Designing for Agency in LLM-Infused Writing Support Tools for Science Journalism

Sachita Nishal*
Northwestern University
Evanston, IL, USA
nishal@u.northwestern.edu

Mina Lee
University of Chicago
Chicago, IL, USA
mnlee@uchicago.edu

Nicholas Diakopoulos
Northwestern University
Evanston, IL, USA
nad@northwestern.edu

Jennifer Wortman Vaughan
Microsoft Research
New York, NY, USA
jenn@microsoft.com

Abstract

Large language models (LLMs) are being integrated into tools to support the creative process of writing, from brainstorming to drafting and editing text. However, studies show that such tools risk undermining human *agency*—the ability to make and execute independent decisions and ensure that outcomes are aligned with personal or professional values. Journalism is a field in which maintaining human agency is particularly crucial. Writing support tools for journalists must be designed to support journalists in gauging interest in a story, independently verifying facts, and drawing connections, all while upholding journalistic values like accountability and independence. In this study, we investigate how to design LLM-infused writing support tools that support rather than impinge on journalists' agency. We engage twenty science journalists in the exploration of four design concepts for hypothetical pitch writing tools that vary in task initiation, scope of LLM usage, and material control over the LLM. We find that participants see the act of pitch writing as a part of their thinking process and are unwilling to delegate tasks that would jeopardize that process. We find that preserving voice and the freedom to explore ideas are central to journalists' sense of agency. We discuss implications for designing LLM-infused tools that support agency and the potential longer-term impact of such tools.

1 Introduction

Large language models (LLMs) are increasingly woven into writing support tools to support the process of writing—all the way from sensemaking and initial brainstorming, to drafting and editing text. These tools have the potential to stimulate new ideas, help users overcome writing blocks, and provide critical feedback [16, 17, 20, 34, 35]. But despite these potential benefits, recent studies have highlighted that co-creating with LLMs can restrict independent problem-solving, cause creative fixation on limited ideas, homogenize writing, reduce writers' sense of ownership, and subtly influence judgment [2, 11, 19, 23, 25, 33]. These effects can undermine human *agency*—defined, for the purposes of this paper, as the ability to make and execute independent decisions and ensure that outcomes remain aligned with personal or professional values [4].

Journalism is a domain in which maintaining human agency is particularly crucial. Journalists are responsible for gauging societal, editorial, or personal interest in a story, independently verifying facts, drawing meaningful connections, and maintaining intellectual independence from external influences, including AI systems themselves [22]. It is important that AI-infused tools for journalists are designed to support these tasks and to support journalistic values more broadly [21]—for instance, accountability to audiences and objectivity (freedom from bias) [10]. Prior work shows that this may not always be a given: journalists view automated writing as lacking "human characteristics" such as a unique voice or a critical approach [8] and as potentially reducing the transparency, objectivity, and creativity of their work [31, 32].

Our work examines how LLM-infused writing support tools can be designed to support, rather than supplant, journalistic agency. We explore this within the specific context of science journalism. Prior writing support tools have been designed to help science journalists brainstorm "newsworthy" story ideas or angles [27–29]. In the context of science writing more generally, interactive interfaces have been designed to generate sparks for writing [15], suggest metaphors for simplifying ideas [20], and create interesting hooks for social media posts [26]. Our work builds on and contributes to this literature by exploring a broader design space of possibilities for LLM-infused writing tools [24], with a focus on supporting different dimensions of agency (e.g., decision making, execution, and value alignment). Specifically, we ask the following questions:

- **RQ1** How do journalists perceive their individual agency when using LLM-infused writing support tools?
- **RQ2** How does journalists' use of these tools come into tension with journalists' values?
- **RQ3** What forms of control do journalists perceive as useful for maintaining agency when using these tools?

We developed four design concepts aimed at supporting journalists in brainstorming, drafting, and/or reviewing their writing. Each was presented as a hypothetical tool, with a mock-up user interface that illustrated potential uses. Such technologically viable (but unimplemented) concepts are aimed at surfacing user needs, goals, or values with respect to design decisions, through rapid, open-ended exploration of the design space [13, 14, 36]. Some of our design choices were also made to provoke discussion by actively taking away journalists' agency, rather than by supporting it. Through semi-structured interviews with 20 science journalists exploring these design concepts, we examined how different design decisions did or did not interface with different facets of journalists' agency.

*This research was conducted while S. Nishal and M. Lee were at Microsoft Research.

Our design concepts and interviews center on the specific task of creating *pitches*, which are short texts written by journalists to propose a story idea, surface preliminary facts, design and communicate a reporting plan, showcase their domain expertise and sources, and highlight why a story needs to be told, all so that an editor might commission them to write the story [3]. Creating a pitch involves an exercise of journalists' creativity in framing a unique angle or perspective, distinguishing the idea from prior stories, presenting a unique writing style and voice—all within certain structural conventions [12]. As Reich [30] puts it, "the mix of formulaic constraints and creative freedom of journalists places them in a situation comparable to that of jazz musicians [...]: the song may be predetermined, but players are free to improvise within a given framework."

