Behind The Paper: A Workflow for Supporting Personal Research Reflection

Tony Zhou
University of Washington
Seattle, Washington, USA
tyzhou05@uw.edu

Tao Long
Columbia University
New York, New York, USA
long@cs.columbia.edu

Gary Hsieh
Human Centered Design & Engineering
University of Washington
Seattle, Washington, USA
garyhs@uw.edu

Ding Li
Adobe Research
Seattle, Washington, USA
ding@dingzeyu.li

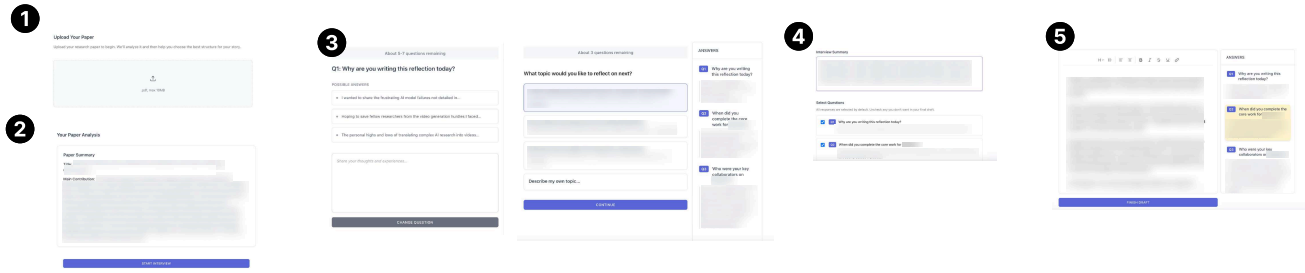

Figure 1: The current workflow of *Behind The Paper*. Users can (1) upload their research paper, (2) review the parsed summary and add notes for the interview. Then, they can (3) answer guided questions and change interview topics at their own pace. Lastly, they can (4) summarize their interview findings, generate a draft, and (5) revise their reflection through a text editor.

Abstract

While researchers publish papers as their primary form of scientific output, the personal journey behind the research process is rarely communicated alongside these polished findings. We posit that these behind-the-scenes reflections on the research process itself might hold significant value, but how are they currently written and shared, if at all? To understand the current practice of *research reflection*—writing and sharing articles describing the personal journey of the research process—we interviewed $n = 11$ authors and collected a corpus of $n = 15,749$ existing research reflections. We analyzed this corpus, derived design goals, and introduce a prototype of *Behind The Paper*, a system that scaffolds the research reflection process, guiding users from paper upload through interview-based elicitation to reflective writing. We discuss implications for supporting future research reflection and conclude with possible directions for future work.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**.

CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3798731

Keywords

research reflection, AI-assisted writing

ACM Reference Format:
Tony Zhou, Tao Long, Gary Hsieh, and Ding Li. 2026. Behind The Paper: A Workflow for Supporting Personal Research Reflection. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3772363.3798731

1 Introduction & Related Work

The advancement of scholarly knowledge through research is inherently uncertain, iterative, and deeply personal [17, 25]. Yet the formal structure of academic publications presents only polished findings, systematically omitting the struggles, surprises, and personal insights that define the research journey [9, 23]. This gap between published papers and lived experience represents a loss of valuable knowledge. How come we never learn about the "hidden curriculum" [14] of research practice, the lessons from failed approaches, and the authentic stories that could mentor future researchers?

We posit that some researchers address this by writing and sharing *research reflections*: first-person retrospective accounts revealing the personal process and lessons learned during the research process, behind what appears in the published paper. For example, a researcher might describe months pursuing a dead-end before a project pivot, providing a story that is never visible in the final

publication (see Appendix A.2). However, to the best of our understanding, no prior work in HCI has systematically characterized and designed tool support for this genre.

