace of NARRIX. Row 1 shows the number of tracks per dimension added by all participants; Row 2 shows the number of distinct strategies added within those tracks from all participants.

	Plot	Character	Information	Emotional	Linguistic	Pacing	Thematic	Engagement	Total
# Tracks	9 (23.68%)	6 (15.79%)	3 (7.89%)	10 (26.32%)	3 (7.89%)	3 (7.89%)	1 (2.63%)	3 (7.89%)	38
# Strategies	29 (30.21%)	19 (19.79%)	5 (5.21%)	18 (18.75%)	10 (10.42%)	5 (5.21%)	0 (0.00%)	10 (10.42%)	96

What dimensions were remixed in the workspace: As summarized in Table 2, participants created 38 total tracks in the Remixer (3.17 per participant) and added 96 distinct strategies overall (8.00 per participant). Track usage was concentrated in the Emotional (26.32%) and Plot (23.68%) dimensions, followed by Character (15.79%), with smaller but notable use of Information, Linguistic, Pacing, and Engagement (each 7.89%). Strategy additions were led by Plot (30.21%), followed by Character (19.79%) and Emotional (18.75%), with Linguistic and Engagement contributing 10.42% each. Information and Pacing were less frequently used (5.21% each), and Thematic showed minimal uptake (1 track, 0 strategies). Overall, participants focused their remixing on plot turns and emotional shaping, while thematic scaffolding was rarely used.

5.3.2 Qualitative Findings. Next, we present key findings from our qualitative analysis of interviews regarding how NARRIX supported participants in remixing narrative strategies during story writing.

Encouraging directed yet open-ended creative exploration: Participants described NARRIX as expanding idea space and enabling more expressive outcomes. Curated, strategy-tagged examples “*remind[ed] me of techniques and spark[ed] ideas*” (P01), while simplifying text into selectable strategies yielded “*faster, higher-signal ideation*” than reading long generations (P02). Several participants reported concrete gains in expressiveness—e.g., automatically calling back earlier motifs for foreshadowing at the ending (P03)—and cited the arc visualizations as a prompt for directions they would not have considered otherwise (P09, P12). Overall, NARRIX made creative exploration feel directed yet open-ended.

Connecting high-level story goals with concrete writing techniques: Participants selected narrative strategies largely based on their storytelling goals. They typically followed a three-step process, starting from (1) using the story-arc visualization to identify examples that pursued similar goals (e.g., turning points, emotional shifts), (2) reading those examples to understand which narrative strategies were used and how, and (3) applying the same strategies in their own drafts to test whether they achieved the desired effect. For example, P08 described planning the overall style first and then selecting strategies to realize it: “*I first think about the style and whether the plot should stay steady or rise and fall... then I look for something that matches that plan and put it into my story.*” P12 used the story arc as feedback to evaluate whether applying a certain strategy shifted the narrative in the intended direction: “*The visualization was rising, but after I rewrote it, it turned downward... the chart made that visible.*” P10 studied how exemplar blocks deploy strategies at specific structural moments (e.g., a happy opening) to identify similar techniques to use: “*If the opening is happy, I find*

the point closest to 1.0 on the arc, then look for strategies there to understand how they’re used.”

Scaffolding targeted, controllable remixing by tracks and block-based edits: Participants used tracks to map priorities (e.g., plot, character, emotion) and to “*spec*” tactics per scene, then invoked block-scoped generation for targeted local changes (e.g., P02, P04, P05, P06, P08, P12). They described dragging strategies tied to the current scene intent, performing segment-level rewrites, and immediately inspecting effects (often via the story arc). For example, P06 emphasized the value of block-scoped rewrites for fine-grained control: “*I would pick a strategy and apply it only to a segment, then rewrite that part... this gave me more control than letting the AI regenerate everything.*” Several highlighted the immediate feedback of the emotion curve, as P12 explained: “*I could see how the curve changed after a rewrite, which made it clear whether the strategy worked.*”

Enhancing controllability with block-level, strategy-steered generation: Participants contrasted the chat baseline’s *summarize-generate* workflow with NARRIX’s block-level, strategy-based steering. They noted that chat was good for quick pre-writing or one-shot follow-ups but offered low strategy salience and limited fine-grained control (e.g., P02, P04, P07, P08, P09, P11, P12). With chat, they often fed an outline or whole draft and then pruned unintended additions; with NARRIX, they targeted specific passages and effects using named strategies (e.g., P04, P06). Several emphasized that NARRIX reduced prompt guesswork and clean-up while increasing ownership of the revision process; chat felt more like “*AI writes for you*” (P08, P12).

6 Discussion

In this paper, we propose and evaluate NARRIX, an interactive system that supports novice writers in discovering, interpreting, and remixing narrative strategies from examples into their own stories. Based on findings from system design and user evaluation, we discuss key design implications for future writing interfaces and creativity support tools.

6.1 Reifying Tacit Knowledge into Visible, Reusable Units

Our results show that making tacit techniques explicit, nameable, and placeable turns “*writing by feel*” into deliberate practice (§5.2.2) and targeted remix (§5.3.2). This design pattern generalizes to other creative domains (e.g., UI design). For instance, Misty [56] supports aspect-level blending (e.g., color, layout, content) from exemplars into work-in-progress designs; while it does not explicitly name strategies, our findings suggest that exposing techniques as explicit units would further strengthen learning, control, and reuse. To this

end, future interfaces could: (i) extract and label tacit techniques directly in examples and user work; (ii) bind each technique to where it occurs (e.g., section/beat, UI element) and why it works (intended effect); (iii) represent techniques as lightweight, manipulatable tokens (e.g., chips on tracks) that can be inserted, reordered, or removed within a work; and (iv) preview local impact before committing (e.g., emotion/arc deltas in writing; simulated user interactions in UI).

6.2 Building Personal “Strategy Libraries” for Transfer and Long-Term Reuse

Participants reported growing awareness and reuse intentions after naming and practicing strategies (§5.2.2). Future systems could therefore support “personal strategy libraries”: save any strategy the user tries (or edits) along with micro-examples, before/after diffs, and tags for goal, tone, and placement. Beyond a single writing sessions, future tools could also let users collect strategies during everyday reading, Pinterest-style [55]. Concretely: provide a lightweight reader/extension to “pin” snippets from the web, PDFs, or e-books; run the paper’s techniques (see §3.4) in the capture flow (e.g., automatic strategy extraction, sentence-level highlights, and affect/arc estimation) to store each pin as a strategy token with provenance (source, author), context (where in the narrative it occurs), intended effect, and representative text. Over time, surface analytics to reveal patterns (e.g., what techniques the author tends to use/avoid) and suggest complementary or under-used strategies. This will support long-term skill development, enabling users to build repertoire across sessions and genres.