Our preliminary findings surface how journalists view the impact of LLM-infused writing support tools on their agency, and where design opportunities exist to support these professionals. Participants revealed a complex view of the pitch writing process: as a fundamental thinking mechanism, a skill to be cultivated, and a personal expression of identity. This led to a nuanced view of LLM assistance, with journalists valuing AI's potential to support writing-adjacent processes like sensemaking of complex documents, information gathering across sources, and receiving suggestions for strengthening their writing. The fact that the pitch entails unpaid labor (prior to being commissioned) led to a need for efficiency through tool use, but not at the cost of creative control over the idea and its presentation (e.g., rejection of LLM-generated first drafts).

The findings highlight a delicate balance between leveraging LLM capabilities and maintaining the unique human elements of journalistic creativity. Our focus on pitch writing offers a contextually rich exploration of agency in the context of LLM use. We believe that the tensions we uncover are likely to resonate across other writing tasks that similarly balance structural constraints with individual expression.

2 Methods

Our goal in this study was to learn about the design opportunities to support journalists' agency, in the context of particular technologies (generative AI, and specifically LLMs) and activities (brainstorming and writing in journalism). We decided to explore multiple types of human-AI configurations to get broader and comparative insights on agency in this context.

Recent work has used approaches like design workbooks and speed dating to attend to concerns similar to ours [5, 9, 18]. Similar to this prior work, we drew on Gaver et al.'s [14] notion of *design spaces*, which represent the landscape of viable possibilities for design that merit exploration and investigation. Design workbooks offer a method to help designers and users explore such design spaces via varied combinations of features and ideas, encouraging speculation on the differences and similarities between different *design concepts* that result from these combinations [14]. This allows designers to move beyond immediate technical constraints and explore a broader range of potential problems and opportunities within a specific domain. Speed dating focuses on generating multiple design concepts for users to rapidly experience and critique, allowing designers to surface unmet needs, uncover latent social barriers, and refine ideas based on real-world feedback that might be missed by traditional methods [36]. Our work is inspired by and draws from these approaches.

We generated design concepts through a month-long period of brainstorming, sketching, and prototyping within the research team, following which we engaged in pilot testing with colleagues and journalists (n=6) to iterate and refine them. Each design concept in the final set represents a different approach to supporting journalistic agency in the context of writing pitches with LLMs: from giving journalists full control over when and how to engage LLM assistance (Pitch Assist), to providing structured opportunities for maintaining editorial control (Pitch Refine), to automatically suggesting improvements while preserving decision-making authority (Pitch Critic), to fully automated assistance that maintains minimal user oversight (Pitch Suggest). Appendix A illustrates these concepts.

Concretely, design concepts varied along three prominent dimensions: task initiation (system vs. user initiation of task), LLM Task Support (whether the LLM supports ideation, drafting, or revision during writing), and material control (degree to which users can manipulate where and how LLMs generate outputs). Table 1 provides an overview.

Our interviewees represented a wide pool of experiences, topical interests, pitching habits, and target audiences. Appendix B presents this information. Interviews lasted 60–80 minutes and consisted of three segments. During the **onboarding** segment, we gathered background information from participants through open-ended questions about their practice, pitching challenges, and current AI use. The substantive portion of our interview centered on exploring the **design concepts** through scenarios, which allowed us to probe how journalists perceived agency, tool utility, workflow changes, and so on. We concluded with a **debrief** during which participants ranked design concepts, explained their preferences, and discussed potential disclosure of AI tool usage.

The first author assigned codes to the interview transcripts through a process of reflexive thematic analysis, while engaging in periodic discussions with other authors on the team [6]. We relied on latent, deductive thematic analysis to generate these codes [7]. As we are currently in the process of generating higher-level themes from this data, this is a work-in-progress.

3 Preliminary Findings

Our preliminary findings reveal how journalists might conceptualize and wish to maintain or cede agency when using LLM-infused writing support tools. We describe a few prominent themes generated from our analysis that showcase journalists' practices and writing approaches, and the ways this shapes their responses to different design concepts.

First, participants emphasized how, in pitching, **the act of writing serves the process of thinking**, i.e., of making sense of a story, its relevance, its newsworthiness. One participant explained:

> The process of iterating and thinking things through and chipping away at my ideas to find the right idea [...] is important and if I were to shorten it with generative AI, then I would have less of a sense of why I'm doing what I'm doing and what the end result is. [P17]

Design Concept	Task Initiation	LLM Assistance Scope	Material Control
Pitch Assist	User-initiated via chat interface	Ideation, drafting, revision	High
Pitch Refine	User-initiated via scaffolded interface	Ideation, revision	Moderate
Pitch Critic	System-initiated via automatic suggestions	Drafting, revision	Moderate
Pitch Suggest	System-initiated via automatic drafting	Ideation, drafting	Minimal

Table 1: Our design concepts vary in how they support agency through different patterns of task initiation, scope of LLM assistance, and material control over process and outcome.

