One Is Not Enough: How People Use Multiple AI Models in Everyday Life

ANONYMOUS AUTHOR(S)

People increasingly use multiple Multimodal Large Language Models (MLLMs) concurrently, selecting each based on its perceived strengths. This cross-platform practice creates coordination challenges: adapting prompts to different interfaces, calibrating trust against inconsistent behaviors, and navigating separate conversation histories. Prior HCI research focused on single-agent interactions, leaving multi-MLLM orchestration underexplored. Through a diary study and semi-structured interviews ($N = 10$), we examine how individuals organize work across competing AI systems. Our findings reveal that users construct primary and secondary hierarchies among models that shift over usage context. They also develop personalized switching patterns triggered by task aggregation to adjust effort and latency, and output credibility. These insights inform future tool design opportunities, supporting users to coordinate multi-MLLM workflows.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Multimodal Large Language Models (MLLMs), Multiple AI Models, Diary study

ACM Reference Format:

Anonymous Author(s). 2025. One Is Not Enough: How People Use Multiple AI Models in Everyday Life. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

AI assistants have become ubiquitous productivity tools, with ChatGPT[1], Gemini[2], and Claude[3] now supporting diverse tasks including writing [5], programming [15], and health information seeking [20]. As these Multimodal Large Language Models (MLLM)—AI systems capable of processing text, images, and code—have matured, people have begun adopting multiple platforms concurrently rather than relying on a single provider [8]. This multi-MLLM practice arises because each system offers distinct strengths: individuals select specific models for particular capabilities, such as one for logical reasoning and another for creative generation [14, 21]. Strategically utilizing specific model capabilities has emerged as a practical skill for optimizing individual workflows and achieving high-quality outcomes [5, 14]. With MLLMs evolving rapidly and multi-platform use becoming routine, users increasingly mix and match models within a single task to leverage their complementary strengths [4, 7]. Yet coordinating across multiple competing MLLM platforms introduces cognitive overhead that existing research has not fully addressed [9].

Prior work on multi-device ecologies shows that people face friction when transferring context across platforms [9]. This context-delegation between platforms challenge intensifies with AI systems, which require users to adapt to prompting styles [21], calibrate expectations against inconsistent behaviors [13], and manage divergent conversation histories [11, 16]. Foundational HCI research has examined trust calibration [10, 18] and collaborative dynamics [6]

[1]https://chatgpt.com/
[2]https://gemini.google.com/
[3]https://claude.ai/

Manuscript submitted to ACM

within single-agent interactions. However, these frameworks assume a stable human-AI dyad rather than scenarios in which individuals switch between competing systems [12]. Recent studies of generative AI adoption in professional contexts focus on individual tool use rather than cross-platform orchestration [17]. Consequently, we have limited empirical understanding of how people develop strategies to allocate tasks, reconcile conflicting outputs, and maintain coherent workflows across multiple MLLM platforms.

This study addresses this gap through a four-day diary study and semi-structured interviews examining how users organize and coordinate multiple MLLMs in everyday practice. We contribute: (1) an empirical characterization of the roles and relational hierarchies users assign to different MLLM systems, and (2) a taxonomy of coordination strategies users employ to manage cross-platform workflows. These findings offer design implications for tools that support users navigating an increasingly heterogeneous AI landscape. Two research questions guide our inquiries:

- **RQ1)** What mental models do users construct when distributing tasks across multiple MLLMs?
- **RQ2)** What strategies do users develop to coordinate their interactions across multiple MLLM platforms?

2 Methodology

2.1 Diary Study and Post-study Interview

To examine how people organize and coordinate multiple MLLMs in everyday practice, we conducted a qualitative design study that combines diary documentation with follow-up interviews, following recent methodological frameworks for evaluating AI tool adoption in-situ [3]. Over four days, participants logged their MLLM use through a web-based diary interface, submitting an entry each time they used an MLLM. Each entry recorded the model used, the rationale for selection, the prompt content, satisfaction, and emotional state, providing in-the-moment accounts of model coordination (see Appendix A). After the diary period, we conducted follow-up interviews (3 in-person, 7 remote) to further examine participants' model choices and workflow strategies. Interviews were facilitated by two researchers, with one lead researcher attending all sessions to ensure protocol consistency; questions probed perceived capabilities, role assignments, and workflow organization. Our study was approved by the Institutional Review Board.

2.2 Participants

We recruited ten adults (6 female, 4 male; aged 23–29, $M = 26.1$, $SD = 1.60$; see Table 1) who met two criteria: (1) regular use of at least two MLLM services and (2) practical experience in using different models complementarily for personal or professional projects. This ensured that all participants possessed sufficient familiarity with cross-platform coordination. Participants were compensated approximately $35 USD in local currency.

