Writing with Generative AI: Multi-modal and Multi-dimensional Tools for Journalists*

Sitong Wang
Columbia University
New York, NY, USA
sw3504@columbia.edu

Lydia B. Chilton
Columbia University
New York, NY, USA
chilton@cs.columbia.edu

Jeffrey V. Nickerson
Stevens Institute of Technology
Hoboken, NJ, USA
jnickers@stevens.edu

1 INTRODUCTION

New generative AI models expand the design space for writing assistants. These systems together with humans form a larger creative system [3]. The authors are building tools that can potentially help journalists write, as part of an NSF grant on the Future of News Work. At the CHI conference, a paper coming out of the grant will be presented: the paper describes a system for generating news angles and its evaluation by journalists [5].

Building on this work, we are exploring ways of augmenting writing assistants for use in journalism. Two complementary directions are discussed here. One direction is building generative writing assistants in conjunction with image generation in order to generate storyboards. A second is building a network graph-based interface so that ideation can be explored in a semi-structured yet non-linear way.

2 MULTI-MODAL WRITING

While many examples of generative AI focus on one system, we think there is value in integrating different AI systems. In particular, we think combining language and image models provides a salient example of multi-modal AI systems.

Moreover, it suggests a solution to a problem in newsrooms: how to generate storyboards for short videos and for investigative documentaries. It also allows us to move the focus of technological exploration from one individual large language model like GPT-3 to the universe of models and their integration, a potentially more productive territory to explore.

An emerging need for news organizations is to reach younger audiences on video-based social media platforms, like Instagram Reels and TikTok. This is challenging because videos on these platforms have narrative norms that are completely different from traditional written, or video news formats. Moreover, journalists are not trained to make these videos, as they are new and emergent forms of communication, and there are no established norms for how to create news for them. It is a domain where exploration and innovation are necessary.

Our use case is storyboarding for journalists who are working on short form videos for social media. This is a more specific use case than that explored in other recent work on storyboarding [4]. Story boarding such videos involves not only generating textual scripts but also placeholder images that can be used to conceptualize videos. The conceptually challenging part of such work is anticipating what kinds of textual prompts will generate scripts that in turn can form the basis for generating graphics.

With generative AI we can help newsrooms quickly explore new narrative forms for the stories. We argue that generative text models and generative image models are best used together to do this exploration. Text can explore conceptual spaces quickly: small words can be used to take great leaps across a semantic landscape. Images can take vague concepts and instantiate them. Together, generative text and image technologies can aid in the early stages of the design process with framing, ideation, and prototyping. They can be evaluated, and can be used in early iterations of a story idea. By speeding up these early design stages, we enable news rooms to explore faster, as well as to move into new media spaces enabled by quickly changing social media platforms.

The use of AI can be combined in more complex workflows. Figure 1 shows how the human designer mediates between language model and image model AI tools.
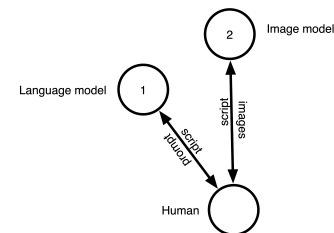
Figure 1: Human spans and integrates two generative AI systems

Figure 2 shows the output of a system called ReelFramer which interfaces to a large language model and an image model.