Another participant likened writing the pitch to "going to the gym," which, while difficult, was a "necessary evil" to build their skills: "You can't have someone do it for you. You still have to do it yourself." [P12]

Consequently, some participants were receptive to LLM suggestions that helped them reflect on their decisions, even if the suggestions themselves were not deemed useful. For instance, toward Pitch Critic's feedback on their writing, one participant noted: "I think it's really good, actually, for making you think about the decisions you're making. Like this bias tool, why are you only thinking about contacting a pharma representative? [...] One concern people have about AI tools is that they reduce the thinking. But here, it actually could make you think more about that." [P19]

To preserve this process of thinking through writing, participants also drew **distinctions between tasks they were willing to delegate to AI and those they saw as central to their cognitive work**. A common trend we observed was that participants welcomed LLM support for information gathering, fact-checking, and making connections across disparate sources. These tasks were seen as peripheral, rather than as central to their brainstorming and writing process.

Participants were especially receptive to design concepts in which the tool supported revisions through offering edits and feedback on their writing (e.g., with Pitch Critic). This was seen as a way to reduce their unpaid labor, by helping them strengthen pitches before sending them out, thereby increasing their chances of receiving a commission. Participants also expressed an openness toward features that could get them out of a creative block. The quality of LLM outputs in this case was besides the point. On their prior experience with using LLMs to come up with a story title, one journalist noted: "It didn't necessarily come up with anything that I used, but it sort of breaks you out of your current thinking patterns." [P15] Another noted that support for generating small parts of the pitch (e.g., angles, headlines) could help them gain momentum in the writing process, by offering text to modify and build on.

On the whole, participants were significantly less receptive to concepts where the tool generated the core idea (e.g., parts of Pitch Assist) or the whole pitch (e.g., in Pitch Suggest). While this might be feasible, it was not always deemed to align with the artistic value of journalistic writing:

> Journalism is like a very practical art form [...] Some people do it incredibly well in a way that I think does make it qualify as an art form, but at the same time, you can get a lot of the functional value from something that is much less art [...] Letting AI in bends the trajectory more towards the latter, more towards making it purely functional, and not something to be celebrated. [P8]

The wide range of tasks participants were open to receiving LLM support for—with a small portion even open to the LLM creating initial drafts that they might then edit—created considerable variation in rankings of the different design concepts during debrief.

A key dimension of writing beyond individual tasks is the writers' style or voice. Participants viewed their **writing style and voice not just as a matter of preference, but as a core part of their professional identity and relationship with readers**. This led to wider concerns about how LLMs might draw on or modify their writing voice. One journalist explained: "Voice is such a hard thing to pin down as a writer [...] it's almost like this sort of imaginary thing that exists that I don't fully understand why mine is the way it is, but it seems to be working. So, again, I would be nervous to sort of even refine it [with the Pitch Refine tool] in a way." [P15] Another linked it to a more personal feeling, "I would prefer to [modify writing style] on my own because I do think [it entails] choosing the words and actually like crafting. Crafting is something that I still feel really strongly about doing myself." [P6]

This concern about voice extended beyond personal style to questions of homogeneity across the field of writers. When noting their disinterest in mechanisms to manipulate voice and tone with LLM support, one participant stated:

> I don't want a world where everyone sounds the same or where people who don't have the same first language just kind of smooth over all of those differences. I think those differences are really interesting. [P19]

Related to homogeneity in writing, participants also worried about how **AI tools might constrain their creative thinking processes too quickly**, especially during initial brainstorming. Several noted that seeing AI-generated suggestions could create a fixation, making it harder to think beyond those initial ideas: "Once you've got that idea, once you've got that angle in your head, you don't see it in any other way. I think, especially, there's something about when it comes from a machine, there's a sense of objectivity that that's the correct answer." [P17]

Latent here is also the notion that journalists exercise agency in determining when an idea might be ready to turn into a pitch or an article, and using LLMs might generally constrain that decision, or make some ideas seem more apt than the journalist might otherwise have judged them to be.

On a more philosophical level, participants grappled with questions about **maintaining their cognitive skills in a future with more AI augmentation**. Improving LLM capabilities create new possibilities for offloading cognitive tasks, but this raises questions of what is appropriate to offload or not. One journalist articulated this tension through an extended metaphor about farming:

> It does seem like [the Pitch Assist tool] could be easy to get dependent on, and I wonder over time if I would get lazy about writing my own sentences [...] Like I don't know how to grow my own food and because other people do it for me, and it would be a big problem if those other people suddenly stopped doing it for me, but does that mean that I should spend my time being a farmer? [P8]

Others noted that automating lower-stakes writing (the pitch being an example, but also short-form news formats) could reduce avenues for skill development for inexperienced journalists.