While existing science communication work focuses on translating published findings through summarization [16], prescriptive implications [7], or short-form formats [3, 12], centering on the final artifact rather than the personal research journey. Prior relevant work has explored graduate student stress [5, 18] and documented value in sharing failures [23, 26], but no existing tools have explored scaffolding reflection, which has long been recognized as crucial for professional development [6, 19], for researchers. Notably, Schön's existing model distinguishes between reflection *in-action* (real-time adjustment) and reflection *on-action* (retrospective analysis) [13]. Another relevant framework operationalizes reflection on-action assessment through five dimensions: *reporting* (describing events), *responding* (attending to emotions), *relating* (connecting to past experiences), *reasoning* (examining causes), and *reconstructing* (deriving future implications) [1]. While scholars have noted differences between research findings and lived experiences [9] and argued that sharing process stories has "profound impact upon both the research process and its findings" [23], current scholarly communication lacks an understanding of how to formalize such reflection.

Recent AI-assisted writing tools have explored AI support for diary journaling, mental health reflection, and biography writing [8, 20, 22], using conversational Q&A [15] or turn-based prompting [11] to develop narratives. Productized platforms such as NotebookLM [4] can support writing tasks in many domains, and even mobile applications like Voicepal [24] enable on-the-go walk-and-talk translation. While these tools are strong for supporting personal narratives, we hypothesize that our class of artifacts might not be best supported by these existing workflows as it is i) inherently reflective in nature and ii) domain-specific to a research project and journey itself. We aim to explore this gap by applying a known framework of reflective theory [1] to scaffold research reflection alongside an existing knowledge base of reflective writing.

First, we contribute i) a formative interview study with $n = 11$ existing authors, and ii) an analysis of a set of $n = 15,749$ existing articles using Bain's 5R framework [1]. We then derived design goals and present iii) a prototype of *Behind The Paper*, a system designed to scaffold research reflection. We aim to answer the following research questions:

- **RQ1:** How do researchers currently write and share research reflections? What support do they need, if any?
- **RQ2:** How can we design a tool that supports novices for deeper reflection on-action?

2 Formative Interviews & Corpus Analysis

To understand the current practice of research reflection, we conducted 1-hour long semi-structured interviews with $n = 11$ researchers who had written and publicly shared research reflections (see further details and demographics in Appendix A). We recruited participants by querying keywords for HCI reflection articles (2015–2025) on public blogging platforms (*e.g.,* Medium, academic webpages, personal websites) via the Google Search API and then directly emailing authors in June & July 2025. During each interview, we asked questions covering participants' experiences with reflective writing, challenges they faced, as well as perceived benefits and opportunities for support that they could envision. Participants were compensated with $30 USD gift cards and our study was pre-approved by our institutional review board (IRB). Study team members analyzed transcripts with affinity diagramming and inductively derived themes [2].

2.1 Findings

2.1.1 A high-quality research reflection must be personal, reflective, and non-performative. Authors valued that *"a good post is about the personal story"* (A5) that is *"deep enough"* (A4) and focused on personal experience *"rather than just marketing or communicating a finished project"* (A10). A2 noted, *"I kind of want my writing to kind of sound like I'm having a conversation. I wanted to feel like someone thinks I'm talking to them."* Authors noted this genre offers *"minimal direct career benefit"* (A11) and isn't *"written for credit or for your CV"* (A7), especially noting that authentic reflection is starkly different from other channels of more performative communication (A2):

> *"I don't look for the same voice in other people's writing, but I do look for sincerity. I really enjoy reading posts from colleagues who are working through their "midpoint struggles" in public, saying, "I'm stuck on this, and I don't know where to go." It's a fascinating kind of process reporting that shows the public what scholarship and science actually look like."* —A11

2.1.2 Authors are motivated by personal clarity and passing along their experiences for future mentorship. While most authors report reflecting with no specific motivation in mind, the primary benefit noted was for personal self-clarity, because *"the biggest benefit of writing it out was for me."* (A7) Secondary motivations included ungatekeeping hidden curricula for future researchers. A6 explained: *"I benefited so much from others who wrote about their experiences, like Philip Guo. 'The Ph.D. Grind' is what convinced me to do a Ph.D. [...] seeing academic heroes share their thoughts gave me visibility into their worlds."* Authors also believed that this genre of reflection held value that was not immediate:

