Research Statement

Thoughtful AI

Technological artifacts are not neutral; they shape how we act, how we see each other, how we think, what we value [15, 16, 18]. This is particularly true in technologies that mediate how people interact with each other, such as media and now generative AI, which injects computational judgments into how we compose our communications and how we perceive others' communications with us. By shaping how we think about each other, these technologies risk doing wrong to each other, such as dishonoring our email recipients and educators through computer-generated simulacra of empathetic and thoughtful prose. But my work embodies a hope that it doesn't have to be that way: shaped rightly, AI could help us *do justice towards one another* through thoughtful communication, perspective-taking, and reflective strategy.[1]

Dissertation Work: Dark side of AI Productivity

Narratives around intelligent systems (especially generative AI) typically center on *productivity*: how can we get more done in less time? My dissertation work challenged this narrative by inaugurating the study of how predictive text affects the content of written communication. Although predictive text was already ubiquitous on smartphones, academic research on it had only evaluated its effects on *productivity* (e.g., text entry speed and accuracy).

Through controlled experiments with a custom mobile predictive-text keyboard I developed, I discovered that these systems nudge writers to conform to system's suggestions in word choice [4], level of detail [5], and stance [3]. Running these experiments required technical work (a custom phrase-suggestion interface, custom n-gram and LSTM-based language modeling backends) and creative experiment design (repeatable writing tasks, sensitive quantitative measures of writing content).

In the years since I published this work, other researchers have replicated and extended these findings [1, 6, 11, 12], confirming broader concerns about how AI systems nudge people to align with system biases [12, 17] and homogenize creative expression [1, 2].

Supporting Thoughtful Writing through Alternative Interactions

Reflecting on my dissertation findings, I conjectured that the problematic nature of the predictive interactions I'd studied was mediated through an implicit value of *fast content production* instead of *cognitive engagement*. By presenting suggestions that humans could accept as their own with minimal thought or interaction, these prediction-based interfaces embody a value of quantity over thoughtfulness. Although disciplined people can still use

---

[1] See https://kenarnold.org/projects.html for papers and demos of the projects mentioned here.

these systems in thoughtful ways, the default is cognitive disengagement; human participation becomes passive. We must study how AI can help us think not *less* but *better*.

Thought Embers: suggestions that require writers to think

What if AI helped us come up with questions, not answers, as Licklider suggested back in 1960 [14]? Can presenting fragments of ideas, rather than complete and usable ideas, help writers retain agency over their work? In Summer 2024, I worked with a team of 4 students (3 undergraduate, 1 high school) to study how the type of information suggested by an AI writing assistant affects writers' cognitive engagement with the suggestion and how they appropriate that suggestion in their draft. We developed a Microsoft Word sidebar that offered next-sentence suggestions expressed in four different ways: in addition to predictive-text-style *examples* they could use verbatim (e.g., by copy-and-paste), we also allowed writers to request *questions* that the next sentence might answer, *vocabulary* they might use, and *rhetorical moves* (such as giving examples or considering counterarguments) that their next sentence might engage with. Students conceptualized these interventions, designed prompting and post-processing approaches for the NLP backend, and designed and implemented the sidebar frontend.

In a pilot study (N=8), writers found questions and rhetorical moves to be useful. Although writers chose to request examples more often, they often rejected the suggested text. Overall, they rated the Questions suggestion type as most compelling (desirable) in post-task surveys, followed by Examples. These preliminary results suggest that writers welcomed AI suggestions that could not be inserted verbatim into their documents but instead required further thought. Overall, by offering intentionally-incomplete suggestions like Questions or Rhetorical Moves, AI systems might become better cognitive partners for writers, enriching thinking rather than circumventing it. The students have presented this work at internal venues; we are designing a follow-up experiment to build on these findings for broader publication.

Interaction-Required Content Suggestions

The prior project manipulated the *type* of AI-generated content but used conventional interaction techniques; what if we presented typical types of content, but through interaction modalities that invite thinking? For example, consider a revision task, like a writer trying to adapt their work for a non-expert audience. An LLM can be used in at least 3 different ways for a task like this: directly generating the complete revised document (requiring no direct interaction), providing the revision as predictive text for the writer one word at a time (showing incremental contextual alternatives and requiring choices at each step), or showing the writer's original document annotated with alternative words the LLM might generate if generating that document (focusing human attention and requiring the writer to make any edits themselves).

