

Towards Full Authorship with AI: Supporting Revision with AI-Generated Views

Jiho Kim^{1,*}, Ray C. Flanagan¹, Noelle E. Haviland¹, ZeAi Sun¹, Souad N. Yakubu¹, Edom A. Maru¹ and Kenneth C. Arnold¹

¹Calvin University, 3201 Burton St SE, Grand Rapids, MI, 49546, USA

Abstract

Large language models (LLMs) are shaping a new user interface (UI) paradigm in writing tools by enabling users to generate text through prompts. This paradigm shifts some creative control from the user to the system, thereby diminishing the user's authorship and autonomy in the writing process. To restore autonomy, we introduce Textfocals, a UI prototype designed to investigate a human-centered approach that emphasizes the user's role in writing. Textfocals supports the writing process by providing LLM-generated summaries, questions, and advice (i.e., LLM views) in a sidebar of a text editor, encouraging reflection and self-driven revision in writing without direct text generation. Textfocals' UI affordances, including contextually adaptive views and scaffolding for prompt selection and customization, offer a novel way to interact with LLMs where users maintain full authorship of their writing. A formative user study with Textfocals showed promising evidence that this approach might help users develop underdeveloped ideas, cater to the rhetorical audience, and clarify their writing. However, the study also showed interaction design challenges related to document navigation and scoping, prompt engineering, and context management. Our work highlights the breadth of the design space of writing support interfaces powered by generative AI that maintain authorship integrity.

Keywords

Human-centered AI, human-LLM interaction, writing tools

1. Introduction

Large language models (LLMs) can produce texts comparable to those written by competent human writers [1, 2, 3]. Two interaction techniques currently dominate: dialogue (e.g., OpenAI's ChatGPT and Google's Gemini) and predictive text completion (e.g., GitHub Copilot). Although these paradigms work well for many tasks, in a writing context, both entail delegating some or all of the creative decision-making to the system. In other words, these systems embody the principle that *the system originates the content*. For example, as demonstrated by OpenAI to introduce ChatGPT, users can specify a desired output based on a goal, such as "help me write a short note to introduce myself to my neighbor," and the chatbot will respond with a


Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA


*Corresponding author.

✉ jihokim8@acm.org (J. Kim); rcf6@calvin.edu (R. C. Flanagan); neh8@calvin.edu (N. E. Haviland); zs35@calvin.edu (Z. Sun); sny2@calvin.edu (S. N. Yakubu); eam43@calvin.edu (E. A. Maru); kcarnold@alum.mit.edu (K. C. Arnold)

🌐 <https://jihokim.dev/> (J. Kim); <https://kenarnold.org/> (K. C. Arnold)

🆔 0000-0002-1434-4440 (J. Kim); 0000-0003-3892-9870 (K. C. Arnold)

 © 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

generated text, such as a generic template for a note to a neighbor¹. This emphasis on LLMs originating written content diminishes the user’s creativity and independence in the writing process, potentially hindering the development of original ideas or unduly influencing the perspectives expressed [4, 5].

In this work, we propose a human-centered approach to designing interactive LLM-powered systems that support the writing process. We introduce Textfocals, a UI prototype conceived with the principle that *the user originates the content*. Our design prevents the LLM from modifying the user’s writing and discourages situations where users might copy and paste content generated by the system into their writing. Instead, our interface motivates users to reflect on their text with LLM-generated summaries, questions, and advice on writing (which we refer to as LLM *views*), helping them discover opportunities for improvement or elaboration. Specifically, compared with a general-purpose dialogue UI like ChatGPT, our prototype provides two UI affordances to reduce both the *physical* and *cognitive* load of generating views. First, since writers often focus their revision effort on small sections such as paragraphs, Textfocals adapts its views to the region where the writer is currently revising. This contextual adaptation, done by integrating into a professional text editing tool, aims to reduce physical effort compared to having the user provide appropriate context to a dialogue interaction (e.g., by copy and paste of users’ writing into the chatbot). Second, since it is cognitively challenging to compose prompts that get LLMs to generate appropriate views, we provide scaffolding for users to select or modify pre-engineered prompts, which also helps users discover LLM capabilities relevant to their current revision needs. Previous works have explored a UI that presented LLM-generated summaries in an interactive sidebar within a text editor to assist with reverse outlining [6], and a UI for supporting template-based prompt engineering for better user-defined feedback on writing [7, 8]. However, Textfocals is the first prototype to explore UI affordances for using LLMs to facilitate human reflection and discovery for making independent and self-directed revisions in writing.

We conducted a formative user study with four participants to qualitatively evaluate the effectiveness of LLM-generated summary, inquisitive, and advisory views in supporting writing revision. Our study revealed that LLM views can help users develop underdeveloped ideas, cater to their rhetorical audience, and improve clarity in writing. Specifically, participants found *summary* views helpful for restructuring and expanding their writing (consistent with Dang et al.’s [6] findings), *inquisitive* views helpful for considering audience interpretations, and *advisory* views helpful for discovering areas for potential improvement.

In summary, our work makes the following contributions:

- **Design and implementation of Textfocals, a UI prototype for a writing tool that facilitates reflection and discovery for making independent revisions in writing.** The prototype provides the following UI affordances: (1) a scaffolding for LLM view-generating prompts that users can use or customize and (2) a sidebar of views that contextually adapts to the part of the text being revised by the user.
- **Formative insights from a user study.** Our findings show that Textfocals can help users maintain creative control and authorship of their writing through LLM views, which

¹<https://openai.com/blog/chatgpt>

help users refine incomplete ideas, tailor their writing to their intended audience, and improve the clarity of their writing.

2. Background and Related Work

2.1. What Does It Mean to Revise?

Our definition of revision largely aligns with that of Jill Fitzgerald, who defines it as “making any changes at any point in the writing process. It involves identifying discrepancies between intended and instantiated text, deciding what could or should be changed in the text and how to make desired changes, and operating, that is, making the desired changes. Changes may or may not affect meaning of the text, and they may be major or minor. Also, changes may be made in the writer’s mind before being instantiated in written text, at the time text is first written, and/or after text is first written” [9]. In other words, revision means critically examining and evaluating, which we refer to as *reflection*, and identifying any opportunities for improvement or further development, which we refer to as *discovery*, and then making the appropriate changes. This can occur at any stage of the writing journey. Our system leverages an LLM to support reflection and discovery to facilitate autonomous revision in writing.