> *'I think it's more useful to share the authentic story. There's gonna be some failures, and there's gonna be some successes. But knowing how the successes came to be after some failures, or just that a success might not be as glamorous as it seems... that's the most more genuine way to sort of teach or to show what the journey is like, right?"* —A5

2.1.3 The hardest challenge is getting started on transforming personal reflections into shareable outputs. While authors mention that reflection happens over time, when it came to write these reflections, they struggled with translating their reflection into words: *"I might have clarity in my own head after reflecting on conversations, but when I try to put it into words, I always struggle with the first sentence. There are so many ways to express these reflections, and it's challenging to know where to begin"* (A4). As A5 explained,

> *"One of the hardest parts is trying to find a niche that I'm uniquely positioned to write about. I might have*

> *access to certain data, a unique experience, or a specific background that qualifies me. Everyone is writing about their own experience, so I don't want to just write "lessons on how to succeed as a Ph.D.," because I only have one experience, just like everyone else."* — A5

2.1.4 Potential tool support should scaffold reflection while preserving authentic writing. When asked about potential tool support, authors mentioned that prompts to overcome the 'blank page' problem (A7, A9), help for *"ideating different structures for reflection"* (A3), and "asking scaffolded questions after a paper is published" (A1) could be useful. However, they cautioned that tools could *"conflict with my approach"* (A11) and that features that explicitly write for users would mean *"it's not my own reflection anymore"* (A4), highlighting tension between a tool as a scaffolding partner and an authentic reflection. A3 stated, *"I really like writing, and I'm kind of arrogant where I'm like, "this is my writing." I know some people don't feel as much ownership over what they're putting out there, but I think I do quite a bit. I want it to be authentic and artisanal."*

Authors caveat that their challenges were not about knowing what they could write about, but just being unsure of how to transform it into an engaging story. One author stated they were stuck with finding the space to reflect, so *"a tool that captures my attention, instead of social media or email, could provide that structure and motivation to just sit down and do it."* (A8)

2.2 Corpus Collection & Analysis

To ground our question generation in real-world examples, we used a Selenium script to collect all $n = 15,749$ research reflection articles from Springer Nature's *Behind The Paper* publication [21] as .txt files in December 2025. We scored each article across Bain's 5R dimensions [1] with Gemini 2.5-Flash to evaluate whether these articles would be representative of good research reflections (see Appendix A.1). Our initial analysis showed that many of these reflections focus on measures of *reporting* and *relating* to experiences:

- reporting (describing events): $M = 0.72$, $SD = 0.18$
- relating (connecting to past experiences): $M = 0.64$, $SD = 0.21$
- responding (emotional reactions): $M = 0.48$, $SD = 0.24$
- reasoning (examining causes): $M = 0.38$, $SD = 0.22$
- reconstructing (future implications): $M = 0.31$, $SD = 0.19$

To inform our system design, we opted to leverage the 5R framework to support prompting inside *Behind The Paper* through evaluating and targeting possible underrepresented dimensions to support deeper reflection. While these articles skew towards natural sciences, we sampled high-scoring articles which showed that this corpus of personal academic reflections is relevant and generalizable to HCI.

3 The *Behind The Paper* System

3.1 Design Goals

From our interview findings, we inductively derived several system design goals alongside our review of prior work.

- **DG1: Implement a pipeline-based workflow to structure and define a research reflection.** Our system should guide novice researchers through step-by-step scaffolding to address the challenge of not knowing how to begin and also to speed up their workflow (§2.1.1)
- **DG2: Ground reflection scaffolding in real examples from a large corpus.** Our system should leverage patterns from existing research reflections to generate contextual, domain-appropriate questions that feel authentic to researchers (§2.2)
- **DG3: Support progression through reflection dimensions.** Our system should guide users from surface-level reporting toward deeper reasoning and reconstructing, grounded in reflective theory [1] (§2.1.3)
- **DG4: Preserve researcher voice and avoid performative-driven output.** Our system's suggestions should scaffold structure and ideation (§2.1.4) without replacing personal voice and ensuring that the output stays true to the researcher's reflection (§2.1.1)

3.2 System Workflow

Behind The Paper leverages a FastAPI backend server and LLM calls via Gemini 2.5-Flash. We built the frontend user interface in TypeScript as a Next.js web application on top of an open-source text editor[1] (see Figure 1).