Table 1. Self-reported Participant Demographics and MLLM Models Used

ID	Age	Gender	Occupation	MLLM Models Used	Other AI Services Used
P1	26	M	Graduate student in Computer Science	ChatGPT, Gemini, Perplexity	Cursor
P2	27	M	Wildlife photographer	ChatGPT, Gemini, Claude	
P3	29	F	Graduate student in Industrial Design	ChatGPT, Gemini, Claude	
P4	25	F	Graduate student in Electrical Engineering	ChatGPT, Gemini	
P5	23	F	Graduate student in Electrical Engineering	ChatGPT, Gemini, Perplexity	Copilot
P6	25	F	Engineer / Software Developer	ChatGPT, Gemini, Claude, Samsung Gauss	
P7	27	F	Visual Development Designer	ChatGPT, Gemini, Claude	Copilot, Adobe Firefly, Freepik
P8	26	M	Full-stack Developer	Gemini, Claude	Antigravity
P9	27	M	Graduate student in Industrial Design	ChatGPT, Gemini	
P10	26	F	Graduate student in Industrial Design	ChatGPT, Gemini, Claude	

2.3 Data Collection and Analysis

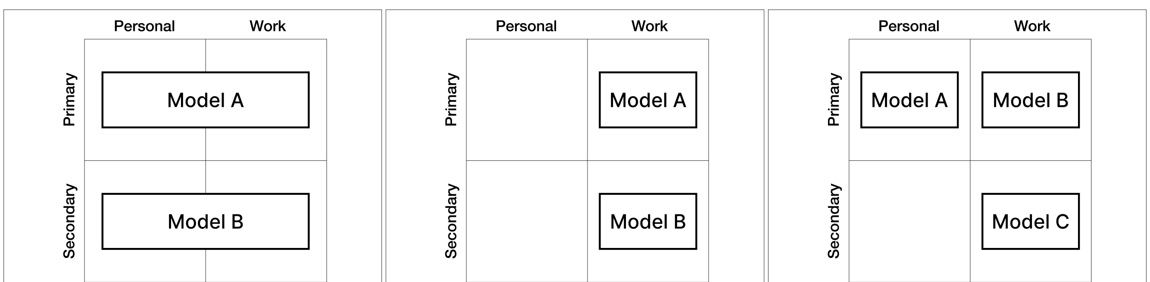
We collected 129 diary entries (M = 12.9, SD = 3.41) and conducted 10 post-study interviews (duration: M = 34.0, SD = 3.4 minutes). Interviews were audio-recorded, transcribed, and translated into English. Three researchers conducted an inductive thematic analysis following Braun and Clarke [2]. We iteratively developed a shared codebook, generated expanded codes, and consolidated them into 11 final codes and four subthemes, which were synthesized into two overarching themes: (1) Individual MLLM Hierarchy Structures and (2) Cross-Platform Coordination Strategies. The coding process was iterative, with all authors holding regular discussions to resolve discrepancies and reach consensus.

3 Findings

Our findings show that participants employed diverse, personalized ways of using and coordinating MLLMs. In the following section, we present: (1) the hierarchy structures participants formed across models, and (2) the strategies they used to navigate and coordinate multi-MLLM workflows.

3.1 Individual MLLM Hierarchy Structures

3.1.1 Different Hierarchy Across Personal and Work Contexts. When coordinating multiple MLLMs, all ten participants chose to utilize the primary model and turned to one or more secondary models as needed. The primary model was used for frequent, core tasks, while secondary models were used for verification (P2, P3, P5, P7), refinement (P3, P6, P9, P10), or more specialized needs (P2, P4, P5, P6, P7, P9, P10). However, its configuration shifted with context—particularly between personal and professional use. Across participants, we identified three recurring hierarchy patterns that describe how model roles were arranged across personal versus work contexts (see Figure 1).



(a) Cross-Context Primary Structure (P2, P6, P7, P10)
The same set of primary and secondary models is shared across both personal and work contexts.

(b) Work-Focused Structure (P3, P4, P5)
Both primary and secondary are composed exclusively of work-oriented models, with no personal-context models included.

(c) Split-Primary Structure (P1, P8, P9)
Distinct primary models are used depending on context (personal vs. work), with each context having its own primary model.

Fig. 1. Three recurring hierarchy patterns: (a) Cross-Context Primay, (b) Work-Focused, and (c) Split-Primary.