2.2. Reflection and Discovery as Facilitators of Independent Writing

In the introduction, we defined *views* as LLM-generated summaries, questions, and advice on writing. This definition is informed by research in writing, which indicates that receiving feedback from peers and teachers can improve the quality of revisions in composition [9, 10, 11]. Although the terms *views* and *feedback* may have similar underlying semantics, we use the former term to emphasize the conclusions of the writing research conducted by Nelson and Schunn [12]. They found that the quality of revision depends on the type of feedback received and whether it contributes to the comprehension of the problem in the text, leading to high-quality revision. Specifically, feedback that summarizes the text helps writers see where the reader might not understand what the writer intended. Additionally, feedback that identifies an issue and offers a potential solution, especially during the earlier draft stages, helps the writer comprehend the problem in the text. Therefore, not all types of feedback are necessarily useful views, and the role of views is to help the writers scrutinize the text at hand to identify parts of the text that could be improved and understand why they could be improved.

Furthermore, previous work by Dang et al. [6] found that automatic summaries help facilitate reflection on writing. However, their system primarily supported structural revision, while our system aims to support content revision. Moreover, a related work by Benharrak et al. [7] investigates a system that helps users define the persona of their rhetorical audience, which can be used to prompt GPT-3.5 to adopt a specific persona when generating feedback. They found that it motivated users to make changes to their text that would resonate better with their readers. Our system aims to generate views that could potentially facilitate thoughtful examination and analysis (i.e., reflection) and potentially lead to the discovery of new insights. This, in turn, potentially motivates high-quality and original revision that better serves the audience. Indeed, according to Hayes, “revisions that are stimulated by the discovery of new

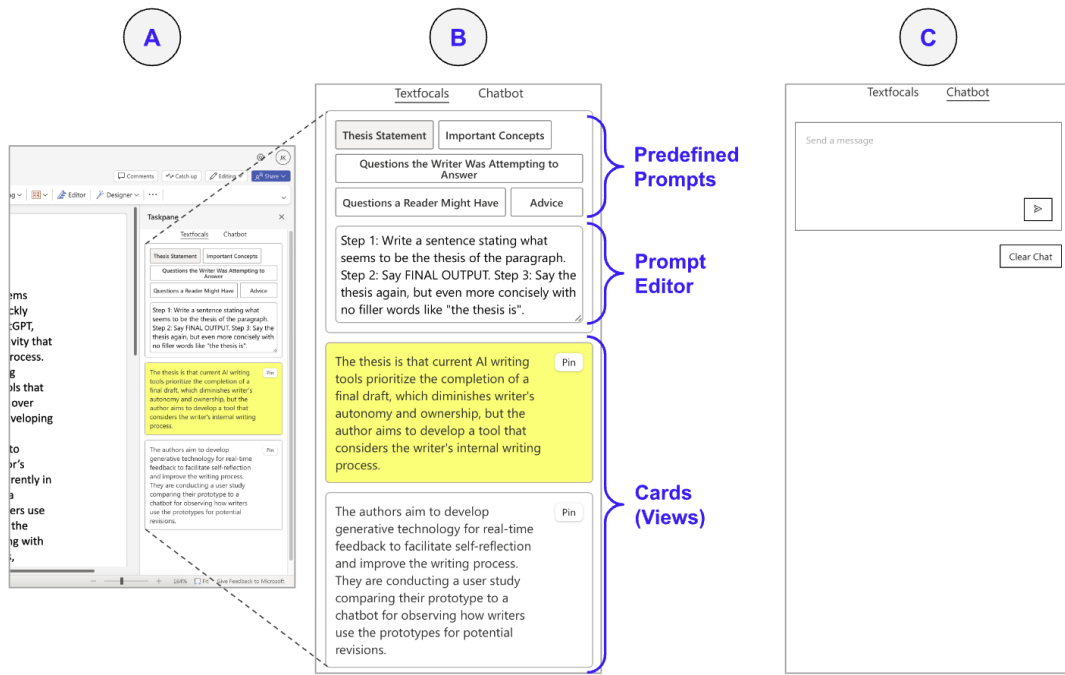


Figure 1: Our Textfocals prototype implemented as a Microsoft Word add-in. Users can interact with the prototype as a “Taskpane” in their Microsoft Word document, as shown in (A). Textfocals includes grouped buttons of predefined prompts, a prompt editor, and cards (views), as shown in (B). The chatbot prototype developed for the pilot user study has a conventional chatbot UI, as shown in (C).

connections, new ideas, or new arguments seem intrinsically more interesting. Such revisions are likely to be associated with improvements in the substance rather than the form of the text. They may mark those occasions when the writer learns something through the act of writing. Indeed, when I was revising this text, there were several occasions when revisions were triggered by discoveries. In many cases, these discoveries were stimulated by editor’s comments but, in others, they were stimulated simply by re-reading the text” [13].

3. Design and Implementation

As outlined in the previous section, we have identified reflection and discovery as core cognitive processes linked to independent revision in writing. However, even though LLM-powered writing tools such as OpenAI’s ChatGPT and Google’s Gemini can generate text similar to that of skilled human writers, these systems burden users with the responsibility of asking the right questions to the LLM for the best response (i.e., prompt engineering) and incorporating the response into their writing while still maintaining their desired level of authorship and creative control over it. To address these challenges, we aim to provide (1) a menu of preprogrammed but customizable prompts that ask the LLM to “observe” the text instead of producing text and (2) a sidebar of cards that help users interact with the LLM’s response (i.e., views) in the context of their writing. Figure 2 illustrates the overall flow of interaction in Textfocals.

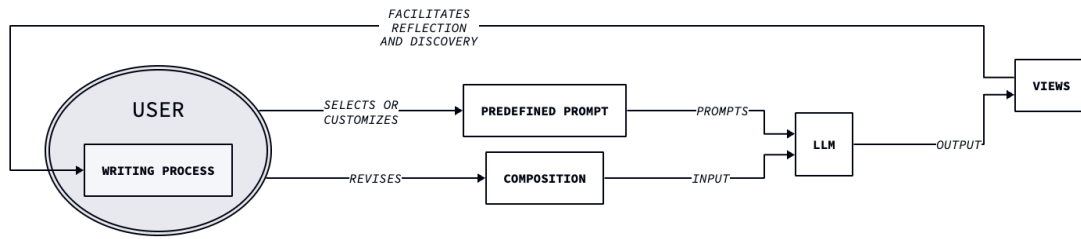


Figure 2: An overview of a user’s interaction with Textfocals: the user selects or customizes a predefined prompt. This prompts the LLM to generate views that the user can utilize for reflection and discovery in their writing process. Ultimately, this may result in the user’s revision of the composition.

