A Selective Summary of *Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence*

Nikhil Singh[§*] **Guillermo Bernal**[§*] **Daria Savchenko**[†*] **Elena L. Glassman**[†]
[§]MIT Media Lab
{nsingh1,gbernal}@mit.edu
[†]Harvard University
{daria_savchenko@g,glassman@seas}.harvard.edu

Abstract

While developing a story, novices and published writers alike have had to look outside themselves for inspiration. Language models have recently been able to generate text fluently, producing new stochastic narratives upon request. However, effectively integrating such capabilities with human cognitive faculties and creative processes remains challenging. We propose to investigate this integration with a multimodal writing support interface that offers writing suggestions textually, visually, and aurally. We conduct an extensive study that combines elicitation of prior expectations before writing, observation and semi-structured interviews during writing, and outcome evaluations after writing. Our results illustrate individual and situational variation in machine-in-the-loop writing approaches, suggestion acceptance, and ways the system is helpful. Centrally, we report how participants perform *integrative leaps*, by which they do cognitive work to integrate suggestions of varying semantic relevance into their developing stories. We interpret these findings, offering modeling and design recommendations for future creative writing support technologies.[1]

1 Introduction

Much remains unexplored about how emerging methods in AI, machine learning, and natural language processing might influence creative writing, in part due to the ambiguity and variability of human writing processes. These processes go beyond the linear projection from idea to a full text; research shows how planning narratives, translating ideas into visible textual material, and reviewing are all happening and interacting throughout the process rather than simple sequential stages (Nold, 1981; Flower and Hayes, 1981). However, this is a very familiar process for humans when communicating through writing; as every writer knows, having good ideas does not automatically produce a good text progression. The need for that "good idea" to be anchored and developed so that the reader can be invested takes a great deal of effort. In today's world, language generation models like GPT-2 (Radford et al., 2019), GPT-3 (Brown et al., 2020), and new ones coming down the line are typically silent on the inner processes of negotiation and decision that a human writer is working through. Additionally, contributions from these systems might take forms to influence writing other than text; writers are able to engage multiple perceptual channels through their work: they may activate multisensory imagination through evocative imagery, invoking auditory and olfactory phenomena, and other forms of sensory description.

We investigate how participants engage with a multimodal writing support system that bridges generated writing suggestions with multimedia retrieval to produce concept representations simultaneously in sight, sound, and language. We pair this interface with an extensive study that combines surveys, interaction, and semi-structured interviews during observed, think-aloud writing sessions. We examine and report in detail how participants receive, consider, and integrate suggestions from an intelligent tool into their writing. We explore prominent axes of individual and situational variation in these integrative behaviors, noting the different kinds of "leaps" participants make to understand suggestions and make the necessary compositional decisions to incorporate new information contained in them, ranging from copying and pasting to re-writing core aspects of their entire story.

In summary, our findings suggest that participants perform different kinds of *integrative leaps*, involving cognitive work to make suggestions useful to their writing. We interpret these and make commensurate design recommendations for future creative writing support tools.

[1]This work is a cross-submission and is published as Singh, Bernal, Savchenko, and Glassman, 2022.

Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022), pages 25 - 26

Acknowledgements

This material is based upon work supported by the NSF under Grant No. 2107391.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College composition and communication*, 32(4):365–387.

Ellen W Nold. 1981. Revising. *Writing: the nature, development, and teaching of written communication*, 2:67–79.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Trans. Comput.-Hum. Interact.* Just Accepted.