# EmoArc: Interactive Emotion Graph for Human-AI Collaborative Writing

Rifat Mehreen Amin
LMU Munich
Munich, Germany
rifat.amin@ifi.lmu.de

Seyit Baran Özdemir
LMU Munich
Munich, Germany
s.ozdemir@campus.lmu.de

Sven Mayer
TU Dortmund University
Dortmund, Germany
info@sven-mayer.com

Andreas Butz
LMU Munich
Munich, Germany
andreas.butz@ifi.lmu.de

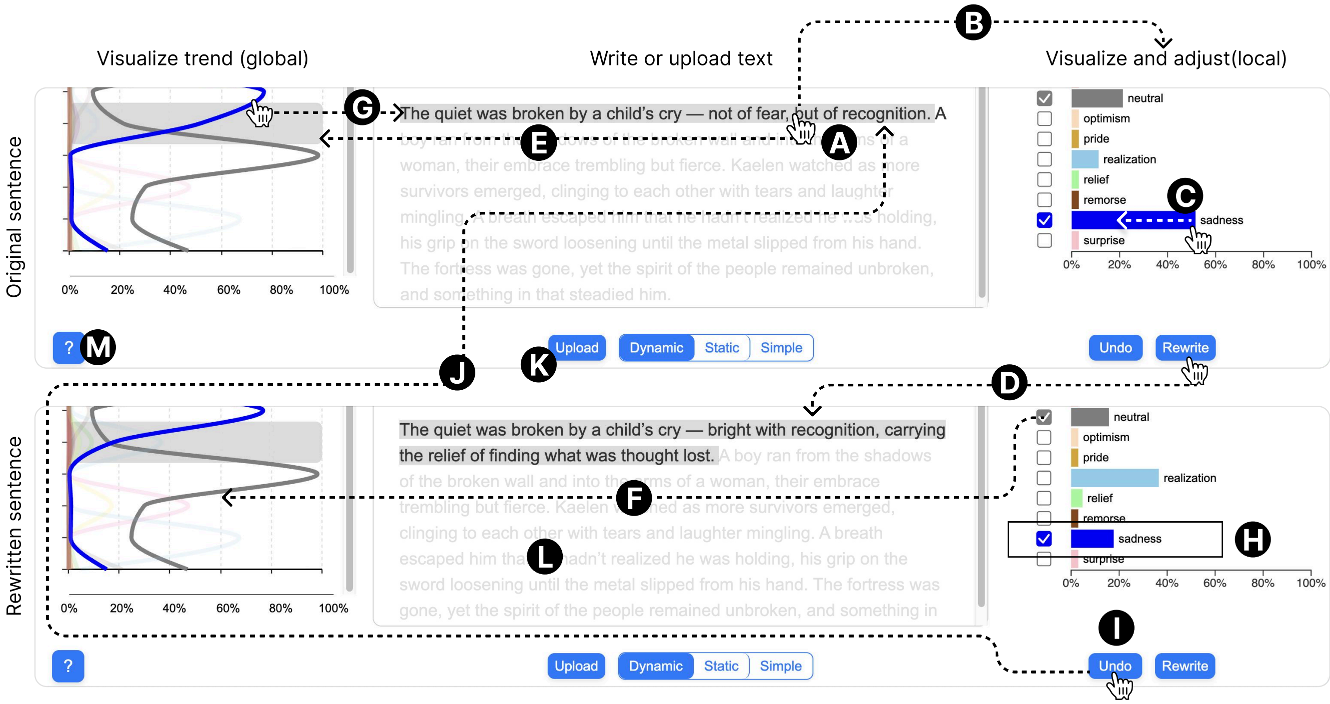


**Figure 1: *EmoArc* consists of three components: the main text editor (center, L), a line graph synchronized with the text (left), and a bar graph (right). Writers can either draft text or (K) upload a file. (A) When a sentence is selected, it is highlighted and simultaneously linked to both (E) the global emotional arc and (B) the local emotion bars. (C) Adjusting the bars and (D) pressing Rewrite generates a new sentence with (H) the desired emotion intensities. (F) Toggles in the legend allow selectively focusing on specific emotions, while (G) clicking on the line chart highlights the corresponding sentence in the editor. (I) Pressing Undo (J) restores the original sentence, and the visualizations readapt.**

## Abstract

With the advent of large language models, creative writers can get support for refinement, co-writing, and text generation. Yet, writers often struggle to understand how model-driven rewrites affect the emotional trajectory of their narratives. We investigate how interactive visualizations can help writers gain awareness, agency, and control over the emotional flow of their texts. First, we designed an interactive visualization-based writing tool that allows writers to adjust emotional pacing, explore different emotional directions, and revise the emotional storyline in real time. Second, we conducted a user study with creative writers ($N = 24$) to evaluate its effectiveness. Our results show a significantly improved enjoyment, exploration, and perceived value. Writers highlighted increased emotional awareness and narrative coherence; simultaneously, they expressed concerns about authorship. Thus, support can enrich the creative process when designed with transparency, user agency, and adaptability in mind, contributing to the understanding of augmenting human writing creativity.


*AVI 2026, Venice, Italy*

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## CCS Concepts

• **Human-centered computing → Visualization**; **Interactive systems and tools**; • **Computing methodologies → Artificial intelligence**.


## Keywords

Human-AI collaboration, Co-creative systems, Visualization, LLMs, GenAI, emotions, creative writing

**ACM Reference Format:**
Rifat Mehreen Amin, Seyit Baran Özdemir, Sven Mayer, and Andreas Butz. 2026. EmoArc: Interactive Emotion Graph for Human-AI Collaborative Writing. In *International Conference on Advanced Visual Interfaces 2026 (AVI 2026), June 8–12, 2026, Venice, Italy*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn


## 1 Introduction

Emotion plays a central role in storytelling, shaping how narratives unfold and how writers intend their work to be experienced. While writers often have an intuitive sense of emotional tone, maintaining awareness of emotional pacing across a text, especially during revision, can be challenging. Subtle tonal imbalances may go unnoticed. Therefore, maintaining a well-balanced emotional flow throughout a story can be challenging [24]. Reflecting on emotional structure often requires rereading and guesswork. Prior work shows that visualizing emotional information, using timelines, color mappings, or circumplex layouts, can make affective patterns in text more accessible [33]. Yet many existing systems analyze emotion post hoc or at coarse granularity, offering limited support for iterative refinement during writing [35]. What remains missing is an approach that combines fine-grained emotion analysis with interactive visualizations and immediate feedback, giving authors agency over their narratives' emotional flow.

To address this gap, we developed and evaluated *EmoArc*, an interactive writing tool that integrates sentence-level emotion detection, visualization, and AI-assisted rewriting. *EmoArc* enables sentence-level exploration of emotional direction without word-level micromanagement, which better accommodates contextual nuance than isolated lexical approaches [32]. Sentence-level emotional shifts play a key role in narrative pacing [21], motivating tools that support fine-grained emotional guidance during writing [22]. *EmoArc* links local edits with global emotional trends and separates emotion analysis from text generation: **all sentences are analyzed post hoc by an emotion classification model**[1]**, while a generative model is used only to propose rewrites conditioned on adjusted emotional targets**. In a study with 24 participants, we compared *EmoArc* to a static visualizing-only tool and a baseline without emotional writing support. We asked participants to complete three writing tasks and evaluated the usability and workload. Our research was guided by the following questions:

**RQ1** How can interactive emotion graphs help authors to adjust the emotional flow of their narratives?
**RQ2** What impact does integrating emotion graphs and AI-assisted rewriting have on writers' creativity and enjoyment?
**RQ3** How does real-time emotional feedback shape authors' perceived authorship, control, and confidence while writing?

[1]https://huggingface.co/SamLowe/roberta-base-go_emotions

The results from our study showed that *EmoArc* significantly improved enjoyment, exploration, expressiveness, and perceived value compared to a simple text editor without support, while maintaining low mental load. It also increased enjoyment over the static visualization-only UI, though tensions around authorship and personal involvement emerged. These findings suggest that emotion-centered visualizations can enrich creative exploration while reshaping how writers reflect on agency and ownership.

## 2 Related Work

We review prior work across two themes: (1) emotion visualization and multimodal systems, and (2) emotion in creative writing and narrative studies.

### 2.1 Emotion Visualization and Multimodal Systems

Building on psychological and computational models of emotion, prior work has explored visualizing and interacting with emotional patterns in text and other media. Many systems adopt dimensional models such as VAD to show emotional variation over time. For example, PEARL [35] visualize emotional dynamics using line- and color-based encodings, while large-scale systems like We Feel [15] and NewsViz [10] focus on aggregate mood patterns in social and narrative data. Related approaches also extend beyond text, visualizing emotion in music or multimodal content using established affective models [8, 23]. However, many of these systems rely on lexicon-based analysis and struggle with context, irony, or cultural variation. More recent work emphasizes interactivity and multimodality. Systems such as StanceVis Prime [14], and EmotionVis [36] provide richer temporal and semantic views of affective information, while multimodal tools like Speak From Heart [16] integrate emotional signals across text, audio, or video.