6.3 Steering Generation with Abstract Strategies and Concrete Intent

§5.3.2 shows users prefer to state an effect (goal-driven) or browse and try (exploratory), then place strategies at specific locations. Future work could further support steering content generation by combining abstract strategies and concrete intent specifications. For example, tools could (i) accept intent specifications (desired effect, narrative position, constraints) and compile them into strategy bundles that guide localized generation; (ii) expose both abstract knobs (like in [15] and [57]) and concrete moves (e.g., “*echo earlier image*,” “*interiority: sensory detail + self-talk*”) so users can steer at the right level; (iii) enable fast retrieval by intent (e.g., “*raise stakes at midpoint*,” “*amplify interiority*”) and recommend previously successful moves when similar contexts recur, helping writers reuse and adapt strategies effectively across different drafts.

6.4 Linking Local Strategies to Global Story Structure

NARRIX primarily emphasizes surfacing strategy applied at the paragraph or block level, which is an intentional first step that gives novice writers more manageable techniques to experiment with before tackling higher-level story structure. While the story-arc visualization provides some connection between local and global reasoning, helping writers reflect on emotional trajectories and structural turning points (§5.3.2), participants sometimes noticed

small inconsistencies when strategies were combined across sections. Because broader narrative coherence depends on linking local moves to global story logic (e.g., character arcs, causal progression, and structures like the Hero’s Journey [62]), a natural next step is to extend our block-scoped, controllable design toward multi-level coherence. Concretely, we envision future extensions on hierarchical tracks (like in [62]) that tie block-level strategies to overarching trajectories, cross-block coherence checks that flag drift or redundancy, and visual summaries of global patterns (such as cumulative emotional arcs or character development) to help writers see how local revisions accumulate into story-wide structure.

6.5 Blending Chat for Pre-Writing with Controllable Remix for Refinement

Participants saw value in both chat-style interfaces (for fast ideation and upfront drafting) and NARRIX’s strategy-steered editing (for precise and fine-grained control) (§5.3.2). Future systems could *hybridize* the strengths: (i) a Pre-write/Chat space for rapid ideation and scaffolding (summaries, beat lists), followed by (ii) a Remix/Edit space where strategy tokens are placed on tracks and applied locally; (iii) provide a seamless handoff (import outline → auto-seed tracks and candidate strategies), and a reversible “promote to draft / demote to sketch” mechanism so users can fluidly move between broad strokes and fine control; and (iv) default to local-first edits to reduce unintended global changes in chat-only one-shot rewrites.

6.6 Supporting Cognitive Apprenticeship with AI for Long-Term Learning

NARRIX operationalizes the five components of cognitive apprenticeship [18]: modeling, coaching, scaffolding, reflection, and exploration (§3.1 and Fig. 8). Previous work has shown that cognitive apprenticeship can foster long-term skill development and knowledge retention [17, 88] and shows promise in computer-based learning environments [32]. Building on this, we can expect that future writing tools like NARRIX can position the AI not merely as a co-author that generates text, but as a long-term “cognitive mentor” for users that externalizes reasoning and augments long-term writing skill development. For example, each suggestion should carry a rationale (modeling), surface provenance from examples (coaching), and support quick “apply this effect” (scaffolding) or “show alternatives” (exploration) actions. Lightweight prompts to encourage reflection, like in [98], would further help users develop strategy awareness rather than deferring entirely to AI outputs. Finally, logging applied strategies into a personal library (§6.2) could transform momentary co-creation into cumulative learning resources and gains, aligning with the apprenticeship cycle of moving from guided observation to independent, strategic practice.

7 Limitations

Our system design has several limitations. First, despite our mitigation strategies, LLM errors cannot be fully eliminated from our technical pipeline due to the statistical nature of these models. We encourage future work to enhance our methods for extracting narrative strategies, such as by establishing benchmarks, developing specialized models, and post-training LLMs to better align with expert narrative analysts. Second, while NARRIX makes individual

strategies explicit and manipulatable, it offers limited support for global-level strategies, such as maintaining coherence across multiple sections or harmonizing style across blocks. Some participants noted inconsistencies when combining strategies across tracks or blocks, suggesting the need for better mechanisms to manage cross-cutting narrative patterns. Third, although the current interface allows users to apply strategies iteratively, it offers limited functionality for comparing different combinations in parallel. Future iterations could address this gap by enabling side-by-side prototyping and evaluation of multiple strategy combinations. Finally, writing excessively long stories (such as a 50-page novel) remains a challenge; future iterations could incorporate hierarchical tracks (from chapter to scene to paragraph), batched apply/rollback of strategies, and cross-block coherence checks (style, voice, references) to keep long-form revision tractable.

Our study methodology also has limitations. First, we recruited a relatively small sample of 12 participants, all non-native English speakers recruited from a single software organization. Additionally, to manage potential confounds and keep sessions within time limits, we controlled task materials with preset across conditions. The fairy-tale examples used are culturally homogeneous, which may not represent the diversity of narratives that writers engage with, limiting the generalizability of our findings to broader global contexts. While we computed statistical significance, we do not claim the results to be conclusive but should be interpreted as promising but preliminary. The specific demographics of our participants may also have influenced how they interacted with the system. As non-native English speakers, who are typically recruited as novice writers in prior work [36, 98], they might have benefited more from the system's explicit scaffolding of narrative strategies and lexical cues, compared to native speakers who could draw more intuitively on their linguistic and cultural repertoire. Similarly, the use of well-known fairy tales, while helpful for controlling familiarity, may have biased participants toward recognizing conventional Western narrative structures and emotional arcs. Future work should examine how NARRIX performs with a more diverse population of writers and story corpora, including different genres, languages, and cultural traditions, to better understand how narrative strategy learning and remixing generalize across contexts.

Finally, our study only provides initial evidence of users' short-term learning gains in awareness and retention. To understand the longer-term impact of NARRIX on writing skill development (e.g., long-term retention of strategies, transfer to independent writing), we plan to conduct a longitudinal study that follows participants over multiple weeks or months, tracking how their strategy use evolves across repeated writing tasks and whether they continue to apply and adapt learned strategies in their independent practice.

8 Conclusion

We presented NARRIX, a writing interface that makes narrative strategies in examples visible, understandable, and ready to remix. Our preliminary user study with 12 novice writers suggests that NARRIX shows promise in shifting AI-assisted writing away from one-shot content generation toward a more deliberate, transparent, and controllable practice, where examples become teachable

moments and edits become opportunities to learn the craft. Looking forward, we envision future writing tools that not only help authors produce text, but also cultivate their narrative repertoire, supporting long-term creative growth, transferable skills, and more collaborative forms of human-AI co-creation.

Acknowledgments

We sincerely thank our participants for sharing their feedback on our system, and we are grateful to all reviewers for their valuable insights and suggestions. This research was conducted during an internship at Adobe Research. We thank the Adobe Research EEL group for their guidance and support throughout this work.

