ReflectEd: Evaluating Reflection-Driven Learning in an AI-Assisted System

Md Nazmus Sakib(✉), Ishika Tarin, Naga Manogna Rayasam, Manas Gaur, and Sanorita Dey

University of Maryland Baltimore County, Baltimore, MD 21250, USA
{msakib1,qp45654,tp34657,manas,sanorita}@umbc.edu

Abstract. In collaborative settings, sustaining momentum between checkpoints (e.g., meetings) is challenging, often causing task drift and reduced preparedness. To address this, we developed *ReflectEd*, an AI-assisted system supporting between-checkpoint reflection via theory-driven prompts with progressive structure and mechanism-based scaffolding. We evaluated *ReflectEd* in a mixed-method study comparing a regular reflection workflow and a deeper workflow with an added transformative reflection activity. Across conditions, participants showed steady early engagement; the deeper condition yielded more actionable, forward-looking reflections but was harder to sustain and more effortful during active work. Partner-visible reflections supported coordination by surfacing differences and facilitating accountability. Overall, results highlight trade-offs between reflection depth, feasibility, and preparedness for subsequent checkpoints, with implications for designing AI-assisted systems that support collaboration readiness and reflection-oriented regulation in time-constrained workflows.

Keywords: AI-assisted reflection · Reflection-on-action · Collaborative learning · Self-regulated learning · LLM · Prompt Engineering

1 Introduction

Academic project work spans multiple days and requires sustained progress, adaptation, and self-regulation. Learning emerges through iterative sensemaking, strategy revision, and action. Reflection supports this process, but it is often difficult to sustain during active work because of competing progress demands [13]. Between checkpoints, asynchronous work and distant deadlines can reduce motivation and increase task drift, leaving learners underprepared without lightweight reflective support. Prior HCI research has explored meeting-centered supports such as summaries, dashboards, and reflective interventions to help groups revisit decisions and strengthen coordination [12]. Recent work also highlights AI-assisted goal reflection during meetings [3,10]. However, these systems largely focus on reflection during meetings and offer limited support for

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
E. G. Blanchard et al. (Eds.): AIED 2026, LNAI 16585, pp. 1–10, 2026.
https://doi.org/10.1007/978-3-032-29770-9_37

reflection-on-action between meetings, when learners must interpret progress, surface unfinished work, and decide next steps with little guidance [8,12]. Its value depends on supporting learning that carries into subsequent action, and it may be most effective when timed and scaffolded to fit ongoing work [2]. To examine these tradeoffs, we ask: **RQ: How do different levels of AI-scaffolded reflection between checkpoints affect learners' reflective engagement, collaboration experience, action planning, and preparedness for subsequent work?** Then, to answer this question, we designed *ReflectEd*, an AI-assisted system for structured reflection between two checkpoints in academic project workflows. After the first checkpoint, *ReflectEd* generates a meeting-grounded summary and delivers lightweight, personalized reflection prompts across the week. The design draws on reflection-on-action and prior HCI work showing that reflection varies in depth and that reflective technologies benefit from concrete scaffolding [2,4]. By distributing prompts across days and grounding them in recent task context, *ReflectEd* supports sensemaking and action planning during the between-meeting interval while keeping added effort low. This paper contributes (1) *ReflectEd*, an AI-assisted reflection system for academic workflows between checkpoints, (2) an empirical comparison of tradeoffs across two reflection structures, and (3) design implications for AIED on supporting reflection-on-action in time-constrained collaborative learning.

2 Related Works

2.1 Learning Through Reflection and Reflection-on-Action

Reflection is often described as a meta-cognitive activity of examining one's own thinking. According to Schön's Theory of Reflective Practice, reflection-in-action occurs during activity, while reflection-on-action occurs afterward [9]. Prior work shows that awareness during activity can shape later reflection-on-action, and that analyzing open-ended reflections can reveal knowledge-building over time [7]. AI-based reflective agents further improve reflection quality, including specificity and depth, while other works emphasize transparency, user control, recommendation strategies, and collaborative scaffolds for engagement, agency, self-regulated learning, structured feedback, and coordination [1]. Building on this direction, our work focuses on reflection-on-action in project-based learning contexts and examines structured reflection as a lightweight scaffold that sustains coordination and self-regulation between checkpoints in short, time-bounded collaborative tasks.