### 2.2 Emotion in Creative Writing and Narrative Studies

Early research on emotion in narrative examined emotional arcs and pacing without AI-driven generation. Mohammad showed how emotional expression varies across genres and authors [20, 21], while systems such as Ashida et al. [2] visualized characters' emotional trajectories to support narrative coherence. Later work integrated computational affect into writing workflows; for example, Mori et al. [22] provided emotion-guided story continuation using sentence-level affective targets. Subsequent systems expanded narrative support through visualization and structural interaction. StoryPrint [34] visualizes narrative structure together with character emotion. More recently, large language models have enabled interactive writing systems that foreground emotional dynamics in co-creation, including Textoshop's editing-inspired manipulation of affective tone [17]. TaleBrush [5] and PatchView [6] enable high-level structural sketching and visual steering of text generation. Recent work on visual story writing further extends this direction by allowing authors to manipulate visual representations of narrative elements to suggest text revisions [18]. Other LLM-based tools explore narrative and emotional support through vocabulary learning, procedural creativity, and multimodal prompting [26, 28].

## 3 EmoArc

We conducted informal design sessions with UX experts and the authors, involving brainstorming, sketching, prototyping, and pilot testing to develop a dynamic, interactive UI for emotion analysis. Through this process, we explored multiple concepts and layouts. The resulting system, *EmoArc*, was guided by four design goals that shaped its architecture and interaction model.

### 3.1 Design Goals

- **DG1: Provide a platform for meaningful text emotion analysis and manipulation**
  Understanding emotional sentiment in writing is inherently nuanced and context-dependent. Generative AI tools often increase metacognitive load by requiring repeated prompt reformulation [29], whereas tools that enable direct analysis and manipulation of emotion can reduce this burden and support more intentional authoring [12].
- **DG2: Ensure real-time responsiveness and interactivity**
  Emotion analysis should offer immediate, interpretable feedback to support fluid writing. Static, linear UIs can disrupt creative flow by requiring repeated prompt reformulation [13], whereas fragment-based and responsive interactions enable more continuous engagement with AI tools [4].
- **DG3: Enable bidirectional emotional consistency**
  Writing tools should preserve narrative coherence while supporting exploration of alternatives. Prior work shows that authorship can diminish when continuity and control are weakened [13, 19]. Consistently reflecting emotional adjustments across text and visualization helps maintain this sense of authorship.
- **DG4: Provide intuitive emotion visualizations**
  Emotion analysis is most effective when presented in clear, interpretable forms. Prior work shows that simple visual encodings such as line and bar graphs support interpretation and exploration [25, 33], while creativity support research emphasizes making temporal and structural aspects of writing visible [3, 7].

### 3.2 User Interface and Interaction

Our system turns emotion from a hidden property of text into a visible and editable layer of narrative structure. The UI (see Figure 1) integrates three primary components: text editor, bar graph, and line graph, plus supporting tools (file upload, Rewrite, Undo, tutorial). Below, we describe the interaction as a sequence of user actions.

*3.2.1 Provide Text (Write or Upload).* The text editor (Figure 1-L) provides a familiar writing environment. Writers can compose text directly or upload a `.txt` file by clicking on the Upload button, which is automatically segmented into sentences for analysis and editing. Emotion analysis runs automatically during pauses in writing, providing timely feedback while maintaining interface responsiveness (DG2), supporting the aim of making progress visible and concrete enough for immediate feedback [30].

*3.2.2 Select a Sentence.* Selecting a sentence applies a grey backdrop in the editor (Figure 1-A) and highlights its emotional intensities across both bar graph (Figure 1-B) and line graph (Figure 1-E), tightly coupling text with emotion analysis through visualizations (DG1).

*3.2.3 Inspect Locally (Bar graph).* Bar graphs provide an intuitive way to represent and compare emotion intensities [33]. In *EmoArc*, they display and allow adjustment of the normalized emotion distribution of a selected sentence, providing clear sentence-level control (Figure 2) and visual clarity (DG4). Hovering reveals exact percentages (Figure 2-F), and dragging a bar updates the target intensity.

*3.2.4 Adjust & Rewrite.* Users begin with the original sentence and its emotion intensities (Figure 2-A) and adjust the bars directly as sliders, for example lowering sadness and increasing surprise to set a target emotional distribution (Figure 2-B). Pressing the Rewrite button (Figure 2-C) rewrites a revised sentence aligned with these targets while preserving context (Figure 2-D) (DG3). The updated sentence and its emotion intensities are shown immediately.

*3.2.5 Undo.* If needed, the Undo button restores the original sentence (and visualizations will follow), which supports safe exploration because it makes it easy to recover from mistakes or dead ends (Figure 2-E).

*3.2.6 Inspect Globally (Line graph).* While the bar graph provides a sentence-level view, the line graph offers a global perspective on emotional flow across the text. Line graphs are well suited for visualizing narrative arcs and emotional variation over a document [33]. In *EmoArc*, each emotion is shown as a line over the document's progression (Figure 3-D), enabling writers to observe global trends. The line graph is scroll-synchronized with the editor (Figure 1-G) so that the visualization reflects the currently visible text.

*3.2.7 Filter & Focus (Legend Toggles).* To reduce clutter, the line graph includes an interactive legend (Figure 3-B) that allows writers to toggle individual emotions. Selecting an emotion highlights its curve. For example, isolating confusion and disappointment to examine their balance over time (Figure 3-C), while non-selected emotions are grayed out but remain visible (Figure 3-E). This filtering supports focused inspection of emotional arcs without overwhelming the display (DG1).

*3.2.8 Navigate & Synchronize.* Hovering over a line in the graph dynamically displays the associated emotion label near the cursor (Figure 3-F), while transparent interactive zones along each row allow hover and click interactions. Clicking anywhere in the graph (Figure 3-A) highlights the corresponding sentence with a semi-transparent rectangle and synchronizes the editor view, automatically scrolling to center the selected text using easing and clamping for smooth navigation.

*3.2.9 Learning and Support.* To lower the entry barrier, a tutorial appears on first use and can be reopened via a help button ? (Figure 1-M). Additional scaffolds, such as tooltips, color coding, and smooth transitions, support quick understanding of the workflow.

The interaction follows a cyclical process: *write* → *visualize* → *adjust* → *rewrite* → *reflect.* **Writers can revise text either by using AI-assisted rewriting or by editing it themselves, with the visualization updating in both cases. They may choose to adjust emotional targets and invoke AI-assisted rewriting, or**