References

- [1] Tazin Afrin, Omid Kashefi, Christopher Olshefski, Diane Litman, Rebecca Hwa, and Amanda Godley. 2021. Effective Interfaces for Student-Driven Revision Sessions for Argumentative Writing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3411764.3445683
- [2] Prithviraj Ammanabrolu, Wesley Cheung, William Broniec, and Mark O. Riedl. 2020. Automated Storytelling via Causal, Commonsense Plot Ordering. arXiv:2009.00829 [cs] doi:10.48550/arXiv.2009.00829
- [3] Mieke Bal. 2004. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, Scholarly Publishing Division, Toronto.
- [4] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies* 4, 3 (May 2009), 114–123.
- [5] Donald J. Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (AAIWS'94)*. AAAI Press, Seattle, WA, 359–370.
- [6] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa
- [7] 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. doi:10.1007/3-540-45420-9_16
- [8] 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. arXiv:2504.07532 [cs] doi:10.48550/arXiv.2504.07532
- [9] Jim Chang and Jason Chang. 2015. WriteAhead2: Mining Lexical Grammar Patterns for Assisted Writing. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. Association for Computational Linguistics, Denver, Colorado, 106–110. doi:10.3115/v1/N15-3022
- [10] Davida H. Charney and Richard A. Carls. 1995. Learning to Write in a Genre: What Student Writers Take from Model Texts. *Research in the Teaching of English* 29, 1 (Feb. 1995), 88–125. doi:10.58680/rte199515358
- [11] 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. doi:10.1145/2617588
- [12] Jean-Peic Chou, Alexa Fay Siu, Nedim Lipka, Ryan Rossi, Franck Démoncourt, 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. doi:10.1145/3586183.3606807
- [13] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3491102.3501819
- [14] 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. doi:10.1145/3491102.3501819
- [15] 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. doi:10.1145/3654777.3676352

- [16] Domenic V. Cicchetti. 1994. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment* 6, 4 (1994), 284–290. doi:10.1037/1040-3590.6.4.284
- [17] Allan Collins, John Seely Brown, and Ann Holum. 1991. Cognitive Apprenticeship: Making Thinking Visible. *American educator* 15, 3 (1991), 6–11.
- [18] Allan Collins and Manu Kapur. 2006. *Cognitive Apprenticeship*. The Cambridge Handbook of the Learning Sciences, Vol. 291. Cambridge University Press, Cambridge, UK.
- [19] Hai Dang, Chelse Swoopes, Daniel Buschek, and Elena L. Glassman. 2025. CorpusStudio: Surfacing Emergent Patterns In A Corpus Of Prior Work While Writing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3706598.3713974
- [20] 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 [cs]
- [21] Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason Weston. 2024. Chain-of-Verification Reduces Hallucination in Large Language Models. In *Findings of the Association for Computational Linguistics: ACL 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 3563–3578. doi:10.18653/v1/2024.findings-acl.212
- [22] Dana Lynn Driscoll, Joseph Paszek, Gwen Gorzelsky, Carol L. Hayes, and Edmund Jones. 2020. Genre Knowledge and Writing Development: Results From the Writing Transfer Project. *Written Communication* 37, 1 (Jan. 2020), 69–103. doi:10.1177/0741088319882313
- [23] Alexandre Duval, Thomas Lamson, Gaël de Léséleuc de Kérouara, and Matthias Gallé. 2021. Breaking Writer’s Block: Low-Cost Fine-Tuning of Natural Language Generation Models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, Dimitra Gkatzia and Djamel Seddah (Eds.). Association for Computational Linguistics, Online, 278–287. doi:10.18653/v1/2021.eacl-demos.33
- [24] Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical Neural Story Generation. arXiv:1805.04833 [cs] doi:10.48550/arXiv.1805.04833
- [25] Robert L. Fowler and Anne S. Barker. 1974. Effectiveness of Highlighting for Retention of Text Material. *Journal of Applied Psychology* 59, 3 (1974), 358–364. doi:10.1037/h0036750
- [26] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, Nevada, USA, 2414–2423.
- [27] 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. doi:10.1145/3532106.3533533
- [28] Pablo Gervás, Belén Díaz-Agudo, Federico Peinado, and Raquel Hervás. 2005. Story Plot Generation Based on CBR. *Knowledge-Based Systems* 18, 4 (Aug. 2005), 235–242. doi:10.1016/j.knosys.2004.10.011
- [29] Steve Graham, Charles A. MacArthur, and Jill Fitzgerald (Eds.). 2007. *Best Practices in Writing Instruction*. The Guilford Press, New York, NY, US. xii, 340 pages.
- [30] Steve Graham and Dolores Perin. 2007. *Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools*. Alliance for Excellent Education, Washington, D.C., US.
- [31] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, Amsterdam, The Netherlands, 139–183. doi:10.1016/S0166-4115(08)62386-9
- [32] Sara Hennessy. 1993. Situated Cognition and Cognitive Apprenticeship: Implications for Classroom Learning. *Studies in Science Education* 22, 1 (Jan. 1993), 1–41. doi:10.1080/03057269308560019
- [33] Md Naimul Hoque, Tasfia Mashiat, Bhavya Ghai, Cecilia D. Shelton, Fanny Chevalier, Kari Kraus, and Niklas Elmqvist. 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. doi:10.1145/3613904.3641895
- [34] 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. doi:10.1145/3313831.3376715
- [35] Yichen Huang, Yizhe Zhang, Oussama Elachqar, and Yu Cheng. 2020. INSET: Sentence Infilling with Inter-Sentential Transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 2502–2515. doi:10.18653/v1/2020.acl-main.226
- [36] Yi-Ching Huang, Hao-Chuan Wang, and Jane Yung-jen Hsu. 2018. Feedback Orchestration: Structuring Feedback for Facilitating Reflection and Revision in Writing. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18 Companion)*. Association for Computing Machinery, New York, NY, USA, 257–260. doi:10.1145/3272973.3274069
- [37] Julie Hui and Michelle L. Sprouse. 2023. Lettersmith: Scaffolding Written Professional Communication Among College Students. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3581029
- [38] Julie S. Hui, Darren Gergle, and Elizabeth M. Gerber. 2018. IntroAssist: A Tool to Support Writing Introductory Help Requests. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173596
- [39] Daphne Ippolito, David Grangier, Chris Callison-Burch, and Douglas Eck. 2019. Unsupervised Hierarchical Story Infilling. In *Proceedings of the First Workshop on Narrative Understanding*, David Bamman, Snigdha Chaturvedi, Elizabeth Clark, Madalina Fiterau, and Mohit Iyyer (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 37–43. doi:10.18653/v1/W19-2405
- [40] Daphne Ippolito, David Grangier, Chris Callison-Burch, and Douglas Eck. 2019. Unsupervised Hierarchical Story Infilling. In *Proceedings of the First Workshop on Narrative Understanding*, David Bamman, Snigdha Chaturvedi, Elizabeth Clark, Madalina Fiterau, and Mohit Iyyer (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 37–43. doi:10.1145/v1/W19-2405
- [41] Takumi Ito, Tatsuki Kuribayashi, Masatoshi Hidaka, Jun Suzuki, and Kentaro Inui. 2020. Langsmith: An Interactive Academic Text Revision System. arXiv:2010.04332 [cs]
- [42] 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. doi:10.1145/3586183.3606810
- [43] 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. doi:10.1145/3544548.3581196
- [44] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* 55, 12 (March 2023), 248:1–248:38. doi:10.1145/3571730
- [45] Hyeon-su B. Kang, Gabriel Amoako, Neil Sengupta, and Steven P. Dow. 2018. Paragon: An Online Gallery for Enhancing Design Feedback with Visual Examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. doi:10.1145/3173574.3174180
- [46] X. J. Kennedy and Dana Gioia. 2016. *Literature: An Introduction to Fiction, Poetry, Drama, and Writing*. Pearson, Boston.
- [47] 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. doi:10.1145/2998181.2998196
- [48] Taewook Kim, Hyomin Han, Eytan Adar, Matthew Kay, and John Joon Young Chung. 2024. Authors’ Values and Attitudes Towards AI-Driven Scalable Personalization of Creative Language Arts. arXiv:2403.00439 [cs] doi:10.1145/3613904.3642529
- [49] Michael Lebowitz. 1984. Creating Characters in a Story-Telling Universe. *Poetics* 13, 3 (1984), 171–194.
- [50] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R. Klemmer. 2010. Designing with Interactive Example Galleries. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2257–2266. doi:10.1145/1753326.1753667
- [51] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambagsans, 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 Sergeevuk, 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. arXiv:2403.14117 [cs] doi:10.1145/3613904.3642697
- [52] 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. doi:10.18653/v1/2022.in2writing-1.9