2.2 Intelligent Systems and Human-Centered AI in Collaborative Work

Intelligent systems have long been explored to support collaborative learning and teamwork by improving interaction processes and providing targeted assistance. Prior work frames reflection as requiring structured support, from depth-based scaffolding to design resources such as temporal framing, discovery, comparison,

and conversation [2], while highlighting gaps between theory and insufficiently structured reflective interfaces [4].

Our work operationalizes staged reflection with AI-assisted, temporally distributed scaffolding to sustain momentum, support team alignment, and enable preparation between checkpoints, positioning reflection as a designable component of collaborative learning systems.

3 Methodology

Theoretical Baseline of ReflectEd: *ReflectEd* draws on prior work on technology-supported reflection. Fleck et al. [5] frame reflection as varying in depth and strength, Bentvelzen et al. [2] identify mechanisms for reflective system design, and Cho et al. [4] argue that reflection is often treated as a design label rather than a scaffolded process. Because reflection is difficult to sustain in real workflows and repeated open-ended prompts can lead to shallow responses or disengagement, *ReflectEd* distributes structured reflection across days through level-based scaffolding and targeted guidance. We used two reflection conditions across two participant groups, both drawing on four reflection levels, Descriptive, Explanatory, Relational, and Transformative [5], and four reflection mechanisms, Temporal, Discovery, Comparison, and Conversation [2]. Since Transformative reflection supports deeper meaning-making and future action, it was excluded from the regular condition and included only on Day 6 in the deeper condition; prompts were otherwise delivered on alternate days to reduce cognitive burden and allow progress between reflections (Table 1 shows prompts of *ReflectEd*).

Study Design: Participants completed a seven-day collaborative poster design task to examine how structured reflection supports collaboration over time. The remote study prohibited LLM use. Pairs were formed by preference or random assignment. Each pair held two online meetings: one to choose a topic and divide responsibilities, and one to review and finalize the poster. Meetings were recommended to last no more than 40 min. We provided example topics, though pairs could choose their own. Between meetings, participants received *ReflectEd* notifications on alternate days and completed reflection activities within 24 h. In the deeper condition, participants could view their partner's Day 3 reflection on Day 5 to support comparison and coordination, while *ReflectEd* conversations remained private. The study focused on collaboration and preparedness rather than the quality of the final poster (Fig. 1).

Participants and Procedure: We recruited 20 students (12 male, 8 female; ages 20–34, $M = 26.40$, $SD = 3.87$) from undergraduate and graduate programs across U.S. universities using convenience and snowball sampling. All had prior experience completing collaborative academic tasks at least twice in academic or professional contexts. Participants were expected to spend about 10 h on the study and received $80. The study protocol, including AI/LLM-based data processing, was approved by the IRB. Participants were randomly assigned to two

Table 1. Reflection workflow (Regular vs. Deeper). Days share the same prompt set, except Day 6 which appears only in the Deeper condition.

Day	Level & Mechanism	Prompts & Purpose
Day 0 (CP1)	*First Checkpoint*	—
Day 1	**Descriptive**	*Comprehensive summary:* Recap the meeting and briefly report progress.
Day 2	—	—
Day 3	**Explanatory** (Discovery + Conversation)	*Explain and guide.* Discovery: (1) What was hard today and why? (2) What was unexpected? Conversation with *ReflectEd* (up to 3 turns): one targeted question and up to two follow-ups to guide next steps.
Day 4	—	—
Day 5	**Relational** (Comparison)	*Explore and connect.* (1) Better/same/worse than yesterday? (2) Why? (3) Compared to your partner, what did you focus on most today? *(Deeper RC only)* Supports self- and partner-comparison.
Day 6	**Transformative** (Temporal)	*Take action and change.* (1) What will you do differently next week? (2) What will you keep the same? (3) What task will you finish before the final checkpoint?
Day 7 (CP2)	*Second Checkpoint*	—

Progression: Descriptive → Explanatory → Relational → Transformative. Checkpoints (CPs) mark meetings for sharing updates.