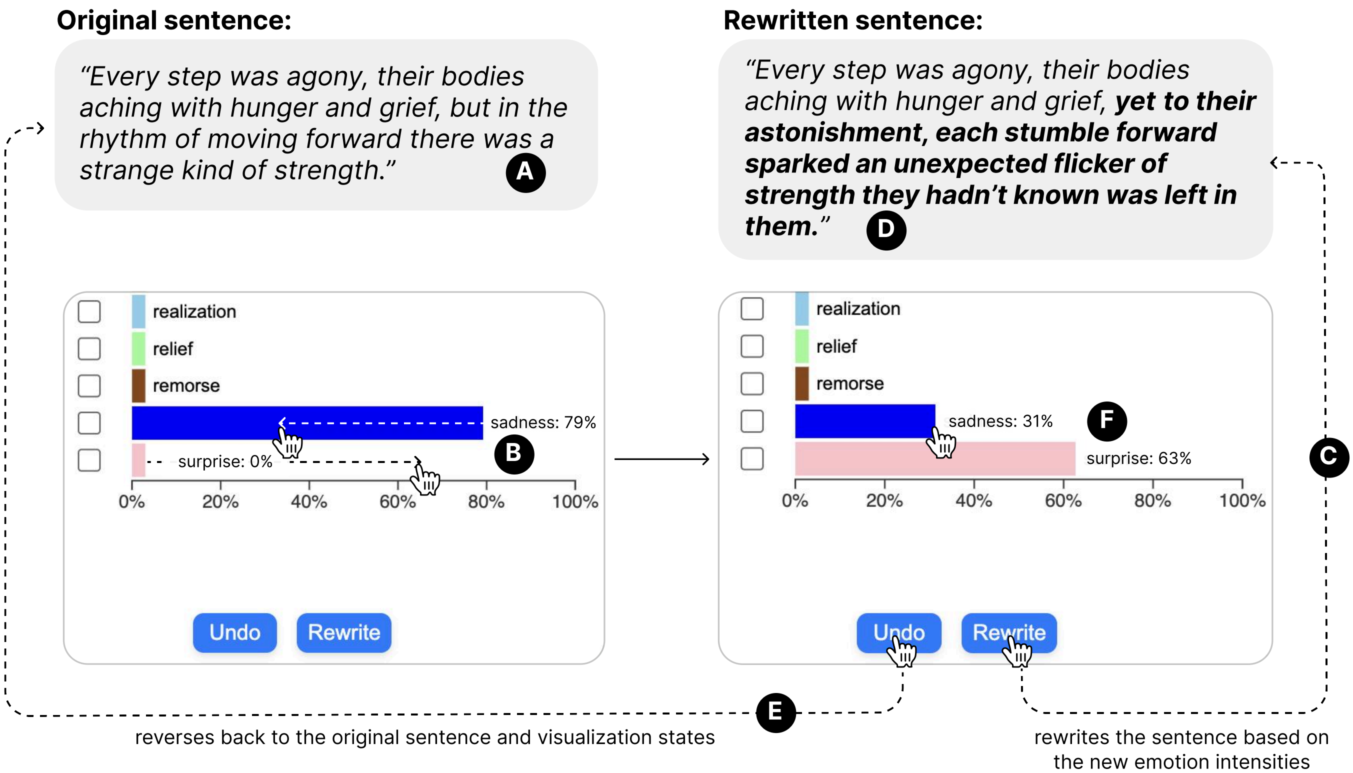


**Figure 2: Bar graph for local inspection and adjustment of emotional tone. (A) The original sentence. (B) Users can manipulate the bars directly. (C) The Rewrite button generates a new sentence to align with the new emotional distribution. (D) The generated sentence is displayed, while (E) the Undo button restores the original sentence and visualization state.**

**instead revise the text themselves and see the visualization update accordingly.** By tightly coupling writing, visualization, and rewriting, the system supports deliberate shaping of affective arcs while preserving writer control (see the supplementary material for an example scenario.).

## 3.3 Technical Implementation

The system is built as a modular client-server application. The frontend is a web interface implemented with D3.js, JavaScript, and CSS, while the backend is developed in Python using Flask. The two sides communicate asynchronously through a DataService module that handles all API calls. See the supplementary materials for more details on how we selected the emotion classification model, the emotion analyzer, and the sentence generator.

# 4 User Study Design

We conducted a within-subjects lab study with 24 participants (ages 18-45, $M = 29.36, SD = 6.70$) with varied creative writing experience and varying familiarity with AI writing tools (more participant details in the supplementary material). Participants were compensated 15€ for a 90-minute session, and the study was approved by our institution’s ethics board.

We evaluated *EmoArc* (dynamic UI) against two comparison conditions: a baseline simple text editor (simple UI, only the center part in Figure 1) and a visualization-only variant without AI-assisted rewriting (static UI). The study included a pre-study questionnaire, a standardized interface tutorial[2], post-task surveys, and a semi-structured interview. Screen and audio were recorded for analysis. Participants completed three short creative writing tasks (see supplementary materials for more details), each requiring the addition of 2-3 sentences to a given text prompt, with task and interface order counterbalanced across participants.

We collected quantitative measures, including the Creativity Support Index (CSI), NASA Task Load Index (NASA-TLX). We also gathered Likert-scale ratings on ease of use, perceptions of authorship and ownership, and the desirability of regular use for each interface. Finally, we recorded participants’ overall preferences among the different tools. The interview gathered feedback on the emotion graph’s usefulness and granularity, perceived creativity and control, authorship and AI collaboration, and overall usability. For the statistical analysis, we first performed a Shapiro-Wilk normality test. Based on the results, we used either a paired t-test or the Wilcoxon signed-rank test to determine the statistical significance of observed differences. All p-values were corrected using the Holm-Bonferroni method to account for multiple comparisons. For qualitative analysis, the first and second authors conducted

[2] submitted as supplementary material

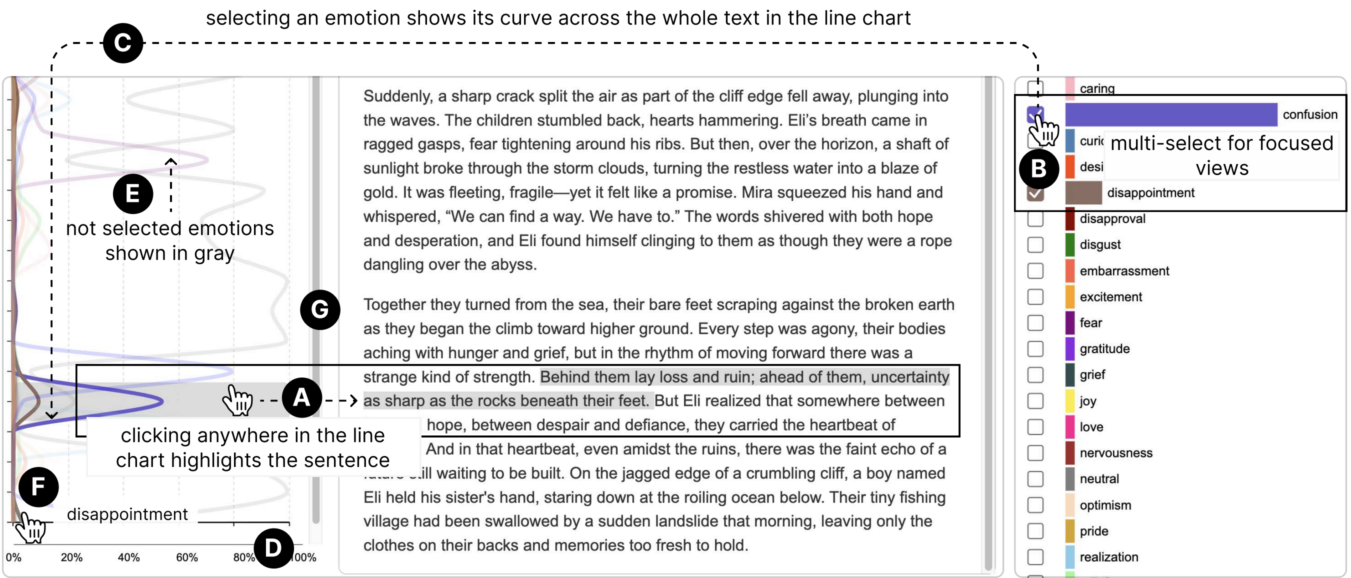


**Figure 3: Line graph for global inspection of emotional trajectories. (A) Clicking anywhere in the graph highlights the corresponding sentence in the editor. (B, C) Users can select or multi-select emotions from the legend to focus on specific trajectories, while (E) non-selected emotions are shown in gray. (F) Individual curves represent normalized emotion intensities across the text, and (G) scroll synchronizes the editor with the visualization.**

open coding of interview data to identify key themes, which were refined through discussion with the co-authors.

## 5 Results

In this section, we present the results of our lab study, structured by the research questions they address.

### 5.1 How can interactive emotion graphs help writers to adjust the emotional flow of their narratives? (RQ1)

Participants described the emotion graphs as a useful and often novel aid for reflecting on and guiding emotional flow. Many noted increased awareness of emotional pacing and support for real-time wording or tone adjustments.

*5.1.1 Changes to Planning and Story Structure.* The degree to which the graph influenced participants' narrative planning varied. A majority indicated that they followed their initial story plans and only occasionally adjusted wording. For instance, P5 explained: *"I followed my initial plan. The graph didn't make me change structure"*. However, others reported concrete changes when the graph highlighted discrepancies between their intended emotion and what was detected. P7 described: *"In the 1st story I was going to write something else. But then, when I looked at the graph, I saw that the sentence was not projecting exactly what I wanted to write. So I had to change one or two words, and then the final outcome was according to what I wanted"*. These suggest that the graph not only supported evaluation but also triggered moments of revision and redirection.

*5.1.2 Intuitiveness of Manipulation.* Overall, participants perceived the interface for manipulating emotions using the bar graph as straightforward. P1 described it as *"simple and intuitive,"* while P10 emphasized: *"It was very easy to use."* Others highlighted the value of immediate responsiveness, as P21 explained: *"I liked how instant it is. For example, the rewriting process takes only 20 seconds. That is impressive... And the thing is, as I'm writing, the graph keeps changing. That's great."* P19 said, *"It was quite simple. The Bar graphs, when I was increasing and decreasing the emotions, the leveling of those were what I could understand."* Nonetheless, several participants identified areas where usability could improve (P14, P17). P17 pointed out that some functions were not self-explanatory: *"The interface was quite intuitive. The only thing not obvious was using 'shift' to select multiple emotions."* Some reported moments of confusion when the system's interpretation diverged from their own. P2 explained: *"Sometimes it delivers tone-deaf sentences... If the tool says a sentence is 80% admiration, I might feel it's actually 50 or 60%."*. These mismatches introduced moments of doubt and minor frustration.