3 MULTI-DIMENSIONAL WRITING

There is a non-sequential workflow and thought process behind planning any larger writing project - regardless of whether the actual plot is sequential in time or not. In non-fiction writing, many authors write papers inside out, starting with a result toward the end of a paper, then backing up to provide the overall conceptual framing that highlights the contributions of the result. Most writing tools have outlines; some like Scrivener have filecard systems, usable for both fiction and non-fiction writing. Both conventions help authors hop across a manuscript. Moreover, we understand from studies of narrative that the sequential stream of words can enclose quite complicated movements across time and space. The raw chronological order of a story is often permuted into a different ordering; many plots start near the end of a story and then rewind to the beginning, not to mention the more complex arrangements made possible by flashbacks [1].
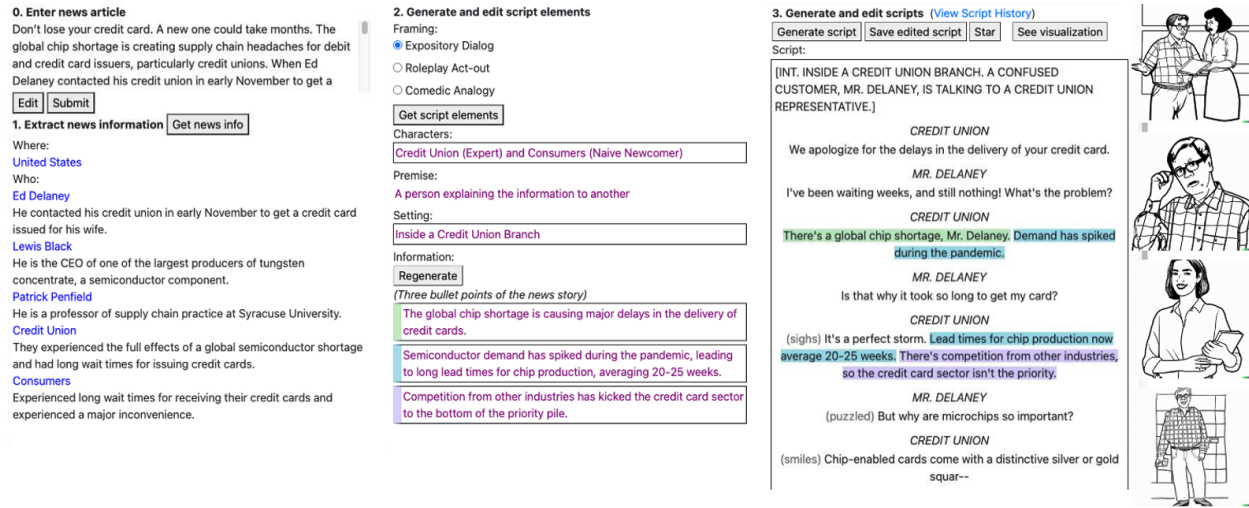

Figure 2: Output from ReelFramer

Current uses of large language models such as GPT-3, though, are challenged by the lack of long-term memory in the models. They tend to use chat-based metaphors [2]. Transformer-based systems are trained using relatively short snippets of text as prompts, and then generate relatively short snippets in response. Getting such models to help generate more than a paragraph of coherent text can be challenging, as the models do not maintain a view of the overall structure of manuscript the same way human writers can. Systems have been created to effectively chain prompts. The problem is that at some point the system cannot be fed the entire manuscript, and therefore cannot really know if it is reintroducing previous examples or copying grammatical constructions from before. The writer is faced with either re-editing by hand, or resubmitting prompts until the computational assistant stumbles on something new.

The idea of multi-dimensional writing is to put a graph-based interface in front the human user as a way of guiding the AI tool. Indeed, it turns out that one can ask AI models such as GPT-3 to return concept triples: source, relation, target. These can be used to build conceptual graphs. Users can then click on concepts to further exploration along a certain dimension.

Figure 3 shows an example of this from an early prototype. GPT-3 is prompted to return structured text which is converted into a network graph. Writers can click on concepts, which triggers another prompt and expands the graph. In Figure 3, the writer first asked for concepts related to glaciers, and then clicked on a return concept, ice sheet, which then generated a set of other concepts, one of which, iceberg, is connected to both focal nodes.

4 CONCLUDING THOUGHTS

The argument here is that AI tools themselves can and should be combined with other AI tools, and in particular there is room to explore the combined use of text and image to generate storyboards. Also, AI tools should be managed through a graph-based interface
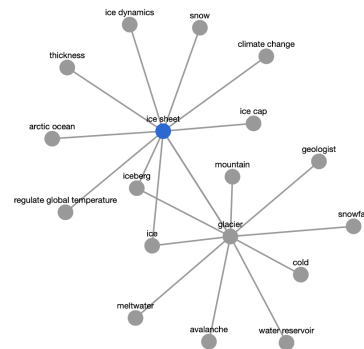

Figure 3: Output from a prototype concept explorer

that allows for exploration along many different dimensions before a narrative ordering is arrived at.

ACKNOWLEDGMENTS

REFERENCES

[1] Gérard Genette. 1983. *Narrative Discourse: An Essay in Method.* Cornell University Press.

[2] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. (Nov. 2022). arXiv:2211.05030 [cs.HC]

[3] Kalle Lyytinen, Jeffrey V Nickerson, and John L King. 2021. Metahuman systems = humans + machines that learn. *Journal of Information Technology* 36, 4 (2021), 427–445.

[4] Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2022. Co-Writing Screenplays and Theatre Scripts with Language Models: An Evaluation by Industry Professionals. (Sept. 2022). arXiv:2209.14958 [cs.HC]

[5] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Stan Jastrzebski, Keren Henderson, Jeffrey V. Nickerson, and Lydia B. Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.*