- [53] Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2016. Rationalizing Neural Predictions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Jian Su, Kevin Duh, and Xavier Carreras (Eds.). Association for Computational Linguistics, Austin, Texas, 107–117. doi:10.18653/v1/D16-1011
- [54] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. 2013. UMUX-LITE: When There's No Time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2099–2102. doi:10.1145/2470654.2481287
- [55] Rhema Linder, Clair Snodgrass, and Andriuid Kerne. 2014. Everyday Ideation: All of My Ideas Are on Pinterest. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2411–2420. doi:10.1145/2556288.2557273
- [56] Yuwen Lu, Alan Leung, Amanda Swearingin, Jeffrey Nichols, and Titus Barik. 2025. Misty: UI Prototyping Through Interactive Conceptual Blending. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. arXiv:2409.13900 [cs] doi:10.1145/3706598.3713924
- [57] Zhuoran Lu, Qian Zhou, and Yi Wang. 2025. WhatELSE: Shaping Narrative Spaces at Configurable Level of Abstraction for AI-Bridged Interactive Storytelling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3706598.3713363
- [58] Deena Mohammad Mansour. 2017. Using COCA to Foster Students' Use of English Collocations in Academic Writing. In *Proceedings of the 3rd International Conference on Higher Education Advances*. Universitat Politècnica València, Valencia, Spain, 600–607. doi:10.4995/HEAD17.2017.5301
- [59] Damien Masson, Young-Ho Kim, and Fanny Chevalier. 2025. Textshop: 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. doi:10.1145/3706598.3713862
- [60] Robert McKee. 1997. *Story: Style, Structure, Substance, and the Principles of Screenwriting*. Dey Street Books, New York.
- [61] James R. Meehan. 1977. TALE-SPIN, an Interactive Program That Writes Stories. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 1 (IJCAI'77)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 91–98.
- [62] 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. doi:10.1145/3544548.3581225
- [63] Saif M. Mohammad. 2025. NRC VAD Lexicon v2: Norms for Valence, Arousal, and Dominance for over 55k English Terms. arXiv:2503.23547 [cs] doi:10.48550/arXiv.2503.23547
- [64] Donald M. Murray. 2004. *A Writer Teaches Writing Revised*. Cengage Learning, Boston, MA.
- [65] M. Asif Naeem, I. Wayan S. Linggawa, Aftab A. Mughal, Christof Lutteroth, and Gerald Weber. 2018. A Smart Email Client Prototype for Effective Reuse of Past Replies. *IEEE Access* 6 (2018), 69453–69471. doi:10.1109/ACCESS.2018.2878523
- [66] Tricia J. Ngoon, Joy O Kim, and Scott Klemmer. 2021. Shōwin: Adaptive Conceptual Guidance Aids Example Use in Creative Tasks. In *Designing Interactive Systems Conference 2021 (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1834–1845. doi:10.1145/3461778.3462072
- [67] Pinelopi Papalampidi, Frank Keller, and Mirella Lapata. 2019. Movie Plot Analysis via Turning Point Identification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojuan Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 1707–1717. doi:10.18653/v1/D19-1180
- [68] Rafael Pérez Y Pérez and Mike and Sharples. 2001. MEXICA: A Computer Model of a Cognitive Account of Creative Writing. *Journal of Experimental & Theoretical Artificial Intelligence* 13, 2 (April 2001), 119–139. doi:10.1080/09528130010029820
- [69] Gerald Prince. 2012. *Narratology: The Form and Functioning of Narrative*. Vol. 108. Walter de Gruyter, Berlin, Germany.
- [70] Anyi Rao, Jean-Peïc Chou, and Maneesh Agrawala. 2024. ScriptViz: A Visualization Tool to Aid Scriptwriting Based on a Large Movie Database. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3654777.3676402
- [71] Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The Emotional Arcs of Stories Are Dominated by Six Basic Shapes. *EPJ Data Sci.* 5, 1 (Dec. 2016), 31. doi:10.1140/epjds/s13688-016-0093-1
- [72] Mohi Reza, Nathan Laundry, Ilya Musabirov, Peter Dushniku, Zhi Yuan "Michael" Yu, Kashish Mittal, Tov 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. arXiv:2310.00117 [cs] doi:10.48550/arXiv.2310.00117
- [73] Mohi Reza, Nathan M Laundry, Ilya Musabirov, Peter Dushniku, Zhi Yuan "Michael" Yu, Kashish Mittal, Tov Grossman, Michael Liut, Anastasia Kuzminykh, and Joseph Jay Williams. 2024. ABScribe: Rapid Exploration & Organization of Multiple Writing Variations in Human-AI Co-Writing Tasks Using Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3613904.3641899
- [74] 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.
- [75] Mark O. Riedl and Robert Michael Young. 2010. Narrative Planning: Balancing Plot and Character. *Journal of Artificial Intelligence Research* 39 (2010), 217–268.
- [76] Daniel Ritchie, Ankita Arvind Kejrival, and Scott R. Klemmer. 2011. D.Tour: Style-Based Exploration of Design Example Galleries. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology - UIST '11*. ACM Press, Santa Barbara, California, USA, 165. doi:10.1145/2047196.2047216
- [77] 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. doi:10.1145/3450741.3465253
- [78] Raymond Scupin. 1997. The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology. *Human Organization* 56, 2 (1997), 233–237. jstor:44126786
- [79] 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. doi:10.1145/3584931.3607492
- [80] 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. doi:10.1145/3637371
- [81] Simeng Sun, Wenlong Zhao, Varun Manjunatha, Rajiv Jain, Vlad Morariu, Franck Dernoncourt, Balaji Vasani Srinivasan, and Mohit Iyyer. 2021. IGA: An Intent-Guided Authoring Assistant. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 5972–5985. doi:10.18653/v1/2021.emnlp-main.483
- [82] Christine Tardy. 2009. *Building Genre Knowledge*. Parlor Press LLC, Anderson, South Carolina, USA.
- [83] Christine M. Tardy, Bruna Sommer-Farias, and Jeroen Gevers. 2020. Teaching and Researching Genre Knowledge: Toward an Enhanced Theoretical Framework. *Written Communication* 37, 3 (July 2020), 287–321. doi:10.1177/0741088320916554
- [84] Yufei Tian, Tenghao Huang, Miri Liu, Derek Jiang, Alexander Spangher, Muhao Chen, Jonathan May, and Nanyun Peng. 2024. Are Large Language Models Capable of Generating Human-Level Narratives? arXiv:2407.13248 [cs] doi:10.48550/arXiv.2407.13248
- [85] 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. doi:10.1145/3170427.3188474
- [86] Scott R. Turner. 1993. *Minstrel: A Computer Model of Creativity and Storytelling*. University of California, Los Angeles, Los Angeles, California, USA.
- [87] Qian Wan, Jiannan Li, Huan Chen Wang, and Zhicong Lu. 2025. PolyMind: Parallel Visual Diagramming with Large Language Models to Support Prewriting Through Microtasks. arXiv:2502.09577 [cs] doi:10.48550/arXiv.2502.09577
- [88] Ruimei Wang, Nurul Nadwa Zulkifli, and Ahmad Fauzi Mohd Ayub. 2024. Investigating the Impact of the Stratified Cognitive Apprenticeship Model on High School Students' Math Performance. *Education Sciences* 14, 8 (Aug. 2024), 898. doi:10.3390/educsci14080898
- [89] Su Wang, Greg Durrett, and Katrin Erk. 2020. Narrative Interpolation for Generating and Understanding Stories. arXiv:2008.07466 [cs] doi:10.48550/arXiv.2008.07466
- [90] Sitong Wang, Samia Menon, Dingzeyu Li, Xiaojuan Ma, Richard Zemel, and Lydia B. Chilton. 2025. Schemex: Interactive Structural Abstraction from Examples with Contrastive Refinement. arXiv:2504.11795 [cs] doi:10.48550/arXiv.2504.11795
- [91] T. B. Ward. 1994. Structured Imagination: The Role of Category Structure in Exemplar Generation. *Cognitive Psychology* 27, 1 (Aug. 1994), 1–40. doi:10.1006/cogp.1994.1010
- [92] Jules White, Qu Chen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. In