*5.1.3 Cognitive Load and Effort (NASA-TLX and General Perception).* We analyzed raw NASA-TLX ratings across the three interfaces to evaluate perceived workload. A Friedman test revealed no significant effects of condition on any NASA-TLX factor, nor on the overall NASA-TLX score (see supplementary materials for details). This indicates that participants experienced all three interfaces as lightweight and manageable.

Qualitative feedback helps explain these patterns. Several participants noted that the tool reduced editing burden by making emotional adjustments more direct. As P17 explained, the system allowed them to "quickly adjust a sentence to portray a certain emotion without overthinking". In contrast, some participants found the static UI more burdensome due to its reduced interactivity. P4

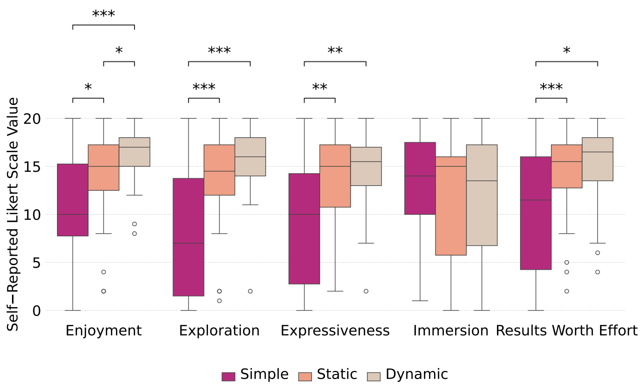


**Figure 4: The CSI scores for our three conditions. In line with prior studies [1, 27], we excluded the Collaboration factor in CSI, as our study did not involve collaborative settings.**

described a sense of "mental pressure" when interaction options were limited, and others noted that static interactions sometimes felt restrictive. These concerns were not raised for the dynamic UI, which participants described as smoother and less obstructive. Overall, while statistical tests did not reveal significant differences, qualitative findings suggest that the dynamic UI offered the best balance of perceived success and manageable effort, whereas static and simple UIs involved trade-offs between effort and engagement.

## 5.2 What impact does integrating emotion graphs and AI-assisted rewriting have on writers' creativity and enjoyment? (RQ2)

We assessed perceived creativity support using the Creativity Support Index (CSI). Results suggest that combining emotion graphs with AI-assisted rewriting increased enjoyment and perceived creative potential compared to simpler designs.

*5.2.1 Creativity Support Index (CSI).* CSI scores showed that the dynamic UI consistently outperformed the simple UI across several dimensions (see Figure 4). A Friedman test revealed significant effects of the UIs on Enjoyment ($\chi^2(2) = 9.18, p = .010$), Exploration ($\chi^2(2) = 20.74, p < .001$), Expressiveness ($\chi^2(2) = 15.58, p < .001$), Results Worth Effort ($\chi^2(2) = 15.93, p < .001$), and the Overall CSI score ($\chi^2(2) = 10.81, p = .004$). No significant effect was found for Immersion ($\chi^2(2) = 1.28, p = .53$).

Enjoyment was significantly higher in the dynamic UI ($W = 238.5$, $Mdn = 17$, $p < .001$) compared to the simple UI ($Mdn = 10$), and also higher in the static UI ($W = 44.5$, $Mdn = 15$, $p = .042$) compared to the simple UI ($Mdn = 10$); the dynamic UI was additionally rated more enjoyable than the static UI ($W = 152$, $Mdn = 17$ vs. $Mdn = 15$, $p = .042$). Exploration followed a similar pattern, with the dynamic UI ($W = 241$, $Mdn = 16$, $p < .001$) rated higher than the simple UI ($Mdn = 7$), and the static UI ($W = 10$, $Mdn = 14.5$, $p < .001$) also surpassed the simple UI ($Mdn = 7$). Expressiveness was likewise higher in the dynamic UI ($W = 204$, $Mdn = 15.5$, $p = .004$) than in the simple UI ($Mdn = 10$), and in the static UI ($W = 14.5$, $Mdn = 15$, $p = .002$) compared to the simple UI ($Mdn = 10$). For results worth effort, the dynamic UI ($W = 211$, $Mdn = 16.5$, $p = .012$) scored higher than the simple UI ($Mdn = 11.5$), and the static UI ($W = 2$, $Mdn = 15.5$, $p < .001$) also outperformed the simple UI ($Mdn = 11.5$).

Overall, CSI ratings reflected the same trend, with the dynamic UI scoring 74.83 ($W = 207$, $Mdn = 77.5$, $p = .004$) significantly higher than the simple UI that scored 52.78 ($Mdn = 54.5$), and the static UI scored 67.92 ($W = 29$, $Mdn = 74.5$, $p = .003$), also higher than the simple UI ($Mdn = 54.5$). No significant differences were found for immersion or between the dynamic and static UIs across most dimensions (see supplementary materials for detailed CSI scores). Overall, both the dynamic and static UIs supported creativity more effectively than the simple UI, with the dynamic UI showing the most consistent benefits for enjoyment, exploration, expressiveness, and overall creative support.

*5.2.2 Exploring New Creative Directions.* A central benefit of combining emotion graphs with rewriting was that it encouraged participants to take their writing in directions they might not have considered otherwise. Many reported that the sliders or rewrites sparked new ideas or opened unexpected possibilities . P1 described: *"it opened new ways of thinking about my sentences."* Similarly, P8 remarked: *"This gave me a wide range of emotions, and it showed me that a sentence can be written in different ways to express different emotions. That really made me feel creative."*. For some, rewrites even inspired iterative improvement, as P14 explained: *"Even though I was aiming for one goal, the first rewrite gave me another, and that inspired a third version that was even better"*. These reflections align with higher CSI ratings for the dynamic UI, which gave participants direct, real-time control over their emotions.

*5.2.3 Alignment with Writers' Intentions.* While the system was generally successful at modifying emotional tone, participants reported mixed experiences in whether the rewrites aligned with their intentions. Many felt the tool reliably reflected their adjustments. For instance, P6 noted: *"Yes, they mostly reflected what I wanted."* However, others found the results inconsistent or stylistically unsatisfying. P21 emphasized: *"I was not very happy with the rewriting option (in dynamic UI)... It didn't have that human touch."* These responses suggest that while the system adjusted emotional content effectively, maintaining stylistic quality and authorial voice remains a challenge.

*5.2.4 Enjoyment and Perceived Creativity.* Overall, participants described the system as enjoyable and creatively stimulating. Many emphasized that it gave them a sense of playfulness and creative exploration, consistent with their CSI ratings. P10 remarked: *"Seeing the real-time reflection somehow made me more confident"*. Yet, a minority felt less creative, often when they relied heavily on the system or when rewrites clashed with their personal style. For example, P9 reflected: *"I feel like I was very lazy in writing it, like I just let the AI do it for me. So I didn't feel too creative."* This illustrates how the system's creative potential was most evident when writers used it as a starting point for exploration, rather than as a replacement for their own authorship.

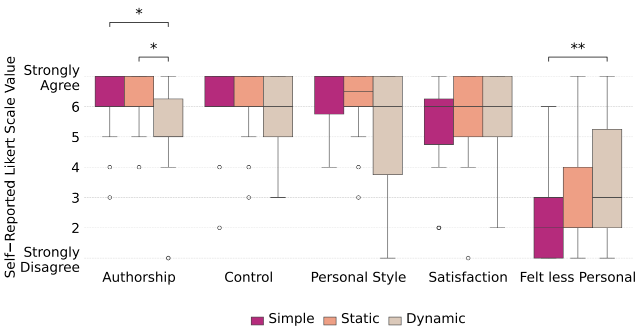


**Figure 5: Self-reported perceptions of authorship, control, personal style, satisfaction, and personal involvement across the three UIs.**

## 5.3 How does real-time emotional feedback shape writers' perceived authorship, control, and confidence while writing? (RQ3)

Participants reflected on how the system influenced their sense of authorship, their control over emotional direction, and how they positioned the tool in relation to their own writing.

*5.3.1 Quantitative Results.* A Friedman test revealed significant effects of condition on participants' sense of authorship ($\chi^2(2) = 8.54, p = .014$) and whether the story felt personal ($\chi^2(2) = 8.96, p = .011$). No significant differences were observed for perceived control over the outcome, the extent to which the ending reflected participants' writing style, or overall satisfaction with the final result.

Post-hoc Wilcoxon results revealed differences in how participants evaluated authorship and personal involvement across the three UIs (see Figure 5). When asked whether they considered themselves the author of the part they added to the story, participants gave higher authorship ratings in both the simple ($Mdn = 7$) and static ($Mdn = 7$) UIs compared to the dynamic UI ($Mdn = 5$). The differences were significant for both the simple UI-dynamic UI comparison ($W = 15.5, p = .040$) and the static UI-dynamic UI comparison ($W = 17.0, p = .024$). This suggests that when the system offered fewer interventions, participants felt more strongly that they remained the primary authors. For personal involvement, where the item stated *"Using the system made the story's continuation feel less personal to me"*, the dynamic UI ($Mdn = 3$) scored higher than the simple UI ($Mdn = 2, W = 152.5, p = .009$). This suggests that dynamic interaction did not help participants feel personally engaged in the continuation compared to simple UI. No significant differences were found for control (*"I had control on the final outcome"*), satisfaction (*"I feel satisfied with the final result"*), or style (*"The ending of the story reflects my personal writing style"*).