(1) **Paper Upload & Retrieval Setup [DG2]:** Users upload their paper PDF, which is parsed[2], embedded[3], and indexed[4] for retrieval-augmented question generation [10] leveraging both the user's paper and our reflection corpus (§2.2).
(2) **Interview Pipeline [DG1, DG2, DG3]:** The system generates questions by retrieving relevant passages from our corpus that match the user's paper content and current reflection depth. Specifically, at each question generation, we prompt *Behind The Paper* to (i) evaluate the user's current responses across all 5R dimensions and (ii) select the next question to target the lowest-scoring or least-represented dimension, creating a gap-filling progression toward deeper reflection (*e.g.,* "How did this challenge your assumptions?" for *reasoning*) and includes three example answer prompts that are reverse-engineered from retrieved segments through high-scoring corpus articles. Users can switch topics, review prior answers in the sidebar, and proceed at their own pace with no time limit.
(3) **Draft Generation & Editing [DG4]:** After completing the interview, users review their responses and can select an audience (*e.g.,* students, peers). The system then leverages the user's responses to generate a structured draft. The draft generation is instructed to not introduce new experiences or claims, and all generated drafts are composed exclusively from the user's interview responses. Users can then modify their draft in a text editor, with interview responses in the sidebar for reference. This aims to ensure that the final reflection maintains the author's authentic voice while benefiting from AI-assisted structure.

[1] https://tiptap.dev/
[2] https://github.com/pymupdf/PyMuPDF
[3] https://ai.google.dev/gemini-api/docs/embeddings
[4] https://github.com/facebookresearch/faiss

3.3 System Walkthrough

David is a third-year PhD student in human-computer interaction. After publishing and presenting a recent paper, he wants to reflect on the research process and share lessons learned with the broader research community, particularly junior PhD students and research assistants. David feels uncertain about how to begin writing a reflection and struggles to organize his thoughts. He uploads his paper to *Behind The Paper*.

Behind The Paper starts with conversational prompts like, "Why did you start this project in the first place?" David answers that he originally thought it would be "a pretty straightforward study," but quickly realized it was messier than expected and went down a rabbithole there. The system follows up with relevant questions such as, "Was there a moment where you thought, 'oh no, this is not what I signed up for'?" David recalls a week where nothing in his qualitative data made sense and he seriously questioned whether the project would get past the finish line at all.

Later, the system asks, "What part of this paper almost didn't make it to publication?" David talks about nearly cutting out a section he cared about because of reviewer pushback. The system then asks, "Looking back now, what did that experience teach you?" He reflects that he learned to better separate what he cared about personally from what the paper needed to communicate.

After completing the interview, David reviews and edits a generated draft tailored to a junior research audience, refining it into a coherent reflection that captures both his struggles and growth, which he then posts on his personal website.

3.4 Pilot Feedback

We informally evaluated the current prototype of *Behind The Paper* with two PhD students. They described that they appreciated the current system's approach in eliciting reflection as they were *"surprised to see it guessing angles I hadn't considered"* and that they liked that the process works by *"forcing you to stay on track."* However, both noted that the workflow felt demanding in its transition from interview to draft generation and editing. For example, they stated, *"the reflective questions are not directly tied to what I want to say"* and that *"there are too many questions."* Participants suggested making the current prototype more iterative (*e.g.,* without strictly asking questions upfront) and also justifying the scaffolding design decisions further in evidence-based writing methods.

4 Next Steps, Discussion, and Conclusion

4.1 Limitations

We present *Behind The Paper* as a prototype. As we continue to develop *Behind The Paper*, formal user study and comparison to a valid baseline condition is necessary to assess effectiveness for reflection depth, writing quality, and user experience.