(a) Cross-Context Primary Structure Four participants (P2, P6, P7, P10) centered their workflows around a single favorite model used across both personal and professional contexts. They perceived the favorite model as a general-purpose assistant and kept interactions consolidated. Participants framed the preference around familiarity and accumulated interaction history—describing their favorite model as one that could interpret intent with minimal re-explanation. P6, for example, used ChatGPT for both emotional support and drafting workplace reports: *"I prioritize it because of the familiarity... when I ask a question, it already knows what I mean. I don't have to explain myself too much."* While these participants primarily stayed within their go-to model, they occasionally brought in secondary models when they wanted a second opinion or when a task called for capabilities they associated more strongly with another system.

(b) Work-Focused Structure However, three participants (P3, P4, P5) kept one primary model dedicated to professional tasks, treating it as a productivity tool. They deliberately did not feel a strong need to use this primary model in personal contexts. P5 said: *"I never use it for personal counseling because I don't think it would be helpful... it just feels like a waste of time."* Instead, they brought in secondary models mainly to improve work efficiency—for example, to validate outputs, compare responses across models, and catch errors during professional workflows. P3 added: *"I paste what Claude said into ChatGPT and ask: 'Claude thinks this, what do you think?' to provoke a debate. I want them to discuss and find a compromise."*

(c) Split-Primary Structure Furthermore, three participants (P1, P8, P9) maintained distinct primary models for personal versus professional contexts. They typically used a lighter or faster option for low-stakes personal tasks, while reserving a stronger model for work tasks that demanded higher reliability. P1 explicitly separates his tools based on the cost of failure: *"For casual queries like pharmacy recommendations or vehicle maintenance, I use Gemini Fast because it's the quickest and doesn't require reasoning. But for work, I strictly use the highest-spec model (GPT-5.2) because fixing errors from a cheaper model wastes more of my time than the token cost."* For these participants, secondary models were primarily considered for task-oriented work needs, whereas personal use tended to remain within the chosen personal primary model. P8 elaborated: *"I switch based on the specific feeling I remember from past usage. If the task requires grasping long contexts, I use Gemini. But if the logic is complex, Claude feels smarter, so I use Claude specifically for that."*

3.1.2 Factors Shaping Model Hierarchies. Across our study, participants formed model hierarchies, with the particular MLLM occupying each role continually shaped by (1) first impressions, (2) expert consensus and social signals, and (3) costs of model use, as participants re-evaluated their options.

First-Impression Lock-in Early experiences with a model often set a reference point that continued to shape later preferences (P3, P4, P8). For instance, P3 established Claude as their primary writing agent based on a strong initial impression, contrasting its calm style with ChatGPT's performative tone: *"ChatGPT felt like it was just showing off with fancy words... whereas Claude felt calm, concise, and clear. Since that moment, that impression became solidified, so I still maintain high trust in Claude for writing."* This lock-in was especially pronounced when the model's style or reasoning process felt aligned with how participants approached the task. P2, a wildlife photographer, favored Gemini because its image analysis process matched their birdwatching expertise: *"Gemini classifies birds exactly like a birdwatcher would, checking critical points like the beak shape or shin feathers first... The way it structures its observation matches my own thinking process."*

Expert Consensus and Social Signals Some participants (P8, P9) stayed attentive to which models were perceived as improving or leading, using expert opinions and community discussions as cues for when to re-evaluate their tool choices. P9 treated expert consensus as a strategic filter, monitoring the market without the burden of constant hands-on testing: *"I can never know better than those AI researchers... So I just trust their reviews. If they say on LinkedIn that a model is rising, that becomes my only standard."* Others (P2, P3, P7, P10) reported trying models after hearing positive recommendations from people around them. In contrast, P1 deliberately bypassed social trends and relied strictly on technical performance evidence: *"I don't look at social reactions. I only trust objective benchmarks like SWE-bench... If a new model scores higher, I switch immediately because relying on an inferior tool wastes my time."*

Costs of Model Use Model preferences were sometimes shaped less by perceived performance, as they balanced task suitability against recurring subscription fees or usage costs (P1, P5). For example, P1 offloads trivial daily queries to Gemini Fast Mode to minimize operational costs, noting that the financial burden of his high-frequency usage necessitates strict budget management *"My usage volume is huge... I already spend over 10 dollars a day. If I calculate that monthly, it is an enormous amount, so cost is a very critical factor in my choices."*

3.2 Cross-Platform Coordination Strategies

3.2.1 Switch Model with Purpose. Participants developed strategies to coordinate their interactions across multiple MLLM platforms, switching intentionally by assigning different task stages to different models, adjusting effort and latency to task demands, and cross-checking credibility-critical information.