*5.3.2 Authorship, Ownership, and Control.* Participants emphasized that authorship remained with them, describing the tool as supportive rather than replacing their creative agency. P7 stressed: *"Even though it is an AI, the original thought is mine. I'm just modifying the way I want my sentence to be perceived by the audience."* Real-time feedback further supported this sense of control; as P4 explained, *"I can control the emotion or direction of the story. Yes, I did feel that."* At the same time, several participants pointed out limitations when the system's output did not align with their intent. P10 noted: *"I would have preferred finer control, because one of the sentences did not reflect the emotion correctly,"*. These experiences suggest that authorship and agency remained intact, though occasional imprecision disrupted participants' sense of control.

*5.3.3 Collaboration with AI.* Most participants described the AI as a helpful assistant that supported their work without taking over creative ownership. P8 explained: *"It was an assistant that helped me to finish my story in a better way."* A smaller group, however, described the tool as taking on a more active role. P14 characterized it as *"someone who gives their own idea to me, someone who shares their insight, someone who's not just going to be a yes-man,"* and P19 went further: *"I think it felt like it's in my head. So it's definitely a co-author."* These differing views suggest that while most writers saw the tool as an assistant, some experienced it as a creative partner, blurring the line between assistance and co-authorship.

## 5.4 Granularity of Control

Participants reflected on the unit of text at which emotional editing should occur. Many were satisfied with sentence-level manipulation, finding it intuitive and sufficient for short passages. P8 noted: *"I think having the opportunity to click on a sentence and know what emotion it's expressing is sufficient, and I can alter the sentence according to the emotion that I'm preferring."* At the same time, a substantial group expressed interest in more flexible granularity, either at the level of phrases, words, or paragraphs. P2 expected selecting specific parts of a sentence: *"The user could select a certain part of a sentence ... and edit those emotions more granularly."* Others preferred paragraph-level control to capture broader context. As P15 explained: *"If it's possible to change the entire paragraph, not only one sentence, I think it will be more convenient, and it will more reflect how people actually write."* These perspectives suggest that while sentence-level control is effective for many use cases, additional options for finer or broader control would better accommodate diverse writing practices.

## 5.5 Overall feedback

In the final survey, 83.3% of participants reported preferring the dynamic UI, while 16.7% preferred the static UI, and none selected the simple UI. This distribution highlights a strong preference for real-time interaction and control. Participants' explanations in the survey responses further show these preferences and align closely with themes observed in the interview data.

*5.5.1 Comparisons Across UIs.* The dynamic UI was consistently described as novel and distinct from conventional sentiment tools. As P3 noted, *"The simple UI is similar to a regular sentiment tool like VOYANT… dynamic UI is something new."* Participants also praised the Dynamic UI for supporting creativity, expressiveness, and control. P14 highlighted its breadth: *"It allows me to explore different ideas, structure of sentences, flow of writing – all of these collectively make the whole process of creative writing more fun and fulfilling."* Others emphasized usability, with P22 stating, *"I like the dynamic UI as it adapts in real time, making interactions feel smooth and more*

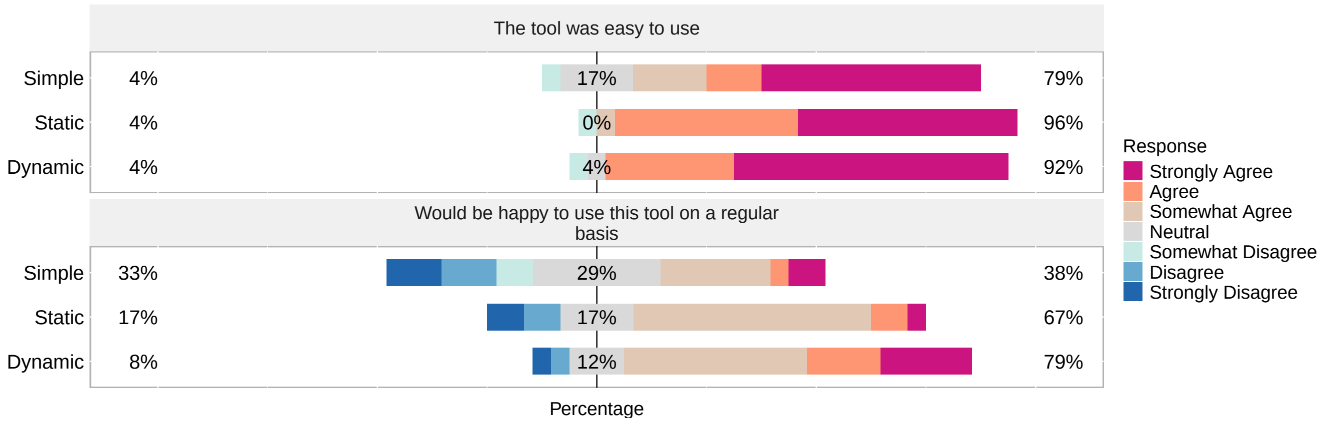


**Figure 6: Perceived usability and adoption across the three UIs.**

*human-like. It makes complex tasks feel simple and intuitive."* In contrast, a minority preferred the Static UI for its simplicity and stronger sense of authorship. As P21 explained, *"I was not happy with the rewriting option (dynamic UI)… the static UI gave me a reading of the emotional tone, and I prefer rewriting/editing myself."* These reflections align with interview findings: visualization and emotional manipulation were engaging, but mismatches between system interpretations and user intent caused frustration.

*5.5.2 Perceived Usability and Adoption.* Participants rated all three UIs as easy to use (Figure 6), with 96% of participants agreeing or strongly agreeing for the static UI, 92% for the dynamic UI, and 79% for the simple UI. Despite its additional interactivity, the dynamic UI was not perceived as harder to use. Willingness to use the tool regularly showed a clearer preference for the dynamic UI (79%), compared to the static (67%) and simple (38%) UIs. Overall, while all interfaces were considered accessible, only the dynamic version combined usability with features that motivated long-term use.

## 6 Discussion & Future Work

In this work, we investigated how interactive emotion graphs can support creative writing. Specifically, we examined (RQ1) how emotion visualizations help writers adjust emotional flow, (RQ2) how such tools influence creativity and enjoyment, and (RQ3) how real-time feedback shapes perceptions of authorship and control. Our findings point to both the potential of emotion-centered interfaces for reflection and exploration and the tensions they introduce around support and authorship.

*Limitations.* One limitation of our study was the limited time and participant pool. In order to keep the study duration manageable, we had to restrict creative writing to the completion of the given texts. While we made an effort to keep these assignments (and also the selection of participants with a degree in literature) as ecologically valid as possible, it remains interesting to see what additional aspects a long-term evaluation with professional creative writers would reveal.

*Visualization as a Creativity Scaffold.* Participants described the graph as both a mirror that revealed unintended emotional shifts and a planning aid for shaping emotional trajectories. This aligns with prior work showing that externalizing abstract dimensions of writing can support metacognition [29]. Even without AI rewriting, the visualization itself was valued for making relationships visible and surfacing alternative interpretations, suggesting that visualization alone can support reflection and divergent thinking.

*AI-Assisted Rewriting as a Catalyst for Exploration.* Reactions to AI-generated rewrites were mixed. Some participants found outputs incoherent or overly "AI-like," echoing critiques of style loss in LLM co-writing [19]. These tensions suggest that productive surprise requires balancing novelty with coherence and authorial style. At the same time, rewrites encouraged playful, divergent exploration, aligning with prior work showing that novelty can catalyze creativity [12], even when used primarily as inspiration.

*Preserving Authorship and Control.* Participants primarily framed the system as an assistant rather than a co-author, aligning with prior work showing that users accept AI suggestions while maintaining ownership of final decisions [4]. Increased system intervention reduced perceptions of sole authorship, reflecting findings from screenplay co-writing systems [19], suggesting that interactivity can both enhance engagement and complicate authorship. Differences in collaboration framing highlight an open design question around system agency. Some participants favored the AI as a critical assistant that offers alternatives, emphasizing transparency and optionality. Opaque prompting has been shown to increase metacognitive burden [1]; making AI reasoning visible and editable may help preserve writer control while supporting collaboration.