Our prototype was developed as a Microsoft Word taskpane add-in using React and Microsoft’s JavaScript API. The add-in listens for cursor position changes in the document, queries the document for the text of the containing paragraph, then queries the backend (a Python FastAPI server) with the prompt and document text. The backend, in turn, queries GPT-3.5 using the OpenAI API and streams its output back to the frontend, which does lightweight parsing (such as Markdown rendering) to show the generated views.

3.1. Predefined Prompts and Prompt Editor

To support users with prompt engineering, Textfocals includes a button group to select from predefined prompts (see Figure 1). These prompts request the LLM to provide observations on the user’s text (i.e., views) instead of generating document text. The prompt buttons are labeled with short summaries of their functionality, but we additionally provide a prompt editor that allows users to see and edit the underlying predefined prompts (see Figure 1). Our intention behind this approach is to inspire users to modify the prompts, encouraging them to ask the LLM to empower their writing process instead of writing for them. We have identified multiple categories of views that can be useful for reflection and discovering areas for improvement in writing. These categories include summary views that summarize the main thesis and important concepts, inquisitive views that pose questions about the text from both a reader’s and writer’s standpoint, and advisory views that provide suggestions for writing, both on a superficial and substantial level. For each category of views, we preprogrammed the prompts as follows:

Summary views:²

- **Thesis Statement:** “Step 1: Write a sentence stating what seems to be the thesis of the paragraph. Step 2: Say FINAL OUTPUT. Step 3: Say the thesis again, but even more concisely with no filler words like ‘the thesis is.’”
- **Important Concepts:** “Step 1: List 10 important concepts in this paragraph, in the format 1. Concept: [concept as a complete sentence] Relevance: [relevance score, 10 best]. Step 2:

²These prompts reflect a lightweight attempt at chain-of-thought prompting; we modified the system to filter out the model’s internal dialogue by discarding text before “FINAL OUTPUT.” Additional prompt engineering could further improve these prompts.

Output FINAL OUTPUT, then a new line, then a Markdown unordered list with the 3 concepts with highest relevance, in short phrases of 2 or 3 words.”

Inquisitive views:

- **Questions the Writer Was Attempting to Answer:** *“List 2 or 3 questions that the writer was attempting to answer in this paragraph.”*
- **Questions a Reader Might Have:** *“As a reader, ask the writer 2 or 3 questions about definitions, logical connections, or some needed background information.”*

Advisory views:

- **Advice:** *“What advice would you give the writer to improve this paragraph? Respond in a bulleted list.”*

3.2. Sidebar of Cards

To support user interaction with the generated views in the context of their writing, we implemented a scrollable sidebar consisting of interactive cards (see Figure 1). Our design draws inspiration from Dang et al.’s “margin annotations” design concept, which presents automated summaries as cards on a sidebar [6]. These cards are an intermediate interface connecting the user’s writing with the LLM’s views. To initiate the interaction process, the LLM is given the first paragraph of the user’s writing as input and the predefined “Thesis statement” prompt discussed earlier. When a specific section of text is later focused or selected by the user in the document, the containing paragraph containing it internally snaps to the nearest paragraph and is given as input for the LLM. The LLM also generates views for the preceding and succeeding paragraphs, enabling users to browse through the views and interact with them within the context of neighboring paragraphs. Additionally, hovering over a card enables them to highlight the associated paragraph by simply allowing for easy navigation of their document based on the generated views.

4. Study Results

To explore how LLM-generated views shape user’s writing processes, we conducted a formative user study with N=4 participants³. Participants, consisting of both staff and professors, were recruited from our university. Many of the participants identified as frequent writers who actively seek feedback on their writing. Each participant brought a draft of writing they were working on, typically about a page. Specifically, P1 was working on a newsletter, P2 was working on a grant proposal, P3 was working on an argumentative essay, and P4 was working on a blog post. They interacted with the Textfocals and chatbot interfaces (see Figure 1). Participants were encouraged to verbalize their thoughts. We analyzed transcripts and screen recordings to gain the following initial insights.

³Our Institutional Review Board approved our study procedures.

4.1. How LLM Views Can Help Writers

Three common themes that emerged in how our study participants used LLM-generated views were (1) discovering underdeveloped ideas, (2) catering to their audience, and (3) identifying opportunities to improve clarity.

4.1.1. Discovering Underdeveloped Ideas

Textfocals included prompts for summarizing the thesis and important concepts in each paragraph (as shown in Section 3.1). Some of our participants found that these summaries could not only help restructure the document (as Dang et al. [6] found) but also help draw attention to parts of their writing that they had not previously considered, thereby helping them identify ideas that could be further developed. For example, when P4 requested a list of important concepts in a short paragraph, the system identified one (“tagging the creator”) that they had mentioned in passing but realized that “Maybe that’s an idea that I could develop a little bit more,” continuing, “I’m really pleasantly surprised that... it could actually help [me] generate new ideas, not just by asking it to give me ideas, but by engaging with it....” This shows that summary views could help users in expanding their writing by highlighting specific areas that could potentially be explored in more depth.

4.1.2. Catering to the Rhetorical Audience

Summary views also helped the participants compare their readers’ understanding of their writing with their intended message and identify any discrepancies. The participants treated the summary views as though they had been written by an external reader. For example, P4 commented, “I think one thing that’s helpful [about the view] is it helps me understand how [my audience] might get one level of understanding of what they’re reading here... that AI can help me understand how readers might take in my writing and misunderstand it, or understand different levels of it.” This shows that summary views could help users revise their composition so that they clearly convey their intended message to their intended audience.