- Proceedings of the 30th Conference on Pattern Languages of Programs (PLoP '23)*. The Hillside Group, USA, 1–31.
- [93] 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. doi:10.1145/3491102.3517582
- [94] Peng Xu, Mostofa Patwary, Mohammad Shoeybi, Raul Puri, Pascale Fung, Anima Anandkumar, and Bryan Catanzaro. 2020. MEGATRON-CNTRL: Controllable Story Generation with External Knowledge Using Large-Scale Language Models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 2831–2845. doi:10.18653/v1/2020.emnlp-main.226
- [95] Xiaotong (Tone) Xu, Rosaleen Xiong, Boyang Wang, David Min, and Steven P. Dow. 2021. IdeateRelate: An Examples Gallery That Helps Creators Explore Ideas in Relation to Their Own. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 352:1–352:18. doi:10.1145/3479496
- [96] 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. doi:10.1145/3490099.3511105
- [97] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3544548.3581388
- [98] 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. doi:10.1145/3706598.3714316
- [99] 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. doi:10.1145/3746059.3747703
- [100] 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. doi:10.1145/3613904.3642647
- [101] Chao Zhang, Shengqi Zhu, Xinyu Yang, Yu-Chia Tseng, Shenrong Jiang, and Jeffrey M. Rzeszotarski. 2025. Navigating the Fog: How University Students Recalibrate Sensemaking Practices to Address Plausible Falsehoods in LLM Outputs. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3719160.3736618
- [102] 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. doi:10.1145/3586183.3606800

A Definition of Narrative Concepts

A.1 Turning Points

A turning point is an event (or plot moment) that significantly influences a plot progression [67]. These turning points are generally in sequential order in a narrative (i.e., Opportunity happens first; Climax happens last). The definitions are shown in Table 3.

Table 3: Definition of turning points [84].

Turning Point	Description
Opportunity	The introductory event that sets the stage for the narrative.
Change of Plans	A pivotal moment where the main goal of the narrative is defined or altered.
Point of No Return	The commitment point beyond which the protagonists are invested in goals.
Major Setback	A critical juncture where the protagonists face significant challenges or failures.
Climax	The peak of the narrative arc, encompassing the resolution of the central conflict.

A.2 Creative Dimensions

We derived a taxonomy of eight creative dimensions from theories in narrative and writing studies (Table 4). These dimensions capture distinct aspects of narrative construction emphasized across narratology, creative writing pedagogy, and rhetorical narrative theory [3, 46, 60, 69]. This taxonomy is not exhaustive; rather, it offers an interpretive lens for examining the diverse narrative strategies employed in story writing.

Table 4: Definition of creative dimensions [3, 46, 60, 69].

Dimension	Description
Plot	Strategies for plot construction and story progression, e.g., causation, escalation, conflict setup and resolution, reversals and twists, and act and beat frameworks.
Character	Strategies for character development and portrayal, e.g., growth, traits, relationships, and archetypal roles.
Information	Strategies for information control and perspective, e.g., revelation, concealment, misdirection, foreshadowing, and point-of-view manipulation.
Emotional	Strategies for emotional effect, e.g., tension, empathy, surprise, catharsis, and atmosphere.
Linguistic	Strategies for language style, e.g., voice, imagery, syntax and rhythm, dialogue, and rhetorical devices.
Pacing	Strategies for pacing at the moment and segment levels, e.g., scene versus summary, time compression and expansion, beat density, sentence and paragraph cadence, cutaways and cross-cutting, time skips, and arrive-late, leave-early trims.
Thematic	Strategies for theme and meaning, e.g., symbolism, allegory, and philosophical exploration.
Engagement	Strategies for reader engagement, e.g., hooks, immersion techniques, curiosity creation, suspense management, and narrative payoffs.

B LLM Prompts Used in NARRIX

In this section, we present the prompts used to instruct GPTs to segment stories, infer narrative strategies, categorize strategies, infer protagonist, and infer emotional adjectives.

B.1 Segmenting Stories into Blocks

We prompt GPT-4o to segment stories into blocks (Fig. 6A).

<p>System Prompt: You are an expert story analyst. Your task is to segment a given story into coherent plot segments that represent distinct narrative beats.</p> <p>**Segment Criteria:** - Each segment should be self-contained enough to understand independently - Aim for 5-10 segments depending on story length and complexity - Each segment should represent a meaningful story progression - Avoid overly short segments (less than 50 words) or overly long segments (more than 300 words)</p> <p>**Output Format:** Return your response as a JSON object with this exact structure: {"plots": [{"title": "Brief, descriptive title (3-8 words)", "plot": "Original text content from this story segment (extracted verbatim from the source)", "summary": "Concise summary of what happens in this segment"}]}.</p> <p>.....</p> <p>Context Prompt: The story title is <insert story title> and the story is: <insert full story content></p>

B.2 Inferring Story Protagonist

We instruct GPT-4o to identify the main character (Fig. 6D).