4 Discussion and Conclusion

Our preliminary analysis suggests several implications for designing LLM-infused writing support tools that support journalists' agency. First, the act of writing serves important cognitive functions beyond text production, even with a pitch. Journalists develop their understanding of the science, the field, and their audiences, and engage in reflection through the writing process. Writing the pitch also involves assessing whether an angle is appropriate for a story, or for an outlet. This suggests AI writing tools need to support these meta-cognitive processes of sensemaking, reflection, and evaluation, rather than focusing solely on generating the most appropriate text output. Supporting these processes would be a way to more broadly support the journalists' exercise of agency over how they develop their stories and their writing.

Second, journalists generally welcomed LLM support for information gathering and editing, while endeavoring to preserve agency over core writing and idea generation. Still, there was some variation in tasks that they wanted to delegate. Beyond AI capability, questions also remain about what tasks writers *should* maintain versus delegate as LLM capabilities change, as well as when task delegation might be feasible and efficient and yet undercut the creative value of the writing for journalists who care about their craft. This highlights an inherent tension between what is *feasible* versus what is *appropriate* to automate, and suggest that design processes must go beyond managerial efficiency needs (a common driver for automation) to consider workers' perspectives and values [1]. Considering temporal perspectives on agency may also be a fruitful direction here. Could offloading tasks and thus increasing dependence on LLM-based tools for ideation, drafting, and editing limit the agency of journalists in the future, e.g., by atrophying core writing skills or reducing opportunities to develop professional judgment through practice?

Third, preserving voice and the freedom to explore ideas emerged as central to journalists' sense of agency, extending beyond stylistic concerns to questions of homogeneity across their field. While journalists valued AI's potential to reduce cognitive load and time spent on mechanical tasks, its use can potentially diminish the scope of creative decisions they take, both individually and at the field level. However, the distinction between mechanical and creative writing tasks itself is increasingly blurred, as seemingly routine decisions like word choice or metaphor selection shape a writer's voice. There remains an open question then: At what point is the LLM that is making these decisions too heavily impacting the journalists' own sense of style or voice?

As we continue our analysis, we will examine how these themes (and others) manifest across different aspects of the brainstorming and writing process in our chosen domain. Through this ongoing work, we aim to better understand how to maintain human agency in LLM-infused writing support tools, enabling journalists to preserve both their professional values and creative satisfaction while serving the public interest.

Acknowledgments

We are grateful to our study participants, without whom this research would not be possible. Thanks to Mandi Cai, Luke Guerdan, Evey Huang, Ben Laufer, Lesia Semenova, and Tejal Wakchoure for participating in pilots of our study. We are also grateful to Shivani Kapania, Ernest Kung, Tzu-Sheng Kuo, Rock Pang, Gonzalo Ramos, and the MSR FATE group for their thoughtful comments and feedback on our work.

References

[1] Mike Ananny and Jake Karr. 2025. How Media Unions Stabilize Technological Hype: Tracing Organized Journalism's Discursive Constructions of Generative Artificial Intelligence. 0, 0 (2025), 1–21. doi:10.1080/21670811.2025.2454516

[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* (New York, NY, USA, 2024-06-23) *(C&C '24)*. Association for Computing Machinery, 413–425. doi:10.1145/3635636.3656204

[3] Martin W. Angler. 2017. Pitching. In *Science Journalism*. Routledge.

[4] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How Does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2023-04-19) *(CHI '23)*. Association for Computing Machinery, 1–18. doi:10.1145/3544548.3580651

[5] Oloff C. Biermann, Ning F. Ma, and Dongwook Yoon. 2022. From Tool to Companion: Storywriters Want AI Writers to Respect Their Personal Values and Writing Strategies. In *Designing Interactive Systems Conference*. ACM, Virtual Event Australia, 1209–1227. doi:10.1145/3532106.3533506

[6] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa

[7] Virginia Braun and Victoria Clarke. 2024. Supporting Best Practice in Reflexive Thematic Analysis Reporting in Palliative Medicine: A Review of Published Research and Introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). *Palliative Medicine* 38, 6 (June 2024), 608–616. doi:10.1177/02692163241234800

[8] Matt Carlson. 2015. The Robotic Reporter. 3, 3 (2015), 416–431. doi:10.1080/21670811.2014.976412

[9] Janet X. Chen, Francesco Vitale, and Joanna McGrenere. 2021. What Happens After Death? Using a Design Workbook to Understand User Expectations for Preparing Their Data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–13. doi:10.1145/3411764.3445359

[10] Mark Deuze. 2005. What Is Journalism?: Professional Identity and Ideology of Journalists Reconsidered. 6, 4 (2005), 442–464. doi:10.1177/1464884905056815