Textfocals also included inquisitive views that ask personified questions about the participant’s writing. These views also helped users acquire a deeper understanding of how their intended audience might interpret their writing differently. For instance, P3 stated that they would make changes to their composition if their current writing could not adequately answer the questions posed by the view, which could reasonably be asked by their audience. P3 likened their experience of reflecting on their writing using the view to the following analogy: “You can sometimes have a vision of ‘I’m writing this piece about this,’ and then someone else would be like, ‘Well, it’s actually about this,’ and you go, ‘Oh, yeah, it is.’” This shows that inquisitive views could also help users to revise their document to better serve their rhetorical audience.

4.1.3. Identifying Opportunities to Improve Clarity

Although the participants found advisory views generally helpful in identifying both superficial and substantive improvements that could be made to their writing, many participants also expressed a desire to see a practical example demonstrating how these improvements could

be implemented in their writing. For example, after reading an advisory view that showed “Reorganize the paragraph to flow logically and smoothly,” and reflecting on the associated text, P4 thought out loud and said, “How can I be more logical?... I’m curious to know what it means by ‘logical’,” implying that they would have appreciated an example. Such demand also led to an unexpected interaction. Most notably, while reflecting on their writing using the view: “Break the paragraph into smaller, more concise sentences to improve readability and flow,” P2 asked, “How do I ask it [Textfocals] to do that? I want the program to show me what it suggests I do.” Thus, for clarity revisions and other advice, system responses should include examples.

Dialogue-style interaction may provide a natural way for users to ask for these sorts of examples, as long as the dialogue agent is given access to sufficient context from the document. For example, P3 copied and pasted text from AI views into our chatbot interface multiple times during the study to get a practical example of the suggestion shown in the view. An improved design might provide an affordance to begin a conversation about any view that is shown, or about any part of the document.

4.2. Interaction Design Considerations

Our pilot study revealed several design challenges and opportunities for how systems can present views and allow users to interact with them.

4.2.1. Display and Navigation of Views

When systems can provide feedback that requires significant space to show, and can apply to areas of different scales in the user’s document, designers need to consider how the relationship between document and feedback is visualized and what affordances the system provides for changing what area of the document they are seeing feedback on. In our study, participants requested views focused on various parts of their document, but not all of the feedback could fit on their screens at once.

Our prototype used highlighting to visualize the part of the document that each view was associated with, but participants found this confusing because the highlights were not directly visually linked with the views and because highlighting in other situations often indicates relevance rather than focus. The cursor interaction was a confusing and misunderstood feature for the participants. For example, P1 misinterpreted the yellow highlight of the text when they hovered the cursor over it as an indication of an error and said, “I’m not quite sure if it’s highlighting it because it’s good or bad.” Similarly, P3 misunderstood the yellow highlight of the card as an indication of the most relevant view, and commented, “If I look at those and especially the ones that are marked in yellow, that I’m assuming are the highest relevancy.” Other participants, P2 and P4, commented that the feature was confusing and unintuitive, and it took some time for them to get used to the feature. So, in situations where the sidebar needs to indicate what part of the document is associated with a sidebar element, a subtle and achromatic outline might be better so as not to imply additional meaning.

4.2.2. Scoping Views to Parts of the Document

When using the sidebar, some participants found it unclear which selection of the text (i.e., the scope of the text) was being provided as input to the LLM. For instance, while scrolling through the view cards, P1 asked whether the prototype was “looking at” the entire document or only the part they had selected. Likewise, P2 asked if the prototype was “looking at” the entire document or only the paragraph they had selected. This suggests that users would have found it useful to have an explicit visual cue that made it clear which subset of the text the LLM was using as an input to generate the views.

4.2.3. Prompt Flexibility

All of our participants used several different predefined prompts for their views, demonstrating the usefulness of variety in views. Some participants additionally tried editing the predefined prompts or writing their own to address issues or needs that go beyond the predefined prompts. For example, one participant edited the “Questions a Reader Might Have” prompt to specify the type of reader; another tried repurposing the view functionality to request that the LLM improve the text for them. However, composing and managing these prompts was challenging for users. A fill-in-the-blanks approach such as that taken by FeedbackBuffet [8] could address part of this challenge, but users may also need support to craft novel types of prompts and recall them later. Further research on helping people express and refine their goals to AI systems could refine this interface.

4.2.4. Contextual Information

Our participants sometimes found the LLM-generated views irrelevant because the view did not account for the information presented elsewhere in the document. For example, the LLM would sometimes generate questions that immediately adjacent paragraphs already addressed. This behavior may have emerged in our prototype because we used only the context of the current paragraphs for generating views (since including the entire document might have exceeded the context length limit of the models we were using). Future work should consider including more context, whether by using LLMs with longer context lengths or by adaptively summarizing the context as needed. Contextual information could also include specification of the audience and other aspects of the rhetorical situation. For example, P2 complained that one of the inquisitive prompts included a question that the audience of the text would already know the answer to. Future work could explore allowing writers to specify this information, perhaps in the form of personas [7].

Our participants also expected that chat conversations in the sidebar of a document should have the content (at least the currently visible part) available for context. Our chatbot prototype did not include this, but future studies should correct this oversight.

5. Discussion

Revision is a writing process that involves the critical examination and evaluation of the writing (i.e., reflection) to identify opportunities for improvement or further development (i.e., discovery)

and make the appropriate changes [9]. By presenting the predefined prompts, which were crafted to prompt the LLM to generate what it “observes” instead of what it can replace or continue, it enabled the users to consider the LLM’s output as an external perspective that they could use to reflect on their writing and make discoveries. We found preliminary evidence that the participants in the pilot user study found the output generated in this manner useful. It helped them draw attention to concepts in their writing that they could develop further. It also assisted them in identifying gaps between their intended message in the writing and their audience’s understanding, allowing them to gain insights on how to cater their writing to their intended readers. Even though these observations are not a comprehensive list of how one can reflect on their writing and make discoveries, we note that they enable users to identify what Hayes refers to as “new connections, new ideas, or new arguments...” which propels “improvements in the substance rather than the form of the text” [13]. Therefore, our results, at least preliminarily, show that Textfocals can facilitate reflection and discovery that supports substantial revision, and help users write for their audience.

Furthermore, our results also underscore the effectiveness of predefined prompts and the prompt editor in guiding participants to prompt the LLM to generate external viewpoints on their writing rather than replacing or continuing their writing. This UI affordance effectively enables participants to leverage the LLM to empower their thinking process in writing rather than relying on the LLM to replace their own thoughts and ideas. Indeed, Hayes also notes that substantial revisions are also “associated with interesting changes... in the thinking of the text’s author” [13]. Therefore, our results also preliminarily demonstrate that Textfocals can help users maintain their full authorship and autonomy when revising their composition.