*Managing Cognitive Demands.* Generative AI tools often increase metacognitive load by requiring users to manage complex prompts [29], yet emotion graphs did not substantially increase workload in our study. NASA-TLX ratings were low across UIs, with the dynamic UI (*EmoArc*) associated with higher perceived success and manageable effort, echoing prior findings on dynamic prompting interfaces [1, 31]. Participants described the graphs as reducing editing burden and enabling quick tone adjustments, though some noted distraction when visualizations diverged from expectations, consistent with prior work on UI overload [9]. These results suggest the importance of customizable feedback levels, allowing writers to balance monitoring and intervention to match their creative flow.

*Design Opportunities and Future Directions.* Our findings point to several routes for extending emotion-centered writing support. First, multilingual emotion analysis is a promising direction, as emotions might carry culturally specific meanings [11]. Examining how emotional arcs are preserved across translations could benefit translators and multimedia applications such as video captioning. Customization and transparency are also critical, as participants requested both finer and broader control. Third, preventing over-reliance is essential. Reflective prompts or justifications for system-suggested changes can help ensure AI supports augmentation rather than substitution, allowing augmented cognition. Finally, linking emotional arcs to recommendation systems (e.g., Goodreads) could enable discovery based on emotional journeys rather than genre.

## 7 Conclusion

In this work, we introduced a system, *EmoArc*, that couples emotion visualization with AI-assisted rewriting to support creative writing. Our study with 24 participants showed that the interface enhanced awareness of emotional pacing, encouraged exploration, and supported reflection on tone, while maintaining manageable cognitive demands. Participants valued the system's transparency and responsiveness, though some raised concerns about authorship and stylistic consistency. These findings suggest that emotion-centered interfaces can enrich human-AI collaboration in writing by helping authors craft more intentional emotional trajectories and opening new directions for emotion-aware design in creative writing.

## References


[1] Rifat Mehreen Amin, Oliver Hans Kühle, Daniel Buschek, and Andreas Butz. 2025. Composable Prompting Workspaces for Creative Writing: Exploration and Iteration Using Dynamic Widgets. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 144, 11 pages. doi:10.1145/3706599.3720243

[2] Atsushi Ashida, Masataka Tokumaru, and Tomoko KOJIRI. 2019. Novel Writing Support System by Target Readers' Story Arcs and Characters' Emotional Changes. *International Conference on Computers in Education* (12 2019). doi:10.58459/icce.2019.283

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

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

[5] 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* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 209, 19 pages. doi:10.1145/3491102.3501819

[6] 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* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 77, 19 pages. doi:10.1145/3654777.3676352

[7] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 98, 13 pages. doi:10.1145/3526113.3545672

[8] Januka Dharmapriya, Lahiru Dayarathne, Tikiri Diasena, Shiromi Arunathilake, Nihal Kodikara, and Primal Wijesekera. 2021. Music Emotion Visualization through Colour. In *2021 International Conference on Electronics, Information, and Communication (ICEIC)*. 1–6. doi:10.1109/ICEIC51217.2021.9369788

[9] Krzysztof Z. Gajos, Mary Czerwinski, Desney S. Tan, and Daniel S. Weld. 2006. Exploring the design space for adaptive graphical user interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Venezia, Italy) *(AVI '06)*. Association for Computing Machinery, New York, NY, USA, 201–208. doi:10.1145/1133265.1133306

[10] E Hanser, P McKevitt, TF Lunney, and J Condell. 2010. NewsViz: emotional visualization of news stories. In *Unknown Host Publication*, D Inkpen and C Strapparava (Eds.). Association for Computational Linguistics, United States, 125–130. Proc. of the NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text ; Conference date: 01-06-2010.

[11] Shreya Havaldar, Sunny Rai, Bhumika Singhal, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023. Multilingual Language Models are not Multicultural: A Case Study in Emotion. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, Jeremy Barnes, Orphée De Clercq, and Roman Klinger (Eds.). Association for Computational Linguistics, Toronto, Canada, 202–214. doi:10.18653/v1/2023.wassa-1.19

[12] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. arXiv:2211.05030 [cs.HC] https://arxiv.org/abs/2211.05030

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

[14] Kostiantyn Kucher, Rafael, M. Martins, Carita Paradis, and Andreas Kerren. 2020. StanceVis Prime: Visual Analysis of Sentiment and Stance in Social Media Texts. *Journal of Visualization* 23, 6 (2020), 1015–1034. doi:10.1007/s12650-020-00684-5

[15] Mark E. Larsen, Tjeerd W. Boonstra, Philip J. Batterham, Bridianne O'Dea, Cecile Paris, and Helen Christensen. 2015. We Feel: Mapping Emotion on Twitter. *IEEE Journal of Biomedical and Health Informatics* 19, 4 (2015), 1246–1252. doi:10.1109/JBHI.2015.2403839

[16] Chenxiao Liu, Zheyong Xie, Sirui Zhao, Jin Zhou, Tong Xu, Minglei Li, and Enhong Chen. 2024. Speak From Heart: An Emotion-Guided LLM-Based Multimodal Method for Emotional Dialogue Generation. In *Proceedings of the 2024 International Conference on Multimedia Retrieval* (Phuket, Thailand) *(ICMR '24)*. Association for Computing Machinery, New York, NY, USA, 533–542. doi:10.1145/3652583.3658104

[17] Damien Masson, Young-Ho Kim, and Fanny Chevalier. 2025. Textoshop: Interactions Inspired by Drawing Software to Facilitate Text Editing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1087, 14 pages. doi:10.1145/3706598.3713862

[18] Damien Masson, Zixin Zhao, and Fanny Chevalier. 2025. Visual Story-Writing: Writing by Manipulating Visual Representations of Stories. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 70, 15 pages. doi:10.1145/3746059.3747758

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

[20] Saif M. Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems* 53, 4 (2012), 730–741. doi:10.1016/j.dss.2012.05.030 1) Computational Approaches to Subjectivity and Sentiment Analysis 2) Service Science in Information Systems Research : Special Issue on PACIS 2010.

[21] Saif M. Mohammad, Tony, and Yang. 2013. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. arXiv:1309.6347 [cs.CL] https://arxiv.org/abs/1309.6347

[22] Yusuke Mori, Hiroaki Yamane, Ryohei Shimizu, and Tatsuya Harada. 2022. Plug-and-Play Controller for Story Completion: A Pilot Study toward Emotion-aware Story Writing Assistance. In *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, Ting-Hao 'Kenneth' Huang, Vipul Raheja, Dongyeop Kang, John Joon Young Chung, Daniel Gissin, Mina Lee, and Katy Ilonka Gero (Eds.). Association for Computational Linguistics, Dublin, Ireland, 46–57. doi:10.18653/v1/2022.in2writing-1.6

[23] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (12 1980), 1161–1178. doi:10.1037/h0077714

[24] Marie-Luise C. R. Schmidt, Julia R. Winkler, Markus Appel, and Tobias Richter. 2023. Tracking Emotional Shifts During Story Reception: The Relationship Between Narrative Structure and Affective Responses. *Scientific Study of Literature*

(Aug 2023). doi:10.61645/ssol.177

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

[26] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 68 (Sept. 2023), 57 pages. doi:10.1145/3511599

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

[28] Sangho Suh, Jian Zhao, and Edith Law. 2022. CodeToon: Story Ideation, Auto Comic Generation, and Structure Mapping for Code-Driven Storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 16 pages. doi:10.1145/3526113.3545617

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

[30] Jason Chew Kit Tham. 2022. *Keywords in design thinking: A lexical primer for technical communicators & designers.* WAC Clearinghouse.

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

[32] Fuhai Wang. 2024. Comparative Evaluation of Sentiment Analysis Methods: From Traditional Techniques to Advanced Deep Learning Models. *Applied and Computational Engineering* 105 (2024), 23–29. https://www.ewadirect.com/proceedings/ace/article/view/16859

[33] Jiaqi Wang, Tianyi Gui, Mingzhi Cheng, Xuan Wu, Ruolin Ruan, and Meng Du. 2023. A survey on emotional visualization and visual analysis. *Journal of Visualization* 26, 1 (2023), 177–198. doi:10.1007/s12650-022-00872-5

[34] Katie Watson, Samuel S. Sohn, Sasha Schriber, Markus Gross, Carlos Manuel Muniz, and Mubbasir Kapadia. 2019. StoryPrint: an interactive visualization of stories. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 303–311. doi:10.1145/3301275.3302302

[35] Jian Zhao, Liang Gou, Fei Wang, and Michelle Zhou. 2014. PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 203–212. doi:10.1109/VAST.2014.7042496

[36] Chris Zimmerman, Mari-Klara Stein, Daniel Hardt, Christian Danielsen, and Ravi Vatrapu. 2016. emotionVis: Designing an Emotion Text Inference Tool for Visual Analytics. In *Tackling Society's Grand Challenges with Design Science*, Jeffrey Parsons, Tuure Tuunanen, John Venable, Brian Donnellan, Markus Helfert, and Jim Kenneally (Eds.). Springer International Publishing, Cham, 238–244.