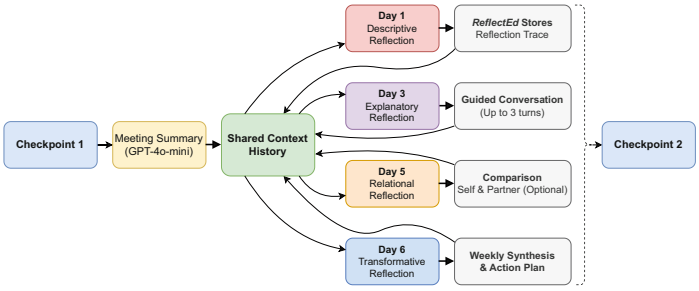
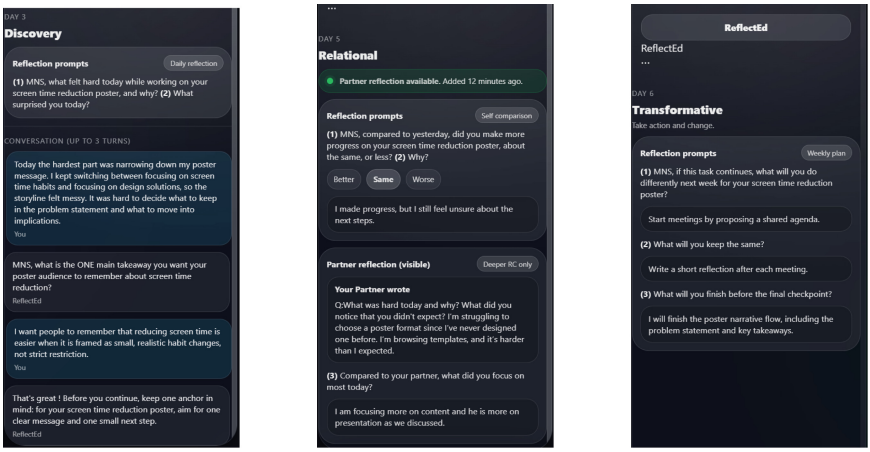


Fig. 1. Overview of the multi-day reflection workflow. Each checkpoint meeting is summarized by GPT-4o-mini and incrementally enriches a shared context history across descriptive, explanatory, relational, and transformative reflection stages. ***Note:*** Day 6 reflection is only applicable in the Deeper condition.

conditions, with five pairs per condition: Group 1 received the regular reflection workflow on alternate days between checkpoints (CPs), and Group 2 received the deeper workflow on alternate days plus an additional Day 6 reflection activity. Each checkpoint corresponded to a scheduled meeting.

System Development: *ReflectEd* is a web-based system for structured reflection across two checkpoints in a short collaborative workflow. It was built with Streamlit (front-end) and MongoDB (data storage). Meetings were conducted within *ReflectEd*, and transcripts were automatically captured and processed via GPT-4o-mini API calls to generate meeting summaries and extract key discussion points for later reflection. We used one-shot and few-shot prompting to

generate meeting summaries, ask short follow-up questions based on participants' latest reflections, and create the Day 6 summary from their earlier reflections and task progress. Example snapshots of *ReflectEd* are shown in Fig. 2.



(a) Day 3 (Regular and Deeper) (b) Day 5 (Regular and Deeper) (c) Day 6 (Deeper only)

Fig. 2. Example snapshots from *ReflectEd* for Day 3, Day 5, and Day 6 reflection activities.

ReflectEd runs a seven-day AI-assisted reflection workflow. Participants engage with the system daily between two checkpoints. At Checkpoint 1, they meet through *ReflectEd*, which generates an AI meeting summary and maintains shared context from reflections and task progress. The workflow moves from **Descriptive recap** (Day 1) to **Explanatory sensemaking** (Day 3), where *ReflectEd* supports a bounded three-turn conversation based on the latest entry using task-grounded prompts and limited follow-ups for clarification or next-action planning. On Day 5, *ReflectEd* supports **Relational reflection** through self-comparison across days and partner-comparison. On Day 6, it supports **Transformative reflection** by generating a weekly synthesis in the deeper condition and an actionable plan before Checkpoint 2. At Checkpoint 2, participants meet again to complete the poster task.