5.1. Limitations

Although our main focus was to assess Textfocals, we structured our pilot study as a within-subjects design to compare the effectiveness of new UI affordances against a chatbot interface. However, due to inconsistent study procedures, we could not draw robust conclusions. For example, even though counterbalancing was attempted to eliminate the learning effect, many participants perceived the two systems as complementary or similar, limiting our ability to assess the effectiveness of the new UI affordances.

Acknowledgments

We would like to thank Saron Melesse for her contribution to the research. We would also like to thank Dr. Keith Vander Linden and Dr. Kristine Johnson for their feedback and advice. Additionally, we would like to express our gratitude to the participants of our pilot study. This research is supported by NSF CRII award 2246145 and the Wierenga Family Foundation Summer Research Fellowship for Sciences.

References

- [1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language Models are Few-Shot Learners, 2020. doi:10.48550/arXiv.2005.14165. arXiv:2005.14165.
- [2] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Al-tenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bog-donoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, R. Campbell, A. Cann, B. Carey, C. Carlson, R. Carmichael, B. Chan, C. Chang, F. Chantzis, D. Chen, S. Chen, R. Chen, J. Chen, M. Chen, B. Chess, C. Cho, C. Chu, H. W. Chung, D. Cummings, J. Currier, Y. Dai, C. Decareaux, T. Degry, N. Deutsch, D. Deville, A. Dhar, D. Dohan, S. Dowling, S. Dunning, A. Ecoffet, A. Eleti, T. Eloundou, D. Farhi, L. Fedus, N. Felix, S. P. Fishman, J. Forte, I. Fulford, L. Gao, E. Georges, C. Gibson, V. Goel, T. Gogineni, G. Goh, R. Gontijo-Lopes, J. Gordon, M. Grafstein, S. Gray, R. Greene, J. Gross, S. S. Gu, Y. Guo, C. Hallacy, J. Han, J. Harris, Y. He, M. Heaton, J. Heidecke, C. Hesse, A. Hickey, W. Hickey, P. Hoeschele, B. Houghton, K. Hsu, S. Hu, X. Hu, J. Huizinga, S. Jain, S. Jain, J. Jang, A. Jiang, R. Jiang, H. Jin, D. Jin, S. Jomoto, B. Jonn, H. Jun, T. Kaf-tan, Ł. Kaiser, A. Kamali, I. Kanitscheider, N. S. Keskar, T. Khan, L. Kilpatrick, J. W. Kim, C. Kim, Y. Kim, H. Kirchner, J. Kiros, M. Knight, D. Kokotajlo, Ł. Kondraciuk, A. Kon-drich, A. Konstantinidis, K. Kosic, G. Krueger, V. Kuo, M. Lampe, I. Lan, T. Lee, J. Leike, J. Leung, D. Levy, C. M. Li, R. Lim, M. Lin, S. Lin, M. Litwin, T. Lopez, R. Lowe, P. Lue, A. Makanju, K. Malfacini, S. Manning, T. Markov, Y. Markovski, B. Martin, K. Mayer, A. Mayne, B. McGrew, S. M. McKinney, C. McLeavey, P. McMillan, J. McNeil, D. Medina, A. Mehta, J. Menick, L. Metz, A. Mishchenko, P. Mishkin, V. Monaco, E. Morikawa, D. Moss-ing, T. Mu, M. Murati, O. Murk, D. Mély, A. Nair, R. Nakano, R. Nayak, A. Neelakantan, R. Ngo, H. Noh, L. Ouyang, C. O’Keefe, J. Pachocki, A. Paino, J. Palermo, A. Pantuliano, G. Parascandolo, J. Parish, E. Parparita, A. Passos, M. Pavlov, A. Peng, A. Perelman, F. d. A. B. Peres, M. Petrov, H. P. d. O. Pinto, Michael, Pokorny, M. Pokrass, V. Pong, T. Pow-ell, A. Power, B. Power, E. Proehl, R. Puri, A. Radford, J. Rae, A. Ramesh, C. Raymond, F. Real, K. Rimbach, C. Ross, B. Rotsted, H. Roussez, N. Ryder, M. Saltarelli, T. Sanders, S. Santurkar, G. Sastry, H. Schmidt, D. Schnurr, J. Schulman, D. Selsam, K. Sheppard, T. Sherbakov, J. Shieh, S. Shoker, P. Shyam, S. Sidor, E. Sigler, M. Simens, J. Sitkin, K. Slama, I. Sohl, B. Sokolowsky, Y. Song, N. Staudacher, F. P. Such, N. Summers, I. Sutskever, J. Tang, N. Tezak, M. Thompson, P. Tillet, A. Tootoonchian, E. Tseng, P. Tuggle, N. Turley, J. Tworek, J. F. C. Uribe, A. Vallone, A. Vijayvergiya, C. Voss, C. Wainwright, J. J. Wang, A. Wang, B. Wang, J. Ward, J. Wei, C. J. Weinmann, A. Welihinda, P. Welinder, J. Weng, L. Weng, M. Wiethoff, D. Willner, C. Winter, S. Wolrich, H. Wong, L. Workman, S. Wu, J. Wu, M. Wu, K. Xiao, T. Xu, S. Yoo, K. Yu, Q. Yuan, W. Zaremba, R. Zellers, C. Zhang, M. Zhang, S. Zhao, T. Zheng, J. Zhuang, W. Zhuk, B. Zoph, GPT-4 Technical Report, 2023. doi:10.48550/arXiv.2303.08774. arXiv:2303.08774.