We focused on text-based reflections published as articles, but research reflections also exist in other media (*e.g.,* podcasts, videos). Additionally, our formative interviews were also with experienced authors and we designed *Behind The Paper* for novices. Future work should consider understanding the needs of these populations. Also, our corpus of existing research reflection articles come from Springer Nature, which may skew toward natural sciences. Expanding to HCI-specific reflections may improve domain relevance. Our prototype supports single-session use, yet authors also noted value in research reflection as a longitudinal process (*e.g.,* thinking on-the-go, revisiting notes or drafts over days, weeks, or months). Future work should explore multi-modal and longitudinal opportunities to support reflection.

4.2 Future Directions

4.2.1 Opportunities for reflection on-action. Our current version of *Behind The Paper* provides scaffolding primarily during the interview phase and we designed this to ultimately generate a draft to speed up the reflective process **(DG1)**. Because our interviewed authors valued writing themselves, future work could explore in-situ support for documents (*e.g.,* a browser extension to overlay on top of all document editors). This specific genre of reflective writing might be a space to explore new lightweight mechanisms to support reflection *on-action*. For example, what might a *reflective autocomplete* interaction that autocompletes your thoughts but not your writing look like? What might a document-native reflective experience look like? These are promising directions future work should consider beyond a pipeline-based workflow.

4.2.2 Adaptive and personalized reflective scaffolding. Our system currently treats all users uniformly, which personal reflection may not work well with. While we explored personalization features, future tools should consider adapting to career stage (an early-year PhD student versus a tenured professor), clarify primary writing goals (mentoring versus a personal reflection), or reflection style (narrative versus analytical), tailoring question selection and depth progression accordingly.

4.2.3 A comparative system evaluation. In our design and iteration, we were motivated to build a structured experience to support novices. But how does this compare to a standard conversational chatbot or blank-page writing? Future studies should systematically evaluate these design choices. Future work should consider developing a reverse-engineered question bank or model for stronger personalized reflective questions that are grounded on real reflection examples instead of off-the-shelf LLM prompting.

Additionally, we selected our interview participants' articles for their reflective quality through manual review, to which our measurements of reflective quality would be redundant for recruitment purposes. However, we could also retrospectively continue to analyze these (and more existing reflections) to compare author-written reflections to system-supported reflections in future evaluation.

4.3 Conclusion

We identified and defined *research reflections* as an underexplored class of written reflections on research projects. To gain an understanding of this class of articles, we interviewed existing authors ($n = 11$), collected and analyzed existing research reflections ($n = 15,749$), and derived design goals. We presented our development of *Behind The Paper* as a prototype to support the research reflection process, and enumerated key directions for future work in this space. Beyond our current system prototype, we see potential for a new class of writing tools: not just for making writing faster, but those that support intrinsically human values [27]. Our work

aims to take a step toward supporting more meaningful activities that might take more time but hold more potential personal value.

References

[1] John D Bain, Roy Ballantyne, Jan Packer, and Colleen Mills. 1999. Using journal writing to enhance student teachers' reflectivity during field experience placements. *Teachers and teaching* 5, 1 (1999), 51–73.

[2] Victoria Clarke and Virginia Braun. 2014. Thematic analysis. In *Encyclopedia of critical psychology*. Springer, 1947–1952.

[3] Meziah Ruby Cristobal, Hyeonjeong Byeon, Tze-Yu Chen, Ruoxi Shang, Donghoon Shin, Ruican Zhong, Tony Zhou, and Gary Hsieh. 2026. PaperTok: Exploring the Use of Generative AI for Creating Short-form Videos for Research Communication. *arXiv preprint arXiv:2601.18218* (2026).

[4] Google. 2025. NotebookLM. https://notebooklm.google.com/

[5] Jane Im, Himanshu Zade, Steve Oney, Pamela Wisniewski, and Kentaro Toyama. 2024. Improving Advising Relationships Between PhD Students and Faculty in Human-Computer Interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 587, 4 pages. doi:10.1145/3613905.3643971

[6] Christopher Johns. 2025. *Becoming a reflective practitioner*. John Wiley & Sons.