<p>System Prompt: Who is the main character of the this story? The output should just be a name or a short phrase. Do not include any other information or context.</p> <p>.....</p> <p>Context Prompt: The story is: <insert full story content></p>

B.3 Inferring Emotional Adjectives

For each content block in a narrative, we ask GPT-4o to infer three adjectives that describe the protagonist's emotions as the plot progresses (e.g., *amused*, *relaxed*, *anxious*) (Fig. 6E).

<p>System Prompt: Use three different words to describe the character's feeling in a given story plot. The output should be a list of words. For example, [happy, sad, joyful]. Do not include other outputs.</p> <p>.....</p> <p>Context Prompt: How does <insert protagonist> feel in this plot? <insert plot content></p>
--

B.4 Inferring Narrative Strategies

We prompt GPT-4.1 to infer narrative strategies (Fig. 6B).

System Prompt:
 You are an expert literary analyst tasked with identifying creative strategies used in story plots. Creative strategies are any storytelling techniques, narrative devices, plot mechanisms, stylistic choices, or structural elements that authors use to create compelling narratives, engage readers, or achieve specific artistic effects.

****Your task:****
 Analyze the given plot and identify all creative strategies employed. For each strategy, provide:
 1. A concise phrase describing the strategy (2-6 words)
 2. A detailed explanation of how and why this strategy is used effectively
 3. Specific lexical features (words, phrases, linguistic patterns) that contribute to or signal this strategy

****What Constitutes a Creative Strategy:****
 Any deliberate creative choice that serves a narrative purpose, including but not limited to:
 - How information is revealed or withheld
 - Character development and interaction patterns
 - Structural and pacing decisions
 - Language and tone choices
 - Conflict creation and resolution approaches
 - Thematic development techniques
 - Reader engagement mechanisms
 - Innovative or unexpected narrative elements

****Lexical Features to Identify:****
 For each strategy, extract EXACT verbatim text from the original plot that contributes to or signals the strategy:
 - Copy exact words, phrases, or sentences as they appear in the plot
 - Include direct quotes from dialogue exactly as written
 - Extract precise descriptive language or imagery
 - Identify specific repeated words or phrases
 - Copy transitional phrases or structural markers verbatim
 - Quote any language that contributes to the strategy's effectiveness
 - Do NOT paraphrase, interpret, or modify the original text - use exact quotations only Return output

****Output Format:****
 Return your response as a JSON object with this exact structure:
 {"strategies":[{"strategy":"Brief strategy name (2-6 words)","reasoning":"1-3 sentence explanation of how this strategy functions and why it's effective.,"lexicon":["word1","phrase2","linguistic pattern3"]}]}.

Context Prompt:
 The story plot is: <insert plot content>

Please analyze this plot and identify all creative strategies employed. Look for any storytelling techniques, narrative choices, or creative elements that serve a purpose in the story - don't limit yourself to traditional categories.

For each strategy you identify:
 1. Name the strategy clearly and concisely
 2. Explain how it functions in the plot and why it's effective
 3. Extract EXACT verbatim words, phrases, or sentences from the plot text above that contribute to or signal this strategy - use precise quotations only, do not paraphrase or modify the original text

Be thorough and creative in your analysis. Consider both obvious techniques and subtle creative choices that make this plot work.

B.5 Categorizing Narrative Strategies

We prompt GPT-4.1 to categorize narrative strategies into the eight creative dimensions in Table 4.

System Prompt:
 You are an expert literary analyst tasked with categorizing creative strategies according to a comprehensive taxonomy. Each strategy should be assigned to one or two primary categories based on its main functions and effects.

****Your task:****
 For each creative strategy provided, determine the PRIMARY CATEGORY (or two categories if the strategy serves multiple major functions) that best describes the strategy's main function(s).

****Taxonomy Categories:****
 <insert definitions of the eight dimensions>

****Guidelines:****
 - Choose 1-2 categories that represent the strategy's primary functions
 - If a strategy clearly serves two major narrative purposes, assign both categories
 - Focus on what the strategy DOES rather than what it contains
 - Consider the strategy's main purpose(s) in the narrative context
 - Only use two categories if the strategy genuinely has dual primary functions - avoid over-categorizing

****Output Format:****
 Return your response as a JSON object with only the category assignment(s):
 {"category": ["PRIMARY_CATEGORY"]} or {"category": ["CATEGORY_1","CATEGORY_2"]}.

Context Prompt:
 Please categorize the following creative strategy:
 <insert strategy name> used in the plot:
 <insert plot content> .

The explanation for this strategy is:
 <insert strategy explanation> .

Assign it to one or two primary categories based on its main function(s) and narrative purpose(s). Use two categories only when the strategy genuinely serves dual major functions.

C Failure Cases from Human Evaluation of LLM-Extracted Strategies

Across the 11 extraction/explanation cases and 8 lexical-cue cases that at least one evaluator flagged as incorrect, raters highlighted points where the LLM analyses failed. Table 5 and Table 6 shows typical failure cases. For instance, describing the “precious sunstone” as a MacGuffin gestures in a sensible direction, but the explanation would benefit from pointing more explicitly to how characters’ choices turn on the object. Likewise, the “chain reaction” reading of one plot summarizes a rich sequence, but could be made more informative by isolating the immediate trigger within the plot that escalates consequences. We also saw a few cases where noticeable surface features, such as an in-medias-res opening (“Reconstructing K.C.’s consciousness...”), a reported line of speech (“Miss Cassandra calls out...”), or setting details (“small-town Florida,” “Mexico City,” “Sumidero Canyon”), were highlighted without fully spelling out their local function (e.g., how the opener withholds specifics to invite inference, what the utterance changes in the scene, or how

Table 5: Failure cases in LLM-extracted strategies and explanations flagged by raters.