- [3] R. Anil, S. Borgeaud, Y. Wu, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, D. Silver, S. Petrov, M. Johnson, I. Antonoglou, J. Schrittwieser, A. Glaese, J. Chen, E. Pitler, T. Lillicrap, A. Lazaridou, O. Firat, J. Molloy, M. Isard, P. R. Barham, T. Hennigan, B. Lee, F. Viola, M. Reynolds, Y. Xu, R. Doherty, E. Collins, C. Meyer, E. Rutherford, E. Moreira, K. Ayoub, M. Goel, G. Tucker, E. Piqueras, M. Krikun, I. Barr, N. Savinov, I. Danihelka, B. Roelofs, A. White, A. Andreassen, T. von Glehn, L. Yagati, M. Kazemi, L. Gonzalez, M. Khalman, J. Sygnowski, A. Frechette, C. Smith, L. Culp, L. Proleev, Y. Luan, X. Chen, J. Lottes, N. Schucher, F. Lebron, A. Rrustemi, N. Clay, P. Crone, T. Kocisky, J. Zhao, B. Perz, D. Yu, H. Howard, A. Bloniarz, J. W. Rae, H. Lu, L. Sifre, M. Maggioni, F. Alcober, D. Garrette, M. Barnes, S. Thakoor, J. Austin, G. Barth-Maron, W. Wong, R. Joshi, R. Chaabouni, D. Fatiha, A. Ahuja, R. Liu, Y. Li, S. Cogan, J. Chen, C. Jia, C. Gu, Q. Zhang, J. Grimstad, A. J. Hartman, M. Chadwick, G. S. Tomar, X. Garcia, E. Senter, E. Taropa, T. S. Pillai, J. Devlin, M. Laskin, D. d. L. Casas, D. Valter, C. Tao, L. Blanco, A. P. Badia, D. Reitter, M. Chen, J. Brennan, C. Rivera, S. Brin, S. Iqbal, G. Surita, J. Labanowski, A. Rao, S. Winkler, E. Parisotto, Y. Gu, K. Olszewska, Y. Zhang, R. Addanki, A. Miech, A. Louis, L. E. Shafey, D. Teplyashin, G. Brown, E. Catt, N. Attaluri, J. Balaguer, J. Xiang, P. Wang, Z. Ashwood, A. Briukhov, A. Webson, S. Ganapathy, S. Sanghavi, A. Kannan, M.-W. Chang, A. Stjerngren, J. Djolonga, Y. Sun, A. Bapna, M. Aitchison, P. Pejman, H. Michalewski, T. Yu, C. Wang, J. Love, J. Ahn, D. Bloxwich, K. Han, P. Humphreys, T. Sellam, J. Bradbury, V. Godbole, S. Samangoeei, B. Damoc, A. Kaskasoli, S. M. R. Arnold, V. Vasudevan, S. Agrawal, J. Riesa, D. Lepikhin, R. Tanburn, S. Srinivasan, H. Lim, S. Hodgkinson, P. Shyam, J. Ferret, S. Hand, A. Garg, T. L. Paine, J. Li, Y. Li, M. Giang, A. Neitz, Z. Abbas, S. York, M. Reid, E. Cole, A. Chowdhery, D. Das, D. Rogozińska, V. Nikolaev, P. Sprechmann, Z. Nado, L. Zilka, F. Prost, L. He, M. Monteiro, G. Mishra, C. Welty, J. Newlan, D. Jia, M. Allamanis, C. H. Hu, R. de Liedekerke, J. Gilmer, C. Saroufim, S. Rijhwani, S. Hou, D. Shrivastava, A. Baddepudi, A. Goldin, A. Ozturel, A. Cassirer, Y. Xu, D. Sohn, D. Sachan, R. K. Amplayo, C. Swanson, D. Petrova, S. Narayan, A. Guez, S. Brahma, J. Landon, M. Patel, R. Zhao, K. Villela, L. Wang, W. Jia, M. Rahtz, M. Giménez, L. Yeung, H. Lin, J. Keeling, P. Georgiev, D. Mincu, B. Wu, S. Haykal, R. Saputro, K. Vodrahalli, J. Qin, Z. Cankara, A. Sharma, N. Fernando, W. Hawkins, B. Neyshabur, S. Kim, A. Hutter, P. Agrawal, A. Castro-Ros, G. van den Driessche, T. Wang, F. Yang, S.-y. Chang, P. Komarek, R. McIlroy, M. Lučić, G. Zhang, W. Farhan, M. Sharman, P. Natsev, P. Michel, Y. Cheng, Y. Bansal, S. Qiao, K. Cao, S. Shakeri, C. Butterfield, J. Chung, P. K. Rubenstein, S. Agrawal, A. Mensch, K. Soparkar, K. Lenc, T. Chung, A. Pope, L. Maggiore, J. Kay, P. Jhakra, S. Wang, J. Maynez, M. Phuong, T. Tobin, A. Tacchetti, M. Trebacz, K. Robinson, Y. Katariya, S. Riedel, P. Bailey, K. Xiao, N. Ghelani, L. Aroyo, A. Slone, N. Houlsby, X. Xiong, Z. Yang, E. Gribovskaya, J. Adler, M. Wirth, L. Lee, M. Li, T. Kagohara, J. Pavagadhi, S. Bridgers, A. Bortsova, S. Ghemawat, Z. Ahmed, T. Liu, R. Powell, V. Bolina, M. Iinuma, P. Zablotskaia, J. Besley, D.-W. Chung, T. Dozat, R. Comanescu, X. Si, J. Greer, G. Su, M. Polacek, R. L. Kaufman, S. Tokumine, H. Hu, E. Buchatskaya, Y. Miao, M. Elhawaty, A. Siddhant, N. Tomasev, J. Xing, C. Greer, H. Miller, S. Ashraf, A. Roy, Z. Zhang, A. Ma, A. Filos, M. Besta, R. Blevins, T. Klimenko, C.-K. Yeh, S. Changpinyo, J. Mu, O. Chang, M. Pajarskas, C. Muir, V. Cohen, C. L. Lan, K. Haridasan, A. Marathe, S. Hansen, S. Douglas, R. Samuel, M. Wang, S. Austin, C. Lan, J. Jiang, J. Chiu, J. A. Lorenzo, L. L. Sjöstrand, S. Cevey, Z. Gleicher, T. Avrahami, A. Boral,