Assigning Models to Specialized Roles A common coordination strategy was to break a task into stages and assign each stage to the model perceived as strongest for that subtask (P2, P5, P6, P7, P9)—for example, ideation with one model, drafting with another, and polishing or refinement with a third. P6 described this transition for their SOP writing workflow: *'I initially brainstormed all my ideas with ChatGPT... it had really nice suggestions. But for the writing style, I took the same essay from ChatGPT and put it in Claude for the polishing."* This sequential switching reflected a deliberate choice to leverage different models' strengths at different points in the workflow.

Managing Effort by Task Difficulty Three articipants (P1, P5, P9) also switched models based on the scale and difficulty of the immediate task to maximize efficiency. This was a deliberate choice to avoid over-investing effort or waiting for a slow, high-performance model to complete a trivial job. P9 switched between modes within a single service to manage latency, treating them as agents with different speeds and ranks: *"If I require logical flow, I use Gemini Thinking Mode. But to change just one word in the result, I switch to Fast Mode immediately... It feels like handing a trivial task to a faster, lower-ranked entity who doesn't need to think deeply."*

Cross-Checking for Credibility When information needed to be reliable, participants (P2, P3, P5, P7) adopted a cross-checking strategy by switching across platforms to verify outputs and reduce hallucination risk. This strategy was often implemented as a simple one-step verification: participants took the initial output to another model to confirm key claims. While many participants relied on this single-step check, P7 used a three-stage sequence (ChatGPT → Gemini → ChatGPT) to locate evidence and then return to the original model to assess validity: *"I paste ChatGPT's output into Gemini and ask it to find supporting evidence. Then, I feed that evidence back into ChatGPT and ask: 'Is this evidence actually correct?'... It is essentially a three-stage validation process."*

3.2.2 Don't Migrate, Iterate Instead. Rather than migrating across platforms, some participants sought to improve output quality by iterating with a single model. They described two reasons for this choice: (1) the model already knew their context, making switching costly; and (2) when outputs fell short, they sometimes located the source of failure in their own input specification rather than model capability, leading them to revise and clarify their requests instead of migrating.

No Resetting the Relationship Four participants (P2, P4, P6, P10) sometimes chose to stick with a familiar model because re-establishing background and personal context elsewhere felt more costly than any likely performance gain. P2 avoided switching away from Gemini for personal hobbies, noting that retraining a new model on his specific persona would require excessive effort: *"Gemini knows I'm crazy about birds... but Claude would just think I'm a weirdo. I could theoretically spend a whole day taming Claude to understand my context, but that is just too annoying."*

Blame the Prompt, Not the Model When results were unsatisfying, participants (P1, P8, P9) sometimes focused on improving how they prompted rather than blaming the model, iterating until the output met their needs. P9 described this mindset explicitly, comparing poor prompting to user error in driving: *"Being unsatisfied after lazy prompting is* *like a bad driver claiming a Tesla is a bad car... I just need to steer it better."* P1 similarly framed prompt refinement as a debugging exercise, iterating on prompts and test cases with the same model until the output met their requirements: *"If the result is unsatisfactory, it is usually because I gave insufficient information. So I don't switch; I just provide more test cases and grill the model until it passes the criteria."*

4 Discussion

Our findings show that participants managed MLLM ecosystems through hierarchical organization and deliberate coordination rather than relying on a single assistant [4]. Across contexts, users treated models as an evolving repertoire: they assigned primary and secondary roles, re-evaluated these roles in response to shifting signals (e.g., first impressions, social cues, and costs), and developed routines for switching, verification, and effort allocation. Whereas much prior work has emphasized single-system use [1, 19], our results highlight cross-model coordination as an ongoing practice of orchestrating an AI ecosystem. Building on this, we highlight two design opportunities. First, tools could better support multi-model use by being workflow-aware, preserving the user's task structure and context while helping them choose an appropriate model for the situation at hand. For instance, we can design a system that users could be able to move from ideation in one model to drafting or polishing in another without having to reconstruct the task framing or re-specify key constraints each time. This would let users leverage complementary strengths with less friction when switching. Second, memory partitioning could help users to protect distinct model contexts. While existing solutions offer a binary choice between full retention and total loss, this approach is considered unoptimal for supporting granular memory management [13]. For example, topic containers could segregate histories by domain (e.g., "Coding," "Health," "Creative") within a single interface. This selective context portability would advance beyond Amershi et al. [1]'s transparency guidelines toward user-controlled curation.

5 Limitations and Future Work

We acknowledge two primary limitations that motivate future work. First, our participants were primarily in their twenties (N=10) and drawn from a limited range of occupations; while this sample size is appropriate for qualitative inquiry, broader age groups and more diverse professional backgrounds would help assess how model hierarchies and coordination practices vary across populations and domains. Second, our four-day diary captured everyday patterns but not longer-term dynamics (e.g., shifts after major updates); longitudinal work could track how releases or pricing changes reshape preferences over time.