# A Supplementary Material

## A.1 Example Scenario: Writing a Short Story

To illustrate how *EmoArc* supports writers in practice, we present a step-by-step walkthrough with some of our features described in Section 3. Marina is preparing a short story for her creative writing group, where the weekly assignment is to write a reunion scene filled with strong emotions. Marina wants her story to convey grief but also a sense of relief, and she turns to *EmoArc* to help refine the emotional flow. The following steps show how she moves through the cycle of write → visualize → adjust → rewrite → reflect, using different components of the interface.

*A.1.1 Write (Intuitive Drafting).* When Marina opens the system, both the **text editor** (center) and the **visualization panels** (bar graph on the right, line graph on the left) are empty. She begins drafting directly in the editor:

> *"When Elena saw her brother at the station, her eyes filled with tears she had been holding back for years."*

*EmoArc* automatically segments the text into sentences. After a short pause, the first analysis runs and the visualizations appear. She then writes a few more sentences. Marina wonders whether the tone she imagined actually comes through.

*A.1.2 Visualize (Externalizing Emotions with Linked Graphs).* She clicks on the first sentence in the **editor**. A grey highlight appears on that sentence; the same sentence is linked in the **line graph** (left) via a synchronized window, and its emotion intensities appear in the **bar graph** (right). The bars reveal sadness as the strongest emotion (hover tooltips show exact percentages). The line graph shows a peak of sadness near the beginning of the story. Both views make Marina realize that while grief is captured, the sense of relief she hoped for is missing.

*A.1.3 Adjust (Exploring Alternatives in the Bar Graph).* To experiment, Marina drags the sliders in the **bar graph**, lowering sadness and raising joy. Intensities normalize across categories and animate smoothly, giving immediate visual feedback. The text in the editor is not changed yet; instead, the bar graph displays the *target* distribution for the selected sentence, and the Rewrite button becomes the next action to apply these changes.

*A.1.4 Rewrite (AI and Manual Editing).* Marina presses Rewrite. *EmoArc* generates a variant that replaces the highlighted sentence in the **editor** for inspection:

> *"When Elena spotted her brother at the station, tears burst forth, but this time they sparkled with relief and happiness."*

The **bar graph** updates to the new emotional balance, and the **line graph** reflects a gentler rise of joy alongside sadness. She compares this with the original by hovering bars and scanning the line graph trend.

She proceeds to the next sentence, typing:

> *"She stepped forward, her chest tight with fear as she reached out to touch his hand."*

She lowers fear and raises surprise in the **bar graph**, then clicks **Rewrite**. The AI proposes:

> *"She stepped forward, startled by the warmth in his hand she had almost forgotten."*

Marina does not like this option. First, she clicks Undo, which restores the previous sentence and returns the **bar** and **line** graphs to their prior states. She then rewrites manually in the **editor**:

> *"She reached out, surprised by the steady warmth of his hand after so many years apart."*

Both graphs update immediately, confirming that manual edits are fully integrated into the feedback loop.

*A.1.5 Reflect (Assessing the Narrative Arc).* To assess the story as a whole, Marina turns to the **line graph**. She uses the **legend** to isolate sadness and joy, while other curves fade into the background. The overview now shows a smoother transition from grief to relief across the reunion passage. She clicks a point in the line graph to jump back to a sentence in the **editor** and skim through the sentence. Satisfied with the pacing and balance, she rereads the passage and prepares it for her writing group.

This process shows how Marina began with intuition, used *EmoArc* to explore and experiment, narrowed down her choices, and ultimately crafted a story whose emotional arc she deliberately shaped. Moving through the cycle of writing, visualizing, adjusting, rewriting, and reflecting gave her both freedom and control.

## A.2 Technical Details

*A.2.1 Model Selection for Emotion Classification.* We based our system on the widely used Google GoEmotions dataset[3] because it provides a fine-grained taxonomy of 28 emotions, far beyond coarse polarity or Ekman's six basic categories[4]. We first compared transformer-based classifiers trained on the GoEmotions dataset (see Table 1). While *bert-base-goemotions*[5] reported slightly higher macro F1, it relied on the older BERT-base architecture and showed weaker accuracy. *modernbert-large-go-emotions*[6] achieved the highest accuracy but demanded substantially more computational resources, making it impractical for interactive use. Balancing accuracy, efficiency, and usability, we selected *roberta-base-go_emotions*[7], which offers reliable accuracy, efficient inference, compatibility with the GoEmotions taxonomy, and strong community adoption. To validate this choice, we conducted multiple tests with different sentences. One example of that is this sentence:

> *"Hagrid let out a booming laugh that made several owls in the nearby shop hoot in alarm."*

However, Ekman-based model misclassified it as overwhelmingly containing fear (0.896), which contradicted the intended light-hearted tone. By contrast, *roberta-base-go_emotions* returned a nuanced distribution: amusement (0.415), neutral (0.238), joy (0.117), and only low fear (0.079), capturing the intended affect more faithfully. This illustrates both the limitations of coarse taxonomies and the advantages of a fine-grained model for creative writing contexts.

[3] https://research.google/blog/goemotions-a-dataset-for-fine-grained-emotion-classification/
[4] https://doi.org/10.1080/02699939208411068
[5] https://huggingface.co/IsaacZhy/bert-base-goemotions
[6] https://huggingface.co/cirimus/modernbert-large-go-emotions
[7] https://huggingface.co/SamLowe/roberta-base-go_emotions

*A.2.2 Emotion Analyzer.* Based on the *roberta-base-go_emotions* trained on the GoEmotions dataset, it segments text into sentences, predicts 28 emotion categories per sentence, and applies optimizations such as caching, parallel batch processing, and GPU acceleration. Results are returned as structured JSON objects that link each sentence to its predicted emotion intensities, original text, and position in the document (see Section A.3). In our benchmarking, the analysis of typical paragraphs (< 300 words) completed in under two seconds, and sentence rewrites were returned in 3-5 seconds.

*A.2.3 Sentence Generator.* The sentence generator uses OpenAI's gpt-4.1-nano API to generate emotionally adjusted sentence variants. It employs structured prompts, batch processing, and filtering based on RMSE and top-3 emotion accuracy (details of the generation procedure and sentence structure in Section A.2.5).

*A.2.4 Connecting Text and System.* A dedicated data service (see Section A.3.2) connects the frontend and backend. It supports three primary actions: (1) analyzing emotions in user text, (2) uploading `.txt` files for batch processing, and (3) generating sentences based on adjusted profiles. All interactions are asynchronous to avoid blocking the interface, ensuring that writing and exploration remain fluid.

*A.2.5 Sentence Generation.*

- *Top-3 Emotion Match* ensures the three most dominant target emotions are preserved within a deviation threshold of 0.12.
- *Root Mean Square Error (RMSE)* quantifies the difference between the target emotional distribution and the generated sentence, with a cutoff of $\leq 0.12$. If no candidate satisfies both thresholds, the best available option is returned. This filtering guarantees emotional fidelity while maintaining responsiveness.
- *Prompt Design:* To guide the language model in producing emotionally aligned rewrites, we developed a structured prompt template. Each prompt includes (1) the original sentence, (2) the top three target emotions and their intensity values, and (3) explicit rewriting instructions. The instructions emphasize retaining the original context and meaning, matching the specified emotional levels, and avoiding clichés or direct references to emotion words. The expected output is a valid JSON object containing the rewritten sentence. The input sent to the model is structured according to a standardized prompt, shown in Table 3.

## A.3 Sentence Structure

The output structure is a JSON object that contains a list of sentences, each with its emotion levels. Every sentence follows this structure:

```
{
  sentence: "A wizard, an' a great one at that, I'd wager," Hagrid
       said,
    his chest swelling with pride as if Harry's achievements were
         his own.
  emotions: {
    "admiration": 0.3197,
    "approval": 0.0800,
    ...
    "pride": 0.2228
  },
  index: 10,
  originalSentence: "A wizard, an' a great one at that, I'd wager,"
       Hagrid..."
}
```

The data object contains the following fields:

- `sentence`: Represents the analyzed version of the sentence.
- `emotions`: Stores the predicted emotion intensities associated with the sentence, with each key corresponding to a specific emotion.
- `index`: Indicates the sentence's original position within the text, enabling alignment with the line graph and the text editor.
- `originalSentence`: Retains the unedited version of the sentence, allowing users to revert any modifications and restore the original text when needed.

*A.3.1 Backend.* Since the system is designed for real-time interaction, several optimizations were introduced:

- *Caching.* Emotion intensities for previously analyzed sentences are cached locally and on the backend to avoid redundant processing. This reduces both latency and server load.
- *Batch Processing.* Sentences are split into chunks and processed in parallel, improving throughput for longer documents. Batch sizes adapt to available hardware to maximize efficiency without overwhelming memory.
- *Rewrite Pipeline.* The Sentence Generator limits candidate sentences to 100 tokens, generates them asynchronously in batches of 50, and uses up to 24 concurrent requests. These measures ensure quick turnaround times while still producing sufficiently diverse variants.