[11] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI Ghostwriter Effect: When Users Do Not Perceive Ownership of AI-Generated Text but Self-Declare as Authors. 31, 2 (2024), 25:1–25:40. doi:10.1145/3637875

[12] Janet Fulton and Phillip McIntyre. 2013. Journalists on Journalism: Print Journalists' Discussion of Their Creative Process. 7, 1 (2013), 17–32. doi:10.1080/17512786.2012.657901

[13] Bill Gaver and Heather Martin. 2000. Alternatives: Exploring Information Appliances through Conceptual Design Proposals. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems. ACM, The Hague The Netherlands, 209–216. doi:10.1145/332040.332433

[14] William Gaver. 2011. Making Spaces: How Design Workbooks Work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1551–1560. doi:10.1145/1978942.1979169

[15] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing Using Language Models. In *Designing Interactive Systems Conference*. ACM, Virtual Event Australia, 1002–1019. doi:10.1145/3532106.3533533

[16] Alicia Guo, Shreya Sathyanarayanan, Leijie Wang, Jeffrey Heer, and Amy Zhang. 2024. *From Pen to Prompt: How Creative Writers Integrate AI into Their Writing Practice*. doi:10.48550/arXiv.2411.03137 arXiv:2411.03137 [cs]

[17] Alicia Guo, Leijie Wang, Jeffrey Heer, and Amy Zhang. 2024. Preserving Writer Values in AI Writing Assistance Tools. In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants* (New York, NY, USA, 2024-10-15) *(In2Writing '24)*. Association for Computing Machinery, 58–61. doi:10.1145/3690712.3690727

[18] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2019. *Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms*. Technical Report.

[19] 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* (2023-04-19). 1–15. doi:10.1145/3544548.3581196 arXiv:2302.00560 [cs]

[20] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh PA USA, 2023-07-10). ACM, 115–135. doi:10.1145/3563657.3595996

[21] Tomoko Komatsu, Marisela Gutierrez Lopez, Stephann Makri, Colin Porlezza, Glenda Cooper, Andrew MacFarlane, and Sondess Missaoui. 2020. AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, Tallinn Estonia, 1–13. doi:10.1145/3419249.3420105

[22] Bill Kovach and Tom Rosenstiel. 2021. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect* (revised and updated 4th edition ed.). Crown, New York.

[23] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (2025-04). ACM. https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/

[24] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024-05-11) *(CHI '24)*. Association for Computing Machinery, 1–35. doi:10.1145/3613904.3642697

[25] Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024-05-11) *(CHI '24)*. Association for Computing Machinery, 1–25. doi:10.1145/3613904.3642625

[26] Tao Long, Dorothy Zhang, Grace Li, Batool Taraif, Samia Menon, Kynnedy Simone Smith, Sitong Wang, Katy Ilonka Gero, and Lydia B. Chilton. 2023. Tweetorial Hooks: Generative AI Tools to Motivate Science on Social Media. arXiv:2305.12265 [cs]

[27] Neil Maiden, Konstantinos Zachos, Suzanne Franks, Rebecca Wells, and Samantha Stallard. 2020. Designing Digital Content to Support Science Journalism. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, Tallinn Estonia, 1–13. doi:10.1145/3419249.3420124

[28] Sachita Nishal, Jasmine Sinchai, and Nicholas Diakopoulos. 2024. Understanding Practices around Computational News Discovery Tools in the Domain of Science Journalism. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 142:1–142:36. doi:10.1145/3637419

[29] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, 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 (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3544548.3580907

[30] Zvi Reich. 2010. Constrained Authors: Bylines and Authorship in News Reporting. *Journalism* 11, 6 (Dec. 2010), 707–725. doi:10.1177/1464884910379708

[31] N. Thurman, K. Dörr, and J. Kunert. 2017. When Reporters Get Hands-on with Robo-Writing: Professionals Consider Automated Journalism's Capabilities and Consequences. 5, 10 (2017), 1240–1259. doi:10.1080/21670811.2017.1289819

[32] Arjen van Dalen. 2024. Revisiting the Algorithms Behind the Headlines. How Journalists Respond to Professional Competition of Generative AI. 0, 0 (2024), 1–18. doi:10.1080/17512786.2024.2389209

[33] Samangi Wadinambiarachchi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. The Effects of Generative AI on Design Fixation and Divergent Thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024-05-11) *(CHI '24)*. Association for Computing Machinery, 1–18. doi:10.1145/3613904.3642919

[34] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. 2024. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. 8 (2024), 84:1–84:26. Issue CSCW1. doi:10.1145/3637361

[35] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2022-03-22) *(IUI '22)*. Association for Computing Machinery, 841–852. doi:10.1145/3490099.3511105

[36] John Zimmerman and Jodi Forlizzi. 2017. Speed Dating: Providing a Menu of Possible Futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (March 2017), 30–50. doi:10.1016/j.sheji.2017.08.003

A Design Concepts Used in the Semi-structured Interviews

Four different design concepts were explored during the semi-structured interviews. The order in which they were presented was randomized across all participants. We describe and present each design concept below along with the scenario we used to illustrate it.