In work in progress, my undergraduate research team is using prototypes we made of these interactions to study how the differences in interaction paradigm expressed in these three interfaces affect the writer's sense of authorship/ownership, how helpful the AI system is

for achieving their goal, and other factors like the writer's awareness of audience and sensitivity to AI bias. To support these interactions efficiently, we needed to implement custom inference code that interfaces with open-weights LLMs at a lower level than is typically used. Specifically, I used the forward pass of the model to compute single-step next-token predictive distributions rather than a batch-generated response, then manually managed the KV cache to generate lookahead tokens.

Supporting Reflection on Writing

Textfocals: AI Views for Revising

Could AI help writers *think about their text from the perspective of their readers* and revise to serve them better? This work builds on theories of writing process and reflection for revision [8, 9]. I worked with 7 undergraduate researchers in Summer 2023 to develop Textfocals, a Microsoft Word sidebar add-in that provides writers LLM-powered tools to reflect on their writing and set revision goals. Built-in tools provide several types of summaries and questions, but writers can also customize the prompts.

A formative user study with Textfocals [13] yielded promising evidence that this approach could help people develop underdeveloped ideas, cater to the rhetorical audience, and clarify their writing. However, the students running the study and analyzing the findings identified interaction design challenges related to document navigation and scoping, prompt engineering, and context management, leading to substantial reworks of the interaction design. Our work shows the breadth of the design space of writing support interfaces powered by generative AI that maintain authorship integrity.

Reflective Dialogue

An advanced undergraduate student is currently working with me and another co-advisor to refine the kind of reflection-promoting interactions that we started exploring in Textfocals. The current work focuses on helping the writer actively reflect on their rhetorical situation (who the writer is addressing, why, and how) and their writing. The AI-powered system can then help the writer set revision goals grounded in their rhetorical situation and consider multiple approaches to achieving those goals. The student is specifically interested in how using *speech* as a modality for reflection might affect the quality of a reflective dialogue.

Although I try to give students specific and reasonably scoped tasks, I also try to involve them at least lightly in the whole research process (from design to writing to reviewing) and let them experience the adventure of uncertain research and face it with courage. The student who is doing this project is an example of the success of that strategy: the vision of the project motivated him enough to start a follow-up project on his own initiative.

Future Directions

My team is currently preparing our Microsoft Word add-in, now called Thoughtful, for a longitudinal study of its real-world use. Other near-term plans include a follow-up study on Thought Embers, a study of how interaction-required suggestions affect ownership in revision tasks, and a study on modality of interaction with AI in reflection for text revision. All these aim for publication in archival venues such as ACM CHI, UIST, or IUI. I also plan to pursue additional funding for when my NSF CRII grant concludes next year.

Visions for Cross-Disciplinary Collaboration

I am particularly interested in three promising areas for collaboration:

1. Working with **writing and rhetoric** scholars to study how AI-assisted tools might explicitly scaffold writing processes that instructors want to emphasize. This collaboration would build on our work on Textfocals while bringing in deeper theoretical frameworks.
2. Working with **education** scholars to develop and study AI systems that help instructors make student-centered revisions to their teaching. For example, I have prototyped LMS extensions that, like my Textfocals work, encourage and empower instructors to **take the perspective of students**: predicting what clarifying questions students will have about assignments, quiz questions, logistics, or policies. I have also done some formative work with a language education scholar on **better understanding our students** by building fine-grained models of learner knowledge to adaptively differentiate instruction within classes.
3. Working with **organizational behavior and communication scholars** to explore how thoughtful AI (reflection-promoting interfaces) might improve group decision-making and communication. Examples could include revision tools for organizational communication materials (perhaps as add-ins for other software like Outlook and PowerPoint) to help think through stakeholder perspectives, and AI assistance for leaders to identify potentially excluded voices and facilitate more inclusive discussions.

Conclusion

We can—and must—build AI to help people think better about other people. To do so, we need cross-disciplinary theoretical grounding, intentional co-design of AI and HCI, and empirical study both in the lab and through deployed systems. My research agenda welcomes students and experts in many disciplines to contribute to building a more caring world through thoughtful AI.