6 Conclusion

We presented a qualitative study of how users navigate and coordinate across multiple MLLMs. Through a diary study and follow-up interviews, we characterized (1) how participants structured model hierarchies across personal and professional contexts and (2) the coordination strategies they used across platforms. This work is timely because, as MLLM ecosystems evolve rapidly, users increasingly need practical skills for selecting, combining, and validating model outputs in everyday practice. Our findings offer an empirical account of how users orchestrate multiple models, complementing prior human–AI interaction work that has primarily examined single-system use. These results point to design opportunities for helping users adapt to evolving AI ecosystems by supporting effective model selection, cross-model coordination, and reliability-oriented use.

References

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[2] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. arXiv:https://doi.org/10.1191/1478088706qp063oa doi:10.1191/1478088706qp063oa

[3] Jenna Butler, Jina Suh, Sankeerti Haniyur, and Constance Hadley. 2025. Dear Diary: A Randomized Controlled Trial of Generative AI Coding Tools in the Workplace . In *2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE Computer Society, Los Alamitos, CA, USA, 319–329. doi:10.1109/ICSE-SEIP66354.2025.00034

[4] Butler, Jenna, Jaffe, Sonia, Janßen, Rebecca, Baym, Nancy, Hecht, Brent, Hofman, Jake, Rintel, Sean, Sarrafzadeh, Bahar, Sellen, Abigail, Vorvoreanu, Mihaela, and Teevan, Jaime. 2025. *Microsoft New Future of Work Report 2025.* Technical Report MSRTR-2025-58. Microsoft Research. https://aka.ms/nfw2025

[5] H. Dang, J. Lehman, and D. Buschek. 2024. Choice Over Control: How Users Write with Large Language Models Using Diegetic and Non-Diegetic Prompting. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA.

[6] Katy Ilonka Gero, Tao Long, and Lydia B Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 245, 15 pages. doi:10.1145/3544548.3580782

[7] Xinyi Hou, Yanjie Zhao, and Haoyu Wang. 2025. LLM Applications: Current Paradigms and the Next Frontier. arXiv:2503.04596 [cs.SE] https://arxiv.org/abs/2503.04596

[8] Imagining the Digital Future Center. 2025. *Close Encounters of the AI Kind: Main Report.* Technical Report. Elon University. https://imaginingthedigitalfuture.org/reports-and-publications/close-encounters-of-the-ai-kind/close-encounters-of-the-ai-kind-main-report/ Accessed: 2026-01-21.

[9] Tero Jokela, Jarno Ojala, and Thomas Olsson. 2015. A Diary Study on Combining Multiple Information Devices in Everyday Activities and Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3903–3912. doi:10.1145/2702123.2702211

[10] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300641

[11] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3313831.3376590

[12] Helena Lindgren. 2025. Emerging Roles and Relationships Among Humans and Interactive AI Systems. *International Journal of Human–Computer Interaction* 41, 17 (2025), 10595–10617. arXiv:https://doi.org/10.1080/10447318.2024.2435693 doi:10.1080/10447318.2024.2435693

[13] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. doi:10.1145/2858036.2858288

[14] Jonas Oppenlaender, Rhema Linder, and Johanna Silvennoinen. 2025. Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. *International Journal of Human–Computer Interaction* 41, 16 (2025), 10207–10229. arXiv:https://doi.org/10.1080/10447318.2024.2431761 doi:10.1080/10447318.2024.2431761

[15] A. Sarkar, A. D. Gordon, C. Negreanu, J. Poetzsch-Heffter, S. S. Ragavan, and B. Zorn. 2024. CollabCoder: A Lower-barrier, Live-coding Environment for Learning to Code with AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA.

[16] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 18 pages. doi:10.1145/3586183.3606756

[17] Macy Takaffoli, Sijia Li, and Ville Mäkelä. 2024. Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1579–1593. doi:10.1145/3643834.3660720

[18] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. 2022. Trust in Human-AI Interaction: Scoping Out Models, Measures, and Methods. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 254, 7 pages. doi:10.1145/3491101.3519772

[19] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376301

[20] Hye Sun Yun and Timothy Bickmore. 2025. Online Health Information–Seeking in the Era of Large Language Models: Cross-Sectional Web-Based Survey Study. *J Med Internet Res* 27 (31 Mar 2025), e68560. doi:10.2196/68560

[21] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. doi:10.1145/3544548.3581388

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009