Pitch Assist. This concept was designed as a chat-based assistant for idea generation, drafting, and revision. In a scenario around pitching a short, environmental science story based on a press release, participants were shown how they could use an LLM to brainstorm angles, headlines, reader concerns, and visualizations, as well as how they could manipulate the style and length of the LLM-generated writing. Using Pitch Assist, the journalist has high material control over the LLM, given their ability to prompt, control decoding parameters, decide when to invoke the LLM, and decide the scope of task support. See Figure 1.

Pitch Refine. This concept afforded more scaffolded interaction with an LLM to support idea generation and revision. The tool could be used to make specific, scoped changed to the news angle and its writing style, suggest interviewees or data sources to shape the story, and adjust the style of the journalists' writing through LLM suggestions on tone and technical specificity. The scenario was based on a short tech-focused story based on a research paper. Using Pitch Refine, the journalist has moderate material control, deciding when to invoke the LLM, and being offered a range of preset prompts for different aspects of the writing (lede, plan, voice). See Figure 2.

Pitch Critic. This concept supported revision by offering both in-text and overall suggestions to the writer in order to better align the pitch to specific journalistic criteria (accuracy, clarity, bias checks, plus custom criteria for templated prompts) or to what editors might want. The scenario was focused on exploring each of these types of suggestions when writing a health story about a new research paper. Using Pitch Critic, the journalist has moderate

material control, being offered a range of preset prompts about different criteria for their writing (accuracy, clarity, bias) along with the opportunity to define custom criteria. However, once criteria have been selected, all individual interactions are initiated by the system as the journalist writes. See Figure 3.

Pitch Suggest. This concept supported idea generation and drafting based on the journalists' writing history, thematic interests, and pitching style. The scenario showed how a journalist might share this information, allowing the tool to surface potential news outlets that might be aligned to their interests and style, allowing the user to pick one and receive an automatic draft to edit themselves. The scenario focused on a longer, mental health-focused piece. Using Pitch Suggest, the journalist has very minimal material control over the LLM beyond selecting what personal data they would like to share and the venue they might want to write for. See Figure 4.

B Participant Metadata

Table 2 contains a brief description of each of our participants, including how many years of experience they have, the beat (i.e., topic area or subject) they write about, and the type of news publications they typically write for.

Most participants had experience with written, rather than broadcast, journalism. A few had experience as staff journalists and editors in newsrooms. Participants worked for a wide range of publications: general interest, science-focused (e.g., WIRED, Scientific American, Eos), and trade publications (e.g., Science, Nature). We consider "local news" to be its own category when mentioned by participants. Years of experience are approximate, collected at the time of the interview, and are not accurate to the month.

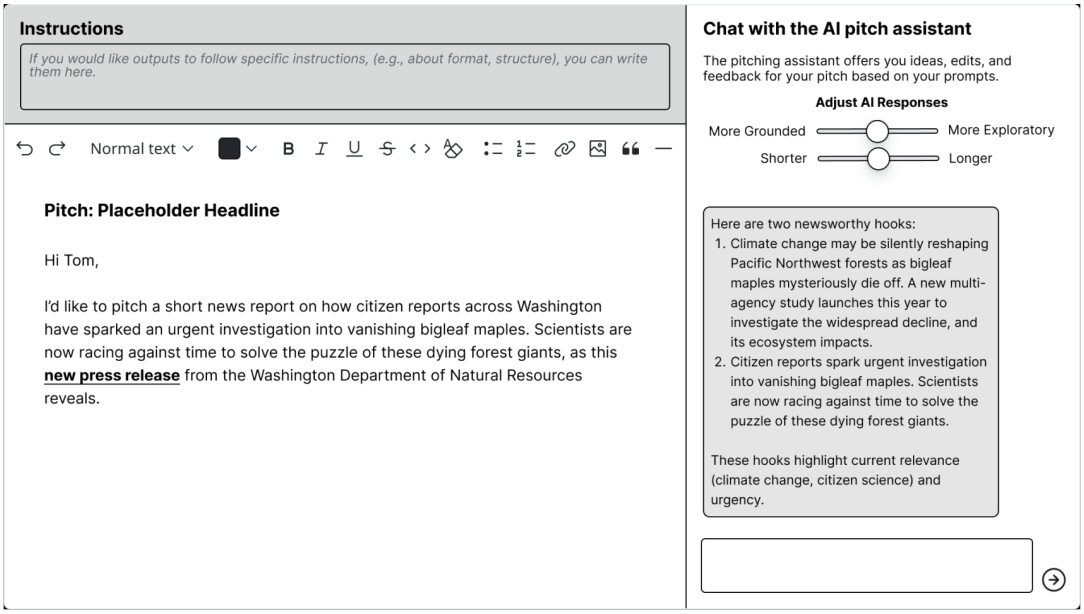

Figure 1: Pitch Assist, a chat-based assistant for idea generation, drafting, and revision.