Together, these measures reduce waiting time and keep the editing process smooth.

*A.3.2 Data Service.*

*A.3.3 Supporting Modules.* Two additional services enhance usability and reliability, and simultaneously provide help as well as information for debugging and plausibility checks during the evaluation:

*Logging Service.* Records all user interactions, such as uploads, rewrites, and emotion edits. Logs are batched on the frontend and filtered on the backend (ignoring changes below 10%). Metadata includes session IDs, intended versus actual emotion values, and action counts.

*Tutorial Controller.* The Tutorial Controller shows a pop-up guide the first time someone uses the interface, helping them understand how to use the tool. This tutorial can also be reopened anytime by clicking the help button, marked with a question mark in the bottom-left corner. When the tutorial appears, the rest of the screen is blurred to help users focus on the guide. To make it accessible for people using screen readers, a short message is announced so they know the tutorial is open. This message is removed right after to keep things running smoothly behind the scenes.

**Table 1: Model Selection for Emotion Classification. The performance of transformer-based models fine-tuned on the GoEmotions dataset, as reported on Hugging Face. Macro F1 is computed as the average across 28 emotion classes.**

| Model | Macro F1 | Accuracy |
|---|---|---|
| SamLowe/roberta-base-go_emotions | 0.450 | 0.474 |
| IsaacZhy/bert-base-goemotions | 0.573 | 0.438 |
| cirimus/modernbert-large-go-emotions | 0.550 | 0.968 |

**Table 2: Overview of `DataService` Functions**

| Function | Endpoint | Method | Description | Input | Output |
|---|---|---|---|---|---|
| `analyzeText(text)` | `/analyze` | POST | Submits raw user text to the backend for sentence-level emotion classification. | `text` (string) | JSON object containing detected emotions per sentence. |
| `uploadFile(file)` | `/upload` | POST | Handles file uploads, allowing users to submit documents for batch emotion analysis or parsing. | `file` (File object) | JSON object confirming upload status or containing parsed content. |
| `modifySentence (sentenceData)` | `/modify` | POST | Generates a modified sentence based on user-adjusted emotion values and optional context. | `sentence` (string), `new_emotions` (object), `context` (string) | JSON object with the modified sentence or an error message. |

**Table 3: Structured prompt design used for sentence rewriting with the OpenAI API.**

| Component | Content / Instruction |
|---|---|
| Original sentence | "A wizard, an' a great one at that, I'd wager," Hagrid said, his chest swelling with pride. |
| Target emotions | { "pride": 0.43, "admiration": 0.32, "joy": 0.21 } |
| Rewrite instructions | • Retain the original context and meaning.<br>• Keep length similar to the original sentence.<br>• Match the top three emotions as closely as possible.<br>• Adjust tone and word choice to align with emotions, without introducing new ideas. |
| Important notes | • Focus only on the top three emotions.<br>• Ensure grammar and natural flow.<br>• Avoid exaggeration, clichés, or direct emotion words. |
| Output format | JSON object with one key `sentence`, e.g., {"sentence": "Your generated sentence here"}. No extra text or formatting allowed. |

## A.4 Participants

In designing our user study, we recruited participants who either held a degree in literature or a closely related discipline or who had writing experience to ensure domain-relevant expertise. A total of 24 participants took part in this study, recruited through convenience sampling. Participants ranged in age from 18 to 45 years (MD = 29.36, SD = 6.70). Of the 24 participants, 10 self-identified as female and 13 as male and 1 preferred not to say. Regarding English proficiency, 2 participants reported native C2-level fluency, 15 indicated C1-Advanced proficiency, 6 reported B2-Upper Intermediate proficiency, and 1 participant had B1-Intermediate proficiency.

In terms of writing experience, 4 participants described themselves as professional writers, 11 identified as hobbyists, and 9 reported writing occasionally for personal or academic purposes. Participants reported various purposes for which they use AI writing tools: 16 used them for editing and proofreading, 14 for idea generation, 12 for seeking alternative phrasings, 12 for content expansion, and 9 for creative writing. Three participants indicated that they had never used any AI for writing purposes.

Regarding specific AI tools used in creative writing, 21 participants mentioned ChatGPT, 12 cited Quillbot, 4 used Claude, 3 reported using DALL-E, and 8 indicated using other tools. Three participants stated they had never used any AI tools in the context of creative writing. Concerning the frequency of AI use, 12 participants reported daily use, 4 weekly, 1 monthly, 5 rarely, and 2 never. 17 participants reported engaging in creative tasks, including 15 who wrote stories, 13 who wrote letters, 12 who maintained blogs, 11 who composed emails, 10 who engaged in copywriting, and 8 who created how-to guides. When asked about their attitudes toward AI in creative writing, 16 participants believed that AI can support creative writing, whereas 5 participants disagreed with this notion, and 3 remained neutral.

Of the 24 participants, 9 either held a degree or were studying in English literature or a closely related field, while the remaining participants came from various other backgrounds.

## A.5 Writing Tasks

## A.6 Additional Measures

**Table 4: Topics for the writing tasks in the user study. Text completion exercises curated by a domain expert in literary studies.**

| Task | Text and instruction |
|---|---|
| 1 | The news felt like a pill to Shaina's heart, liberating her lungs to breathe in the air that had become fresh in the blink of an eye. The moments that suffocated her almost to death had vanished like a nightmare as she woke up to see her father still alive and walking out of the confinement. She had no tears in her eyes; it was all disbelief. Still, she stepped ahead, trembling to touch him. Was it a fantasy? **Now, in 2-3 sentences, complete the story. What does Shaina do or feel next in this moment of relief?** |
| 2 | In 1987, right after I recovered from the life-threatening injuries I got in an accident, I made myself a promise to visit graveyards and hospitals more often. I didn't keep that promise. Now, once again, I lie here, unable to move, to stand, to breathe. In my thoughts, I wait quietly for someone to take me from this hospital bed, once more, to the graveyard. This time. . . **Now, in 2-3 sentences, complete the story. What thoughts or emotions pass through the narrator's mind in these final moments?** |
| 3 | Riyad chose to embrace the depth of tragedy, to live as a dark figure in the background. He was willing to learn from pain and rise above hatred with quiet strength. When people admired him, they helped him grow into who he became. But he never saw it coming, that the same people could turn against him, punishing him for trying to bring change for the better. It started with a sense of pain but. . . **Now, in 2-3 sentences, complete the story. How does Riyad respond to the betrayal, and what final decision does he make?** |

**Figure 7: Raw NASA-TLX scores by factor (N=24).**

**Table 5: The raw NASA-TLX results from the lab study.**

| | Simple | | Static | | Dynamic | |
|---|---|---|---|---|---|---|
| **Factor** | **M** | **SD** | **M** | **SD** | **M** | **SD** |
| Mental Demand | 3.96 | 2.14 | 3.88 | 1.94 | 4.00 | 1.96 |
| Temporal Demand | 2.75 | 1.70 | 2.58 | 1.67 | 2.54 | 1.74 |
| Success | 4.58 | 1.93 | 4.92 | 1.82 | 5.42 | 1.64 |
| Effort | 4.54 | 2.06 | 3.29 | 1.76 | 3.67 | 1.95 |
| Annoy | 2.12 | 1.54 | 2.12 | 1.54 | 1.54 | 1.06 |
| Physical Demand | 2.46 | 2.02 | 1.88 | 1.75 | 1.83 | 1.40 |
| Overall NASA-TLX (Sum) | 20.42 | 6.88 | 18.67 | 5.54 | 19.00 | 6.06 |
| Overall NASA-TLX (Mean) | 5.83 | 1.97 | 5.33 | 1.58 | 5.43 | 1.73 |

**Table 6: The Creativity Support Index (CSI) results from the lab study.**

| | Simple | | Static | | Dynamic | |
|---|---|---|---|---|---|---|
| **Factor** | **M** | **SD** | **M** | **SD** | **M** | **SD** |
| Enjoyment | 11.04 | 5.37 | 13.58 | 4.99 | 16.33 | 3.18 |
| Exploration | 7.96 | 6.62 | 13.58 | 5.41 | 15.58 | 3.76 |
| Expressiveness | 9.33 | 6.83 | 13.17 | 5.48 | 14.13 | 4.51 |
| Immersion | 13.21 | 5.63 | 12.00 | 6.22 | 11.54 | 6.84 |
| Results Worth Effort | 10.21 | 6.71 | 13.79 | 4.73 | 15.17 | 4.45 |
| Overall CSI Score | 52.78 | 27.39 | 67.92 | 21.75 | 74.83 | 18.59 |