[7] Taewook Kim, Dhruv Agarwal, Jordan Ackerman, and Manaswi Saha. 2024. Steering AI-Driven Personalization of Scientific Text for General Audiences. *arXiv preprint arXiv:2411.09969* (2024).

[8] Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. 2024. DiaryMate: Understanding user perceptions and experience in human-AI collaboration for personal journaling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[9] Baz Lecocq. 2002. Fieldwork Ain't Always Fun: Public and Hidden Discourses on Fieldwork1. *History in Africa* 29 (2002), 273–282.

[10] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* 33 (2020), 9459–9474.

[11] Hyunseung Lim, Ji Yong Cho, Taewan Kim, Jeongeon Park, Hyungyu Shin, Seulgi Choi, Sunghyun Park, Kyungjae Lee, Juho Kim, Moontae Lee, et al. 2024. Co-Creating Question-and-Answer Style Articles with Large Language Models for Research Promotion. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 975–994.

[12] 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 preprint arXiv:2305.12265* (2023).

[13] Hugh Munby. 1989. Reflection-in-action and reflection-on-action. *Current issues in education* 9, 1 (1989), 31–42.

[14] Kendall Nakai and Philip J Guo. 2023. Uncovering the hidden curriculum of university computing majors via undergraduate-written mentoring guides: A learner-centered design workflow. In *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 1*. 63–77.

[15] Soya Park, Hari Subramonyam, and Chinmay Kulkarni. 2024. Thinking Assistants: LLM-Based Conversational Assistants that Help Users Think By Asking rather than Answering. arXiv:2312.06024 [cs.HC] https://arxiv.org/abs/2312.06024

[16] Marissa Radensky, Daniel S. Weld, Joseph Chee Chang, Pao Siangliulue, and Jonathan Bragg. 2025. Papers-to-Posts: Supporting Detailed Long-Document Summarization with an Interactive LLM-Powered Source Outline. arXiv:2406.10370 [cs.HC] https://arxiv.org/abs/2406.10370

[17] Peter Reason and Judi Marshall. 1987. Research as personal process. *Appreciating Adult Learning. London: Kogan Page* (1987), 112–126.

[18] Fujiko Robledo Yamamoto, Janghee Cho, Amy Voida, and Stephen Voida. 2023. "We are Researchers, but we are also Humans": Creating a Design Space for Managing Graduate Student Stress. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 75 (Sept. 2023), 33 pages. doi:10.1145/3589956

[19] Donald A Schon and Vincent DeSanctis. 1986. The reflective practitioner: How professionals think in action.

[20] Inhwa Song, SoHyun Park, Sachin R Pendse, Jessica Lee Schleider, Munmun De Choudhury, and Young-Ho Kim. 2025. Exploreself: Fostering user-driven exploration and reflection on personal challenges with adaptive guidance by large language models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.

[21] Springer Nature. 2025. Behind The Paper. https://communities.springernature.com/channels/behind-the-paper

[22] Shayan Talaei, Meijin Li, Kanu Grover, James Kent Hippler, Diyi Yang, and Amin Saberi. 2025. StorySage: Conversational Autobiography Writing Powered by a Multi-Agent Framework. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–26.

[23] Susan Thomson, Ann Ansoms, and Jude Murison. 2012. *Emotional and ethical challenges for field research in Africa: The story behind the findings*. Springer.

[24] Voicepal. 2026. A ghostwriter in your pocket. https://voicepal.me/

[25] Anita F Young. 2004. Becoming a practitioner-researcher: A personal journey. *British Journal of Occupational Therapy* 67, 8 (2004), 369–371.

[26] Annie Li Zhang and Hang Lu. 2025. When Scientists Share Their Struggles: How Scientists' Self-Presentation on Social Media Influences Public Perceptions, Support for Science, and Information-Seeking Intentions. *Science Communication* (2025), 10755470251322902.

[27] Haoqi Zhang. 2024. Searching for the Non-Consequential: Dialectical Activities in HCI and the Limits of Computers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–13.