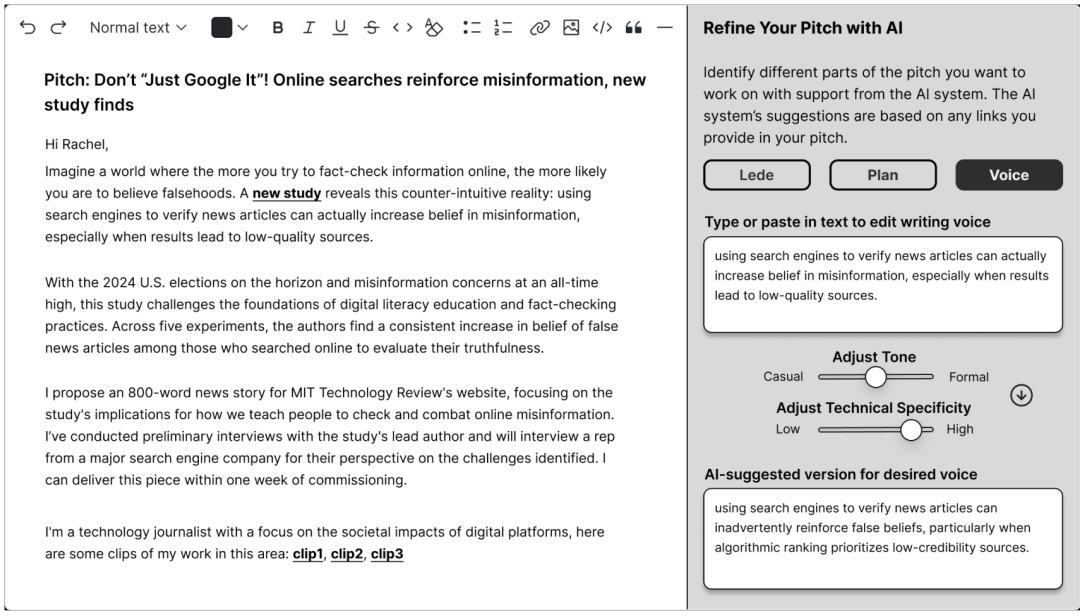

Figure 2: Pitch Refine, a scaffolded interaction for idea generation and revision.

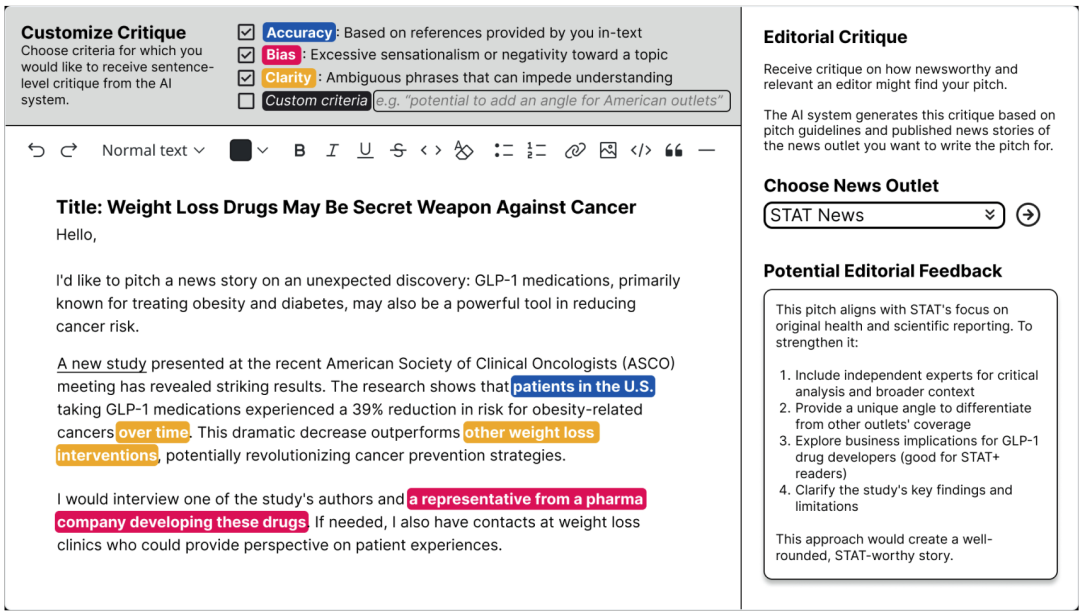

Figure 3: Pitch Critic, a system offering in-text and overall suggestions based on journalistic criteria.