Evaluation: To evaluate *ReflectEd*, we combined self-reports with trace-based measures, including reflection completion time, and artifact-based measures from reflections and checkpoint meetings. After the seven-day workflow, participants completed a post-study survey assessing (1) collaboration preparedness for CP2 (preparedness, progress clarity, remaining tasks, and partner alignment), (2) reflection usefulness (sensemaking, identifying next actions, staying focused, and improving collaboration), and (3) system perception (ease of use, prompt relevance, structure, and workflow fit) on a 5-point Likert scale (Strongly Disagree–1

to Strongly Agree–5), along with open-ended feedback. In the survey, reflection was defined as looking back on progress, challenges, and next steps in relation to the ongoing collaborative task. We coded reflection traces using Kember et al.'s four-level framework [6] and compared shared days (Day 1/3/5), since only the Deeper condition included Day 6. To capture learning-to-action beyond critical reflection, we also coded actionability on a 0–2 scale indicating no next step, a vague intention, or a concrete task plan, and examined whether Day 6 plans were referenced in CP2.

For qualitative analysis, we analyzed open-ended survey responses on interface usability, reflection usefulness for planning and collaboration, partner visibility, CP2 experiences, and suggestions for improvement using inductive thematic analysis grounded in grounded theory methods [11]. The first and second authors, both experienced in qualitative HCI coding, conducted open coding, refined themes through discussion, achieved high initial agreement (Cohen's Kappa = 0.86), and resolved remaining differences collaboratively. Using Kember et al.'s framework, we also coded whether responses described events, showed understanding, reflected on experience, or critically reconsidered it.

4 Findings

4.1 Quantitative Analysis

Reflection Progression: We coded written reflections using a four-level scheme, with acceptable inter-rater reliability on a subset of the data (Cohen's $k = 0.84$); remaining disagreements were resolved through discussion. Reflection levels were similar across conditions on Day 1 (Regular: $M = 1.90$; Deeper: $M = 1.86$) and Day 3 (Regular: $M = 2.18$; Deeper: $M = 2.11$). By Day 5, the Deeper condition reached higher coded levels (Deeper: $M = 2.89$; Regular: $M = 2.20$). Across shared reflection days (Day 1, Day 3, Day 5), mean reflection levels were slightly higher in the deeper condition (Regular: $M = 2.09$, $SD = 0.41$; Deeper: $M = 2.25$, $SD = 0.44$).

Feasibility and Cognitive Burden: Across the seven-day workflow, the regular condition showed higher completion rates than the deeper condition. Most participants completed a larger proportion of assigned reflections in the regular condition (Regular: $M = 0.97$, $SD = 0.10$; Deeper: $M = 0.80$, $SD = 0.22$; Table 2), suggesting that greater reflection depth and the additional Day 6 activity were accompanied by lower adherence. The deeper condition also showed longer time-on-task per reflection (Regular: $M = 180$s, $SD = 32$s; Deeper: $M = 228$s, $SD = 46$s), suggesting higher effort during individual reflection activities.

Actionability of Reflection Outcomes: Actionability scores were similar across conditions on Day 3 (Regular: $M = 1.11$, $SD = 0.57$; Deeper: $M = 1.01$, $SD = 0.71$), indicating comparable next-step planning. By Day 5, the deeper condition showed slightly higher actionability scores (Regular: $M = 1.17$, $SD = 0.51$; Deeper: $M = 1.31$, $SD = 0.47$). In the Transformative reflection activity

(deeper condition only), several Day 6 reflections emphasized forward-looking planning, including intermediate deadlines, communication routines, and task allocation, and showed moderate actionability (Day 6: $M = 1.46$, $SD = 0.50$ on a 0–2 scale). Additionally, 82% of participants in the deeper condition mentioned at least one plan of action during CP2. Post-study surveys also showed higher self-reported preparedness for CP2 in the deeper condition (Regular: $M = 3.84$, $SD = 0.56$; Deeper: $M = 4.15$, $SD = 0.51$), while reflection usefulness (Regular: $M = 4.06$, $SD = 0.63$; Deeper: $M = 4.13$, $SD = 0.58$) and system perception (Regular: $M = 4.22$, $SD = 0.43$; Deeper: $M = 4.08$, $SD = 0.52$) were similar across conditions.

Table 2. Reflection progression (shared days: Day 1/3/5), actionability, and self-reported outcomes by condition. Values are mean (SD), except day-level reflection means which are reported as means.