Story Excerpts	Extracted Strategies & Explanations
<p>“...Maya and Willy went to the glow worms, but they took Maya and Willy up to wreck the hive and the precious sunstone, much to the displeasure of the Queen...”</p> <p>“...Will reaches a lower road and encounters a man who turns out to be a Frank’s guard, but talks him into letting him pass. He finds Frank’s camp in the woods, a derelict farm, the main building of which is an extensive but rudimentary drug lab. Most of the inhabitants are unconcerned at his presence...”</p> <p>“...Reconstructing K.C.’s consciousness, the Moon’s operating system appears to him as his cat, Fuzz Aldrin, and his mother, remarking that they must now ‘get started’...”</p> <p>“...After Lydell is arrested in a DEA sting, a desperate Liza asks Neel for help paying for the surgery but he exhorts her to use Phoebe’s situation as ‘fire’ for inspiration, like he did when his wife was dying from cancer and he came up with Lonafen. When a friend’s husband dies from a Lonafen overdose and her condolences are rejected, Liza agrees to testify to the U.S. Attorney’s office that’s investigating Zanna. She admits her involvement in Zanna’s speaker and bribery programs...”</p> <p>“...Meanwhile at the hive, Crawley tries to fix the sunstone, but to no avail, much to his dismay. Miss Cassandra calls out that Maya and Willy has left the meadow...”</p> <p>“...Alone at last, Shirley allows Stanley to read her work on Hangsaman. He declares it to be a work of genius. Shirley acknowledges his praise. It’s obviously a familiar pattern in their peculiar partnership. The two celebrate by drinking and dancing — together and yet alone in their cluttered house...”</p> <p>“...The nun then shows the brothers that Eloise really is an actual plane built by Flavio to pass down to his sons. Now having a way of getting home, Renato gets ready to part ways with Asher and the goat; Renatito. However, he changes his mind and invites them both to the wedding. They fly away together and return to Mexico as they pass above the Sumidero Canyon...”</p> <p>“...Ava Faulkner is a recovering addict and former soldier turned assassin...”</p> <p>“...Among the guests are Rowena’s housekeeper Olga Seminoff, Drake family doctor Leslie Ferrier and his son Leopold, and Joyce’s Romani assistant Desdemona Holland; they are joined by Maxime right before the séance, and during it Poirot reveals Desdemona’s half-brother Nicholas—and Joyce’s second assistant—hiding in the chimney...”</p> <p>“...In small-town Florida, Chris is working at a car wash when a woman, Maria, arrives. Chris tells the man whose car he’s washing that she is his high school crush. He begins to vacuum the car, but the overpowered vacuum sucks his clothes off, leaving him naked. He panics and hides in the car. The man has a conversation with Maria, attempting to get her number for Chris, but she declines and leaves...”</p> <p>“...In Mexico City, Sebastián Silva is a depressed, ketamine-addicted artist and filmmaker contemplating suicide by taking pentobarbital...”</p>	<p>Use of MacGuffin: The “precious sunstone” functions as a MacGuffin, an object that propels the plot and is central to the characters’ actions and the conflict.</p> <p>Unexpected Lack of Conflict: Rather than a tense or violent encounter, Will persuades the guard and the camp’s inhabitants are unconcerned, subverting expectations for dramatic confrontation and promoting intrigue about their motives or broader circumstances.</p> <p>Abrupt Narrative Start: The action begins in medias res, with K.C.’s consciousness being reconstructed and immediate engagement with the central mystery. This technique hooks the reader and leaves important details to be filled in later.</p> <p>Chain Reaction Causality: The plot unfolds through a tight sequence of cause and effect, where each character’s action directly provokes consequential reactions, propelling the narrative forward and intensifying momentum.</p> <p>Direct Character Dialogue: Involving Miss Cassandra’s speech adds immediacy and a sense of presence to the scene, making the action feel ongoing and interactive while also revealing plot developments through character voices.</p> <p>Symbolic Setting: The ‘cluttered house’ setting at the conclusion operates symbolically, reflecting the characters’ internal lives and the complexity of their relationship.</p> <p>Symbolic Setting Transition: Flying above the Sumidero Canyon serves as a visual marker of transition—both geographic and emotional—underscoring the theme of crossing thresholds into new phases of life.</p> <p>Economical Exposition: The plot delivers significant backstory and premise in a single sentence, using minimal language to maximum effect. This directness creates instant engagement without over-explanation, trusting the audience to infer dramatic stakes.</p> <p>Romani Identity Signification: Distinctly identifying Desdemona as ‘Romani’ marks cultural and social difference, potentially invoking theme, bias, or stereotype as part of the plot’s societal commentary.</p> <p>Setting as Character: Locating the plot in ‘small-town Florida’ uses regional specificity to evoke a sense of place, potential small-town gossip, and intimacy, affecting both character behavior and plot plausibility.</p> <p>Setting Establishment: By explicitly setting the story in Mexico City, the narrative immediately grounds the plot in a specific cultural and urban environment, providing context and atmosphere that can shape the reader’s expectations and the protagonist’s experiences.</p>

the locale shapes stakes and behavior). On evidence selection, raters were concerned when quoted cues were very general (e.g., quoting only the noun “the precious sunstone”), leaned on temporal or expository phrases (“The next morning...”, “with his emotionally estranged wife”) rather than the textual signals of the claimed device, or lifted long recap-style spans that restate events instead of pinpointing the terms that create causal or thematic pressure.

D User Study

D.1 Baseline System Interface

Given that no existing solutions are directly comparable to NARRIX in terms of supporting interaction with narrative strategies in

Table 6: Failure cases in LLM-extracted lexical cues flagged by raters.

Story Excerpts	Extracted Strategies & Explanations
<p>“...Maya and Willy went to the glow worms, but they took Maya and Willy up to wreck the hive and the precious sunstone, much to the displeasure of the Queen...”</p>	<p>Use of MacGuffin: The “precious sunstone” functions as a MacGuffin, an object that propels the plot and is central to the characters’ actions and the conflict.</p>
<p>“...Reconstructing K.C.’s consciousness, the Moon’s operating system appears to him as his cat, Fuzz Aldrin, and his mother, remarking that they must now ‘get started’...”</p>	<p>Lexical Cues: “the precious sunstone” Abrupt Narrative Start: The action begins in medias res, with K.C.’s consciousness being reconstructed and immediate engagement with the central mystery. This technique hooks the reader and leaves important details to be filled in later.</p>
<p>“...After Lydell is arrested in a DEA sting, a desperate Liza asks Neel for help paying for the surgery but he exhorts her to use Phoebe’s situation as ‘fire’ for inspiration, like he did when his wife was dying from cancer and he came up with Lonafen. When a friend’s husband dies from a Lonafen overdose and her condolences are rejected, Liza agrees to testify to the U.S. Attorney’s office that’s investigating Zanna. She admits her involvement in Zanna’s speaker and bribery programs...”</p>	<p>Lexical Cues: “Reconstructing K.C.’s consciousness” Chain Reaction Causality: The plot unfolds through a tight sequence of cause and effect, where each character’s action directly provokes consequential reactions, propelling the narrative forward and intensifying momentum. Lexical Cues: “After Lydell is arrested in a DEA sting, a desperate Liza asks Neel for help”, “When a friend’s husband dies from a Lonafen overdose and her condolences are rejected, Liza agrees to testify”</p>
<p>“...Hutch Mansell sits inside of an interrogation room, heavily bruised and injured. Two FBI agents interrogate him about his identity, before Hutch responds that he’s simply “nobody.” In a flashback to a few days earlier, Hutch leads an ordinary, mundane life as an office worker with his emotionally estranged wife Becca, teenager Blake and a much younger daughter, Abby...” “...Meanwhile at the hive, Crawley tries to fix the sunstone, but to no avail, much to his dismay. Miss Cassandra calls out that Maya and Willy has left the meadow...”</p>	<p>Emotionally Distant Familial Relationships: By describing Hutch’s home life as ‘emotionally estranged,’ the story hints at inner conflicts and character motivations, adding depth and potential for emotional development. Lexical Cues: “with his emotionally estranged wife Becca” Direct Character Dialogue: Involving Miss Cassandra’s speech adds immediacy and a sense of presence to the scene, making the action feel ongoing and interactive while also revealing plot developments through character voices. Lexical Cues: “Miss Cassandra calls out that Maya and Willy has left the meadow”</p>
<p>“...Her brother Kevin is caught dancing in the restroom by some older students who record him and bully him, but Clancy’s and Kevin’s mother Margot, the lunch monitor, scares them away. The bullies upload a remix video of Kevin dancing and Margot admonishing them to YouTube, where it quickly goes viral, gaining over two million views...”</p>	<p>Authority Figure Intervention: Margot, as a lunch monitor and mother, intervenes to protect her son, adding a layer of authority and adult presence that both disrupts the bullying and sets up later plot developments. This juxtaposition of child vulnerability and adult protection intensifies reader investment in the outcome. Lexical Cues: “Margot, the lunch monitor, scares them away”</p>
<p>“...Alone at last, Shirley allows Stanley to read her work on Hangsaman. He declares it to be a work of genius. Shirley acknowledges his praise. It’s obviously a familiar pattern in their peculiar partnership. The two celebrate by drinking and dancing — together and yet alone in their cluttered house...” “...The next morning, Kath awakens to find everyone gone. Venturing into the surrounding woods, she finds Al, crying. He explains that he found Greta and Max having sex and that the two left together...”</p>	<p>Symbolic Setting: The ‘cluttered house’ setting at the conclusion operates symbolically, reflecting the characters’ internal lives and the complexity of their relationship. Lexical Cues: “in their cluttered house” Withholding Information: Key details about what happened the night before are initially omitted, heightening suspense and encouraging curiosity. This deliberate omission motivates the reader to follow Kath as she seeks answers. Lexical Cues: “The next morning, Kath awakens to find everyone gone.”</p>