H. Srinivasan, V. Selo, R. May, K. Aisopos, L. Hussenot, L. B. Soares, K. Baumli, M. B. Chang, A. Recasens, B. Caine, A. Pritzel, F. Pavetic, F. Pardo, A. Gergely, J. Frye, V. Ramasesh, D. Horgan, K. Badola, N. Kassner, S. Roy, E. Dyer, V. Campos, A. Tomala, Y. Tang, D. E. Badawy, E. White, B. Mustafa, O. Lang, A. Jindal, S. Vikram, Z. Gong, S. Caelles, R. Hemsley, G. Thornton, F. Feng, W. Stokowiec, C. Zheng, P. Thacker, Ç. Ünlü, Z. Zhang, M. Saleh, J. Svensson, M. Bileschi, P. Patil, A. Anand, R. Ring, K. Tshilas, A. Vezer, M. Selvi, T. Shevlane, M. Rodriguez, T. Kwiatkowski, S. Daruki, K. Rong, A. Dafoe, N. FitzGerald, K. Gu-Lemberg, M. Khan, L. A. Hendricks, M. Pellat, V. Feinberg, J. Cobon-Kerr, T. Sainath, M. Rauh, S. H. Hashemi, R. Ives, Y. Hasson, Y. Li, E. Noland, Y. Cao, N. Byrd, L. Hou, Q. Wang, T. Sottiaux, M. Paganini, J.-B. Lespiau, A. Moufarek, S. Hassan, K. Shivakumar, J. van Amersfoort, A. Mandhane, P. Joshi, A. Goyal, M. Tung, A. Brock, H. Sheahan, V. Misra, C. Li, N. Rakićević, M. Dehghani, F. Liu, S. Mittal, J. Oh, S. Noury, E. Sezener, F. Huot, M. Lamm, N. De Cao, C. Chen, G. Elsayed, E. Chi, M. Mahdieh, I. Tenney, N. Hua, I. Petrychenko, P. Kane, D. Scandinaro, R. Jain, J. Uesato, R. Datta, A. Sadovsky, O. Bunyan, D. Rabiej, S. Wu, J. Zhang, G. Vasudevan, E. Leurent, M. Alnahlawi, I. Georgescu, N. Wei, I. Zheng, B. Chan, P. G. Rabinovitch, P. Stanczyk, Y. Zhang, D. Steiner, S. Naskar, M. Azzam, M. Johnson, A. Paszke, C.-C. Chiu, J. S. Elias, A. Mohiuddin, F. Muhammad, J. Miao, A. Lee, N. Vieillard, S. Potluri, J. Park, E. Davoodi, J. Zhang, J. Stanway, D. Garmon, A. Karmarkar, Z. Dong, J. Lee, A. Kumar, L. Zhou, J. Evens, W. Isaac, Z. Chen, J. Jia, A. Levskaya, Z. Zhu, C. Gorgolewski, P. Grabowski, Y. Mao, A. Magni, K. Yao, J. Snaider, N. Casagrande, P. Suganthan, E. Palmer, G. Irving, E. Loper, M. Faruqui, I. Arkatkar, N. Chen, I. Shafran, M. Fink, A. Castaño, I. Giannoumis, W. Kim, M. Rybiński, A. Sreevatsa, J. Prendki, D. Soregel, A. Goedeckemeyer, W. Gierke, M. Jafari, M. Gaba, J. Wiesner, D. G. Wright, Y. Wei, H. Vashisht, Y. Kulizhskaya, J. Hoover, M. Le, L. Li, C. Iwuanyanwu, L. Liu, K. Ramirez, A. Khorlin, A. Cui, T. LIN, M. Georgiev, M. Wu, R. Aguilar, K. Pallo, A. Chakladar, A. Repina, X. Wu, T. van der Weide, P. Ponnappalli, C. Kaplan, J. Simsa, S. Li, O. Dousse, F. Yang, J. Piper, N. Ie, M. Lui, R. Pasumathi, N. Lintz, A. Vijayakumar, L. N. Thiet, D. Andor, P. Valenzuela, C. Paduraru, D. Peng, K. Lee, S. Zhang, S. Greene, D. D. Nguyen, P. Kurylowicz, S. Velury, S. Krause, C. Hardin, L. Dixon, L. Janzer, K. Choo, Z. Feng, B. Zhang, A. Singhal, T. Latkar, M. Zhang, Q. Le, E. A. Abellan, D. Du, D. McKinnon, N. Antropova, T. Bolukbasi, O. Keller, D. Reid, D. Finchelstein, M. A. Raad, R. Crocker, P. Hawkins, R. Dadashi, C. Gaffney, S. Lall, K. Franko, E. Filonov, A. Bulanova, R. Leblond, V. Yadav, S. Chung, H. Askham, L. C. Cobo, K. Xu, F. Fischer, J. Xu, C. Sorokin, C. Alberti, C.-C. Lin, C. Evans, H. Zhou, A. Dimitriev, H. Forbes, D. Banarse, Z. Tung, J. Liu, M. Omernick, C. Bishop, C. Kumar, R. Sterneck, R. Foley, R. Jain, S. Mishra, J. Xia, T. Bos, G. Cideron, E. Amid, F. Piccinno, X. Wang, P. Banzal, P. Gurita, H. Noga, P. Shah, D. J. Mankowitz, A. Polozov, N. Kushman, V. Krakovna, S. Brown, M. Bateni, D. Duan, V. Firoiu, M. Thotakuri, T. Natan, A. Mohananey, M. Geist, S. Mudgal, S. Girgin, H. Li, J. Ye, O. Roval, R. Tojo, M. Kwong, J. Lee-Thorp, C. Yew, Q. Yuan, S. Bagri, D. Sinopalnikov, S. Ramos, J. Mellor, A. Sharma, A. Severyn, J. Lai, K. Wu, H.-T. Cheng, D. Miller, N. Sonnerat, D. Vnukov, R. Greig, J. Beattie, E. Caveness, L. Bai, J. Eisenschlos, A. Korchemniy, T. Tsai, M. Jasarevic, W. Kong, P. Dao, Z. Zheng, F. Liu, F. Yang, R. Zhu, M. Geller, T. H. Teh, J. Sanmiya, E. Gladchenko, N. Trdin, A. Sozanschi, D. Toyama, E. Rosen, S. Tavakkol, L. Xue, C. Elkind, O. Woodman, J. Carpenter, G. Papamakarios, R. Kemp, S. Kafle, T. Grunina, R. Sinha, A. Talbert, A. Goyal, D. Wu,