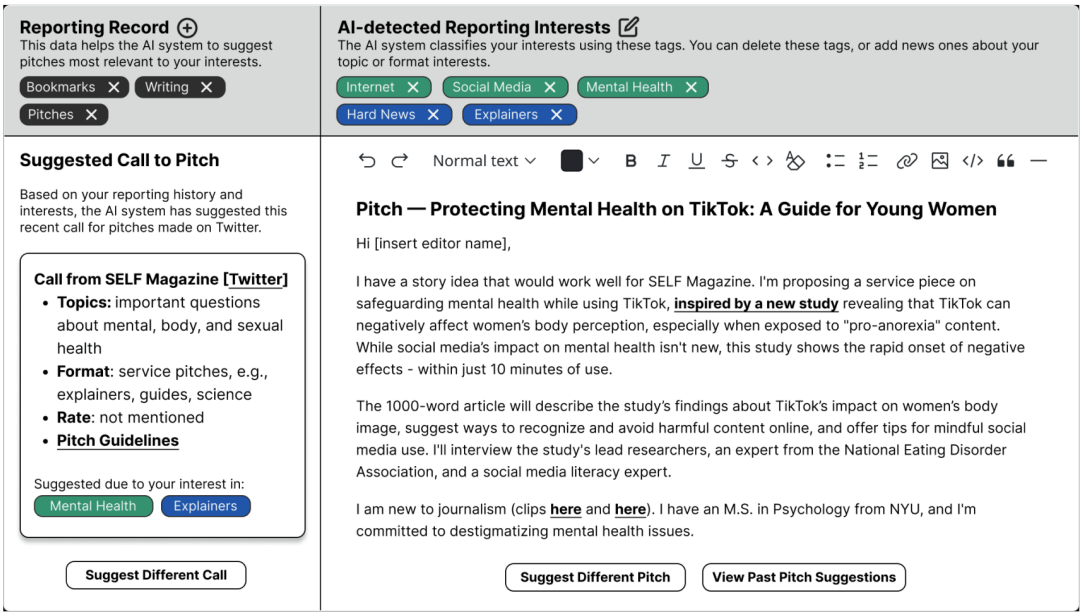

Figure 4: Pitch Suggest, a system supporting idea generation and drafting based on journalists' reading, writing, and pitching history.

ID	Experience	Beat	Story Type	Publication Type
P1	3 years, freelancing	Biology, neuroscience	News, study stories, trend stories	Science-focused, trade publications
P2	30 years, freelancing and editorial experience	Biology, diseases, healthcare	Features, occasional study stories	General interest, science-focused, trade publications
P3	8 yrs, freelancing and staff experience	Biology, animals	News stories, features; videos and books as well	General interest, science-focused publications
P4	2 yrs, freelancing	Science, nature, environment	Study stories, news, features	Science-focused publications, podcasts
P5	Over 30 years, freelancing (7 yrs) and staff, editorial experience	Earth science, environmental issues, health	Study stories, news, features	Science-focused publications
P6	16 yrs, freelancing	Environmental science	Features, news, study stories	General interest, science-focused publications
P7	14 yrs, freelancing and staff experience	Life science, technology	Features, occasional study stories	General interest, science-focused publications
P8	3 yrs, freelancing	Earth science, life science, climate change, diseases, women's health	Features, occasional study stories	General interest, science-focused, trade publications
P9	4 yrs, freelancing	Environmental science, climate, health, wildlife, agriculture, ecology	Features, occasional study stories	Science-focused publications
P10	10 yrs, freelancing	Health science, diseases, space	Features	General interest publications
P11	6 yrs, freelancing and staff experience	Mental health, health justice, health equity	Features, occasional news stories	General interest, science-focused publications, local news
P12	18 yrs, freelancing and staff experience	Neuroscience, physics, other related topics	News, occasional features	Trade, science-focused publications; occasionally general interest
P13	22 yrs, freelancing (recent, 4 yrs) and staff experience	Life science, ocean science	News, study stories, profiles, features	Science-focused and general-interest publications; occasionally trade publications
P14	10 yrs, freelancing	Science, technology, A.I.	Features, podcasts	General interest, trade publications
P15	9 yrs, freelancing	Human evolution, fossils, anthropology	Features, news, study stories	Science-focused publications
P16	7 yrs, freelancing	Space exploration, astronomy, cosmology, general physics	Features, news, study stories, profiles	Science-focused publications
P17	3 yrs, freelancing	Technology, media, healthcare, A.I.	Features	Science-focused publications
P18	6 yrs, freelancing and editorial experience	Astronomy, physics, chemistry, geology, science and culture	Features, news, podcasts	General interest publications
P19	8 yrs, freelancing	Environment, health	Features	General interest publications
P20	25 yrs, freelancing and staff experience	Biomedical science, health science, engineering, technology	Features (more recently)	General interest, science-focused publications, local news

Table 2: Years of experience, beat, and target outlet type for each of our interview participants.