A Formative Study Participants

We utilized the Google Search API to query existing research reflections in HCI from 2015–2025. To ensure participant anonymity, we do not report our exact criteria, but our criteria included that:

- Articles were publicly available on blogging websites (*e.g.,* Medium, Substack, Wordpress) or academic communication channels (*e.g.,* personal academic website, university-affiliated lab website).
- Articles were written by a sole author in first-person and the primary goal of the blog was for personal reflection.
- Articles revealed additional, personal insights and content that would not have been apparent from otherwise just reading the author's publications.

Authors initially reported challenges as lack of time (n = 7), unsure about what content would be interesting or appropriate (n = 6), difficulty getting started (n = 5), and concerns about sharing failures publically (n = 5). All except one reported maintaining an active social media account for discussing or viewing research.

A.1 Article scoring

```
Analyze this article for personal reflection using Bain's 5R
      reflection framework, detailed below. Focus on the
     author's personal journey, not technical content.

Score each dimension from 0.0-1.0 and return a sum of the
     scores. Ensure that promotional or overly technical
     content is scored low.

Reporting (0-1): Personal experiences and specific moments
- Score high: "We spent weeks debugging", "I remember the
     moment we realized"
- Score low: Generic methods, "we developed X", promotional
     language
- Must be personal storytelling, not technical summary

Responding (0-1): Personal emotions and thoughts
- Score high: "I felt frustrated", "We were surprised", "
     This made me doubt"
- Score low: No emotions, or formulaic ("we were excited to
     publish")
- Must express genuine feelings about the process

Relating (0-1): Connections to past personal experiences
- Score high: "This reminded me of my PhD", "Similar to when
      I worked on Y"
- Score low: Generic citations, no personal past connections
- Must explicitly connect to author's own history

Reasoning (0-1): Personal analysis of why things happened
- Score high: "Looking back, I realize we failed because I
     assumed X"
- Score low: Technical mechanisms, "the system works because
      of X"
- Must be reflection on personal decisions/actions, not
     technical explanations

Reconstructing (0-1): Personal lessons and changed practices
- Score high: "Next time I would start by", "I learned to
     always X"
- Score low: "Future work could", "This opens pathways for",
      generic implications
- Must be specific personal insights about changed behavior

Additional rules:
- Promotional content -> all scores near 0.0
- Technical explanations -> Reasoning = 0.0 (not reflective)
- Generic future work -> Reconstructing = 0.0 (not personal
     lessons)
- No emotions mentioned -> Responding = 0.0
- No past connections -> Relating = 0.0

Return your answers in standard JSON format:
{"reporting": 0.0,
"responding": 0.0,
"relating": 0.0,
"reasoning": 0.0,
"reconstructing": 0.0,
"score_sum": 0.0}
```

Table 1: Formative study participant demographics. We asked authors to rate their usage of common AI tools in writing-related tasks on the following: Never; Rarely (a few times a month); Occasionally (a few times a week); Frequently (a few times a day).

Participant	Domain	Experience	Article (Year)	Article Length	AI Writing Usage
A1: Professor	HCI	5-9 years	Personal Website (2025)	<1000 words	Frequently
A2: PhD Student	HCI	5-9 years	Medium (2019)	1000-2000 words	Occasionally
A3: PhD Student	NLP	5-9 years	Medium (2021)	1000-2000 words	Rarely
A4: PhD Student	HCI	1-4 years	Personal Website (2025)	<1000 words	Frequently
A5: Professor	HCI	10-16 years	Personal Website (2022)	>2000 words	Never
A6: Professor	HCI	10-16 years	Medium (2021)	1000-2000 words	Rarely
A7: Professor	NLP	16+ years	Medium (2021)	>2000 words	Never
A8: PhD Student	HCI	5-9 years	Personal Website (2020)	1000-2000 words	Rarely
A9: Professor	HCI	16+ years	Academic Website (2022)	>2000 words	Occasionally
A10: Professor	HCI	16+ years	Medium (2018)	1000-2000 words	Occasionally
A11: Professor	HCI	16+ years	Medium (2025)	>2000 words	Never

A.2 Article .txt collection

A sample post scored as considerably reflective.