Metric	Regular (n=10)	Deeper (n=10)
Completion rate (assigned reflections)	0.97 (0.10)	0.80 (0.22)
Time-on-task per reflection (sec)	180 (32)	228 (46)
Reflection level (Day1, mean)	1.90	1.86
Reflection level (Day3, mean)	2.18	2.11
Reflection level (Day5, mean)	2.20	2.89
Mean reflection level (shared days, 1–4)	2.09 (0.41)	2.25 (0.44)
Day 3 actionability (0–2)	1.11 (0.57)	1.01 (0.71)
Day 5 actionability (0–2)	1.17 (0.51)	1.31 (0.47)
Day 6 actionability (0–2)	–	1.46 (0.50)
Day 6 action carryover into CP2 (%)	–	82%
CP2 preparedness (self-report)	3.84 (0.56)	4.15 (0.51)
Reflection usefulness (self-report)	4.06 (0.63)	4.13 (0.58)
System perception (self-report)	4.22 (0.43)	4.08 (0.52)

Note: Reflection level coded using a four-level scheme (1–4). Actionability coded on a 0–2 scale. Self-reported outcomes used 5-point Likert scales (1–5). Day 6 metrics were collected in the deeper condition only.

4.2 Qualitative Analysis

Micro-checkpoints that Sustain Momentum: Across conditions, participants (n=13) described reflections as lightweight pacing mechanisms that reduced drift between meetings and made progress more explicit. Some (n=4) valued brief pauses to review completed work and identify a next step. One participant described these check-ins as micro-deadlines: *When I reach the check-in, it feels like a small deadline... I choose one next step and write it down... and I begin with less hesitation."* Some also suggested adding simple task-support

features, such as *a small checklist next to the prompt and a progress bar we both could see,"* to reduce coordination overhead and support shared awareness. Several noted that reminders became more useful as CP2 approached.

Feasible Reflection Requires Flexible Depth: Across conditions, participants valued reflection, but some (n=3) noted that deeper prompts on Day 6 could feel demanding on busy days. One participant explained: *On busy days, the deeper prompt (Day 6) felt a bit heavy, though it often led to useful reflection when I had the time."* This suggested a preference for mixed rhythms, where regular prompts supported continuity and deeper prompts supported sensemaking at key moments. As another participant noted: *The factual question (Day 3) helped me remember my progress and keep procrastination in check... That little conversation (Day 5) helped me see the difficulties... "*. These accounts point to the need to balance reflection depth with feasibility in time-constrained workflows.

Reflection-Driven Alignment Supports Readiness and Transfer: Participants in the deeper condition (n=5) described partner-visible reflections as a coordination resource that clarified roles and supported accountability, especially before CP2. Day 6 transformative reflection also helped turn weekly progress into concrete next steps before re-entering collaborative work. During CP2, this appeared to reduce catch-up time, which some regular-condition participants said slowed poster completion. More broadly, participants noted that differences in partners' reflections surfaced alternative strategies and that stating planned changes before CP2 helped them return with clearer intent.

5 Discussion and Future Work

Design Tradeoffs in Reflection Depth: Our findings suggest that reflection depth should be treated as a design parameter rather than a universal target. While both configurations helped sustain momentum between checkpoints, the deeper condition supported somewhat higher reflection levels, greater actionability later in the workflow, and higher preparedness for CP2, but was less feasible, as participants completed fewer reflections and spent more time per response. Prior work has largely examined reflection in synchronous meetings, retrospective reviews, or personal informatics contexts [4,8]. In contrast, our study focuses on the understudied interval between collaborative checkpoints, where learners must sustain momentum asynchronously, interpret partial progress, and prepare to re-engage with shared work. Qualitative findings suggest that regular prompts acted as "micro-checkpoints" that reduced drift and helped participants re-engage as deadlines approached, while deeper prompts were more useful near key moments. This points to a mixed-depth design: regular prompts for continuity and occasional deeper prompts for sensemaking and next-step planning. Participants also noted risks around visibility and impression management, suggesting that selective sharing may support coordination without making reflection feel performative [2,8].