examples, we implemented a baseline system (Fig. 10) that closely resembled NARRIX in UI but excluded the key features unique to our approach. Both systems shared the same Markdown text editor, ensuring a consistent writing environment. However, in the baseline, we removed the interactive story arc visualization. In the Browser panel, example stories were shown as story cards; each card could be clicked to reveal the full text of the example story, but without highlighting or extracting narrative strategies. The Remix panel in NARRIX was replaced by an AI chat assistant. This assistant, powered by the same underlying LLM but designed to mimic mainstream chat-based writing interfaces (such as ChatGPT Canvas):

users could interact via free-form chat prompts, and the system would respond conversationally in a dedicated output panel. This baseline followed design conventions used in prior work [59, 73, 98], which also implemented chat-based AI writing assistants as comparison conditions when evaluating novel writing tools.

In summary, this baseline design (1) retains basic non-contributory features of NARRIX (e.g., the Markdown editor) to minimize interface confounds, (2) uses the same underlying model to isolate interaction design effects, (3) mirrors real-world practices where users have access to general-purpose AI tools like ChatGPT, and (3) integrates story editing, example browsing, and AI assistance within a unified workspace to avoid unnecessary window switching.

Table 7: Demographic information of participants and their usage of examples and AI writing tools in daily writing practice. *Writing proficiency; **English proficiency (both on Likert scales from 1/low to 7/high); *Frequency of referring to others' work in writing (from Never to Always).**

ID	Gender	Age	Writing*	English**	Example***	AI Writing Tools	AI Usage Freq.
P01	Female	27	1	4	Always	ChatGPT, Grammarly, DeepL	Weekly
P02	Male	28	2	5	Always	ChatGPT, Grammarly, Cursor	Daily
P03	Male	25	3	5	Sometimes	ChatGPT, Cursor	Daily
P04	Female	25	2	6	Often	ChatGPT, Grammarly	Daily
P05	Female	28	1	5	Always	ChatGPT	Weekly
P06	Male	25	2	5	Always	ChatGPT, Grok, Gemini, Grammarly	Daily
P07	Female	26	3	6	Always	ChatGPT, NotebookLM, Grammarly	Weekly
P08	Female	—	2	3	Sometimes	ChatGPT, Grammarly	Daily
P09	Female	28	3	7	Sometimes	ChatGPT, Grammarly	Daily
P10	Male	28	4	4	Often	ChatGPT, Notion AI, Grammarly	Daily
P11	Female	23	4	5	Sometimes	ChatGPT, Grammarly	Weekly
P12	Female	26	1	5	Often	ChatGPT	Daily

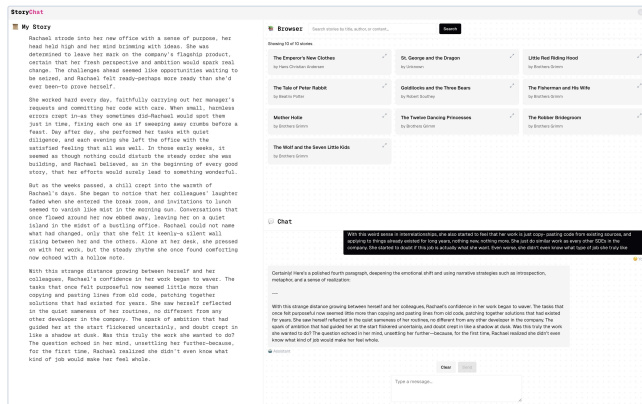


Figure 10: Baseline system interface, which includes the same Markdown editor as NARRIX, example stories presented as whole-text cards without surfacing narrative strategies, and a chat-based AI assistant, similar to interfaces such as ChatGPT Canvas.

D.2 Participant Information

We recruited 12 participants (Table 7; 8 female, 4 male), ages 23–28 ($M = 26.27$, $SD = 1.68$), from a large software organization in the United States via internal communication channels and word of mouth. We sought novice writers and, following prior work that recruited ESL writers as novices for writing tasks [36, 98], targeted non-native English speakers during recruitment. All participants reported regularly engaging in writing and wanting to improve

their creative writing skills; each had prior creative writing experience and self-rated their expertise on a 7-point scale ($M = 2.33$, $SD = 1.07$; 1 = beginner, 7 = professional). On a 5-point scale, participants reported actively seeking and referring to examples in writing ($M = 4.08$, $SD = 0.90$; 1 = never, 5 = always) and regularly using generative AI tools (e.g., ChatGPT) in their writing activities ($M = 4.67$, $SD = 0.50$; 1 = never, 5 = daily).

D.3 Interview Questions

- (1) What was your overall experience with the two tools? Was there anything that excited or frustrated you?
- (2) Did you have any “aha” moments or memorable experiences while working with either tool?
- (3) Which features did you find most beneficial in each tool, and in what scenarios were they especially useful?
- (4) How did your strategies for working with or using AI differ between the two systems?
- (5) Before this study, how did you typically use examples in writing? After using the tool, did your approach to using examples in writing change? If so, how?
- (6) Can you describe any specific narrative strategies or techniques you learned from the examples and how you tried to use them in your own story?
- (7) What was your approach to selecting narrative strategies that best fit your goals?
- (8) Do you think the tool will contribute to your long-term writing skill development? If so, how?
- (9) Do you have any suggestions or ideas to improve the tool?