> The paper in Nature is here: http://go.nature.com/2ikJmLn
> I was doing my Ph.D. in what I guess would be called a 'classic' microbiology lab in 2005, when I read a first article on sequencing-based gut microbiota analyses – a new methodology that had yet to evolve into microbiome research. While the potential of the technique was clear from the start, to be honest, my colleagues and I were not too impressed with the first results presented. There we were, spending our days and nights following up fermentation experiments with fecal isolates - diluting, plating, and counting. And suddenly, there was this article, published in a major journal that we thought out of reach for gut microbiologists, describing microbiota compositional variation on taxonomical levels that oscillated between self-defining units and all-encompassing phyla. The authors did not even attempt to quantify any of the observations described. Everything became proportional – pure horror to us, dedicated followers of Drigalski.
> In 2010, I joined the Raes Lab, back then still in its start-up phase at the VIB - Vrije Universiteit Brussel (Belgium). Even in those days, gut microbiome research was still in its infancy. I witnessed the field mature and was lucky enough to be able to contribute my bit to the process. Yet, we never seemed to get rid of the elephant in our microbiome room: all data remained relative. Compositional, as I was taught to term it correctly. Sure, we had some ideas on how to bypass compositionality. And of course, there are some statistical methods available to minimize its impact on results. Still, the integration of computational and wet lab side of microbiome research that we try to achieve at the Raes Lab made it tempting to try to tackle the problem at the source. We just needed some time and resources - and a truckload of fresh fecal samples.
> Things started moving in 2014, when the Raes Lab moved to Leuven (Belgium). While it is not always easy to integrate method development in valorization-targeted research projects – the standard among Belgium funding opportunities – the KU Leuven, our new host institution, offered us a way out by providing credit for postdocs wanting to set up high risk/high gain research lines. With this funding secured, we could start thinking about logistics. Wanting to quantify microbial loads in fecal samples, we needed fresh material. Clearly, there was no choice: we had to abandon our frozen sample Flemish Gut Flora Project comfort zone and adapt the lab's logistics to the real stuff. We invited locally recruited volunteers to deposit anonymously their coded contribution in a cardboard mailbox just outside the lab. The incentive consisted in an assortment of fruits and cakes placed on top of the mailbox. When a sample arrived, lab hell broke loose as we rushed to complete our analyses within the hour after egestion.
> Of course, our protocols became a lot more relaxed when we found cell counts in fresh and frozen fecal material correlated. Working with frozen samples not only facilitated flow cytometry analyses, it also increased the applicability of the method dramatically. Indeed, frozen fecal material is the raw material and starting point for analyses in most microbiome labs, including ours. The method we proposed is rather straightforward: multiple aliquots of frozen fecal samples are used to parallelize flow cytometry assessment of microbial loads with microbiome sequencing. Next, count data allow correcting the sequence matrix for sampling depth. Simple, but not trivial. We first observed an up to tenfold difference in fecal cell densities between healthy individuals – increasing to a fifty-fold when patients were included in the analyses. Obviously, until now, we had been over-sampling low load samples when compared to high density ones. Unexpectedly, our analyses also showed that

differences in microbial loads might be differentiating enterotypes. Next, we demonstrated how quantitative microbiome profiling affected the outcome of both taxa co-occurrence analyses and associations with metadata. Finally, we identified microbial loads to be a key feature of disease-associated microbiota dysbiosis, specifically in Crohn's Disease.

Within the Raes Lab, quantitative microbiome profiling was quickly adopted to become the standard in our analyses. While it does not necessarily affect all previously described microbiome signals, it has already generated some exciting results that we hope to be able to share with the community soon. We do hope that our excitement regarding the technique will be shared by many in the field. We are not claiming to have resolved every outstanding methodological issue in the microbiome field – or that the method proposed is the only strategy conceivable to deal with compositionality and variation in cell densities. We do think that our paper and the publication platform it received will make it impossible to continue ignoring the elephant that has been stomping around for over a decade in microbiome research. Quantitative analyses are what the field needs to start moving from the bench to the bedside.