D. Owusu-Afriyie, C. Du, C. Thornton, J. Pont-Tuset, P. Narayana, J. Li, S. Fatehi, J. Wieting, O. Ajmeri, B. Uria, T. Zhu, Y. Ko, L. Knight, A. Héliou, N. Niu, S. Gu, C. Pang, D. Tran, Y. Li, N. Levine, A. Stolovich, N. Kalb, R. Santamaria-Fernandez, S. Goenka, W. Yustalim, R. Strudel, A. Elqursh, B. Lakshminarayanan, C. Deck, S. Upadhyay, H. Lee, M. Dusenberry, Z. Li, X. Wang, K. Levin, R. Hoffmann, D. Holtmann-Rice, O. Bachem, S. Yue, S. Arora, E. Malmi, D. Mirylenka, Q. Tan, C. Koh, S. H. Yeganeh, S. Pöder, S. Zheng, F. Pongetti, M. Tariq, Y. Sun, L. Ionita, M. Seyedhosseini, P. Tafti, R. Kotikalapudi, Z. Liu, A. Gulati, J. Liu, X. Ye, B. Chrzaszcz, L. Wang, N. Sethi, T. Li, B. Brown, S. Singh, W. Fan, A. Parisi, J. Stanton, C. Kuang, V. Koverkathu, C. A. Choquette-Choo, Y. Li, T. J. Lu, A. Ittycheriah, P. Shroff, P. Sun, M. Varadarajan, S. Bahargam, R. Willoughby, D. Gaddy, I. Dasgupta, G. Desjardins, M. Cornero, B. Robenek, B. Mittal, B. Albrecht, A. Shenoy, F. Moiseev, H. Jacobsson, A. Ghaffarkhah, M. Rivière, A. Walton, C. Crepy, A. Parrish, Y. Liu, Z. Zhou, C. Farabet, C. Radebaugh, P. Srinivasan, C. van der Salm, A. Fidjeland, S. Scellato, E. Latorre-Chimoto, H. Klimczak-Plucińska, D. Bridson, D. de Cesare, T. Hudson, P. Mendolicchio, L. Walker, A. Morris, I. Penchev, M. Mauger, A. Guseynov, A. Reid, S. Odoom, L. Loher, V. Cotruta, M. Yenugula, D. Grewe, A. Petrushkina, T. Duerig, A. Sanchez, S. Yadlowsky, A. Shen, A. Globerson, A. Kurzrok, L. Webb, S. Dua, D. Li, P. Lahoti, S. Bhupatiraju, D. Hurt, H. Qureshi, A. Agarwal, T. Shani, M. Eyal, A. Khare, S. R. Belle, L. Wang, C. Tekur, M. S. Kale, J. Wei, R. Sang, B. Saeta, T. Liechty, Y. Sun, Y. Zhao, S. Lee, P. Nayak, D. Fritz, M. R. Vuyyuru, J. Aslanides, N. Vyas, M. Wicke, X. Ma, T. Bilal, E. Eltyshev, D. Balle, N. Martin, H. Cate, J. Manyika, K. Amiri, Y. Kim, X. Xiong, K. Kang, F. Luisier, N. Tripuraneni, D. Madras, M. Guo, A. Waters, O. Wang, J. Ainslie, J. Baldridge, H. Zhang, G. Pruthi, J. Bauer, F. Yang, R. Mansour, J. Gelman, Y. Xu, G. Polovets, J. Liu, H. Cai, W. Chen, X. Sheng, E. Xue, S. Ozair, A. Yu, C. Angermueller, X. Li, W. Wang, J. Wiesinger, E. Koukoumidis, Y. Tian, A. Iyer, M. Gurumurthy, M. Goldenson, P. Shah, M. K. Blake, H. Yu, A. Urbanowicz, J. Palomaki, C. Fernando, K. Brooks, K. Durden, H. Mehta, N. Momchev, E. Rahimtoroghi, M. Georgaki, A. Raul, S. Ruder, M. Redshaw, J. Lee, K. Jalan, D. Li, G. Perng, B. Hechtman, P. Schuh, M. Nasr, M. Chen, K. Milan, V. Mikulik, T. Strohmaier, J. Franco, T. Green, D. Hassabis, K. Kavukcuoglu, J. Dean, O. Vinyals, Gemini: A Family of Highly Capable Multimodal Models, 2023. doi:10.48550/arXiv.2312.11805. arXiv:2312.11805.

- [4] K. C. Arnold, K. Chauncey, K. Z. Gajos, Sentiment Bias in Predictive Text Recommendations Results in Biased Writing, in: Graphics Interface 2018, Toronto, Ontario, Canada, 2018, pp. 8–11.
- [5] M. Jakesch, A. Bhat, D. Buschek, L. Zalmanson, M. Naaman, Co-Writing with Opinionated Language Models Affects Users' Views, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–15. doi:10.1145/3544548.3581196. arXiv:2302.00560.
- [6] H. Dang, K. Benharrak, F. Lehmann, D. Buschek, Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries, 2022. doi:10.1145/3526113.3545672. arXiv:2208.09323.
- [7] K. Benharrak, T. Zindulka, F. Lehmann, H. Heuer, D. Buschek, Writer-Defined AI Personas for On-Demand Feedback Generation, 2023. doi:10.48550/arXiv.2309.10433. arXiv:2309.10433.
- [8] S. MacNeil, A. Tran, J. Kim, Z. Huang, S. Bernstein, D. Mogil, Prompt Middleware: Mapping

Prompts for Large Language Models to UI Affordances, 2023. doi:10.48550/arXiv.2307.01142. arXiv:2307.01142.

- [9] J. Fitzgerald, Research on Revision in Writing, *Review of Educational Research* 57 (1987) 481–506. doi:10.2307/1170433. arXiv:1170433.
- [10] A. S. Horning, A. Becker (Eds.), *Revision: History, Theory, and Practice*, Reference Guides to Rhetoric and Composition, Parlor Press, West Lafayette, Ind, 2006.
- [11] D. M. Murray, *The Craft of Revision*, 5th ed., Cengage Learning, Boston, MA, 2003.
- [12] M. M. Nelson, C. D. Schunn, The Nature of Feedback: How Different Types of Peer Feedback Affect Writing Performance, *Instructional Science* 37 (2009) 375–401. arXiv:23372520.
- [13] J. R. Hayes, What Triggers Revision?, in: G. Rijlaarsdam, L. Allal, L. Chanquoy, P. Largy (Eds.), *Revision Cognitive and Instructional Processes*, volume 13, Springer Netherlands, Dordrecht, 2004, pp. 9–20. doi:10.1007/978-94-007-1048-1_2.