Reflection as a Learning Scaffold: Rather than showing learning gains directly, our findings show how *ReflectEd* supports regulation processes tied to learning in open-ended collaborative work, where learners must monitor progress, identify stalls, and decide next steps with limited guidance [1]. Between-checkpoint reflection was both retrospective and forward-looking, helping participants track progress, surface unresolved issues, and articulate feasible next steps. This aligns with recent work on AI-assisted reflection in meetings, but shifts attention from synchronous goal reflection to asynchronous re-engagement between meetings [3,10]. Participants described *ReflectEd* as lightweight guidance that supported planning, comparison, and retrospection, especially near deadlines. At the same time, the value of reflection depended on workload and task demands, suggesting that fixed schedules may not suit all learners or contexts. Future systems may therefore need adaptive reflection timing and depth based on workload, milestone proximity, or learner readiness.

Limitations and Future Work: We evaluated *ReflectEd* in a short seven-day, two-checkpoint poster task, which limits generalization to longer, less structured projects, and instructor-mediated settings. Several outcomes relied on self-reports and coded reflections, capturing engagement, feasibility, actionability, and preparedness, but not longer-term learning or transfer. Future work should examine longer deployments, adaptive reflection schedules, and privacy-aware sharing that supports alignment without reducing autonomy or psychological safety. Future work should also include a no-support or non-AI control to better isolate the role of AI assistance.

References

1. Barria-Pineda, J., Sonmez Unal, D., Akhuseyinoglu, K., Brusilovsky, P., Walker,E.: Using self-regulated learning theory to inform the design of educational recommender systems for introductory programming. In: International Conference on Artificial Intelligence in Education. pp. 276–284. Springer (2025)
2. Bentvelzen, M., Woźniak, P.W., Herbes, P.S., Stefanidi, E., Niess, J.: Revisiting reflection in HCI: four design resources for technologies that support reflection. Proc. ACM Interact. Mobile Wearable Ubiquit. Technol. **6**(1), 1–27 (2022)
3. Chen, X., Tankelevitch, L., Vanukuru, R., Scott, A.E., Panda, P., Rintel, S.: Are we on track? AI-assisted active and passive goal reflection during meetings. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. pp. 1–22 (2025)
4. Cho, J., Xu, T., Zimmermann-Niefield, A., Voida, S.: Reflection in theory and reflection in practice: an exploration of the gaps in reflection support among personal informatics apps. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. pp. 1–23 (2022)
5. Fleck, R., Fitzpatrick, G.: Reflecting on reflection: framing a design landscape. In: Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction. pp. 216–223 (2010)

6. Kember, D., McKay, J., Sinclair, K., Wong, F.K.Y.: A four-category scheme for coding and assessing the level of reflection in written work. Assess. Eval. High. Educ. **33**(4), 369–379 (2008)
7. Li, G., Chen, L., Tang, C., Deguchi, D., Yamashita, T., Shimada, A.: From reflections to motifs: a graph-based analysis of learners' knowledge construction. In: International Conference on Artificial Intelligence in Education. pp. 299–307. Springer (2025)
8. Park, S., Kulkarni, C.: Retrospector: rapid collaborative reflection to improve collaborative practices. Proc. ACM Hum. Comput. Interact. **7**(CSCW2), 1–20 (2023)
9. Schon, D.A., DeSanctis, V.: The reflective practitioner: how professionals think in action. J. Contin. High. Educ. **34**(3), 29–30 (1986). https://doi.org/10.1080/07377366.1986.10401080
10. Scott, A.E., Tankelevitch, L., Rintel, S.: Mental models of meeting goals: supporting intentionality in meeting technologies. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. pp. 1–17 (2024)
11. Thornberg, R., Charmaz, K., et al.: Grounded theory and theoretical coding. SAGE Handb. Qual. Data Anal. **5**(2014), 153–169 (2014)
12. Wang, R., Qiu, L., Cranshaw, J., Zhang, A.X.: Meeting bridges: designing information artifacts that bridge from synchronous meetings to asynchronous collaboration. Proc. ACM Hum. Comput. Interact. **8**(CSCW1), 1–29 (2024)
13. Zhang, N., Yan, Y., Suzuki, R.: From following to understanding: Investigating the role of reflective prompts in ar-guided tasks to promote user understanding. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. pp. 1–18 (2025)