References

[1]   Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2024. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. https://doi.org/10.48550/arXiv.2409.11360

[2]   Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization Effects of Large Language Models on Human Creative Ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (*C&C '24*), June 23, 2024. Association for Computing Machinery, New York, NY, USA, 413–425. https://doi.org/10.1145/3635636.3656204

[3]   Kenneth C Arnold, Krysta Chauncey, and Krzysztof Z. Gajos. 2018. Sentiment Bias in Predictive Text Recommendations Results in Biased Writing. In *Graphics Interface 2018*, 2018. Toronto, Ontario, Canada, 8–11. Retrieved from http://graphicsinterface.org/wp-content/uploads/gi2018-7.pdf

[4]   Kenneth C. Arnold, Krysta Chauncey, and Krzysztof Z. Gajos. 2020. Predictive text encourages predictable writing. In *Proceedings of the 25th international conference on intelligent user interfaces* (*IUI '20*), March 17, 2020. Association for Computing Machinery, Cagliari Italy, 128–138. https://doi.org/10.1145/3377325.3377523

[5]   Kenneth C Arnold, Krzysztof Z. Gajos, and Adam T. Kalai. 2016. On Suggesting Phrases vs. Predicting Words for Mobile Text Composition. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*, 2016. 603–608. https://doi.org/10.1145/2984511.2984584

[6]   Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (*IUI '23*), March 27, 2023. Association for Computing Machinery, New York, NY, USA, 436–452. https://doi.org/10.1145/3581641.3584060

[7]   Arghavan Moradi Dakhel, Vahid Majdinasab, Amin Nikanjam, Foutse Khomh, Michel C. Desmarais, Zhen Ming, and Jiang. 2023. GitHub Copilot AI pair programmer: Asset or Liability? https://doi.org/10.48550/arXiv.2206.15331

[8]   Linda Flower and John R. Hayes. 1981. A Cognitive Process Theory of Writing. *Coll. Compos. Commun.* 32, 4 (1981), 365–387. https://doi.org/10.2307/356600

[9]   Linda Flower, John R. Hayes, Linda Carey, Karen Schriver, and James Stratman. 1986. Detection, Diagnosis, and the Strategies of Revision. *Coll. Compos. Commun.* 37, 1 (1986), 16–55. https://doi.org/10.2307/357381

[10]   Krzysztof Z. Gajos and Lena Mamykina. 2022. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (*IUI '22*), March 22, 2022. Association for Computing Machinery, New York, NY, USA, 794–806. https://doi.org/10.1145/3490099.3511138

[11]   Jess Hohenstein, Dominic DiFranzo, Rene F. Kizilcec, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeff Hancock, and Malte Jung. 2021. Artificial intelligence in communication impacts language and social relationships. https://doi.org/10.48550/arXiv.2102.05756

[12]   Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*

(*CHI '23*), April 19, 2023. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3544548.3581196

[13] Jiho Kim, Ray C. Flanagan, Noelle E. Haviland, ZeAi Sun, Souad N. Yakubu, Edom A. Maru, and Kenneth C. Arnold. 2024. Towards Full Authorship with AI: Supporting Revision with AI-Generated Views. In *Joint Proceedings of the ACM IUI 2024 Workshops*, March 18, 2024. CEUR-WS.org, Greenville, South Carolina, USA. Retrieved from https://ceur-ws.org/Vol-3660/paper17.pdf

[14] J. C. R. Licklider. 1960. Man-Computer Symbiosis. *IRE Trans. Hum. Factors Electron.* HFE-1, 1 (March 1960), 4–11. https://doi.org/10.1109/THFE2.1960.4503259

[15] Neil Postman. 1998. Five Things We Need to Know About Technological Change.

[16] L. M. Sacasas. 2021. The Questions Concerning Technology. *The Convivial Society*. Retrieved August 26, 2024 from https://theconvivialsociety.substack.com/p/the-questions-concerning-technology

[17] Tobias Werner, Ivan Soraperra, Emilio Calvano, David C. Parkes, and Iyad Rahwan. 2024. Experimental Evidence That Conversational Artificial Intelligence Can Steer Consumer Behavior Without Detection. https://doi.org/10.48550/arXiv.2409.12143

[18] Langdon Winner. 2003. Do artifacts have politics? *Technol. Future* 109, 1 (2003), 148–164. https://doi.org/10.2307/20024652