

Exploring a Design Space of Interventions That Promote Reflective Use of Conversational Generative AI Among University Students

by

Fairuz Shaiara

A thesis submitted to
The Faculty of Graduate and Postdoctoral Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

Master of Science

Department of Computer Science
The University of Manitoba
Winnipeg, Manitoba, Canada
August 2026

© Copyright 2026 by Fairuz Shaiara

Thesis advisor

Author

Andrea Bunt

Fairuz Shaiara

Exploring a Design Space of Interventions That Promote Reflective Use of Conversational Generative AI Among University Students

Abstract

The use of conversational Generative AI (GenAI), such as ChatGPT, in higher education raises concerns around cognitive offloading and academic integrity, particularly when students engage with AI-generated content with limited reflection. This thesis uses a Research Through Design (RtD) approach to explore how reflective support can be integrated into conversational GenAI interfaces for academic use. Through low-fidelity sketching, engagement with prior literature, and research team discussions, we formulate an initial design space around two dimensions: degree of externalization and level of friction. We instantiate this space through four contrasting design concepts and examine students' perceptions of them in a qualitative elicitation study with 20 undergraduate students from diverse disciplines. Findings show that no single intervention is universally preferred; rather, receptiveness is shaped by perceived reflective abilities, academic integrity concerns, external accountability, disciplinary context, task characteristics, and workflow constraints. Building on

these findings and prior literature, we extend the design space with the *type of motivation* as a potential third dimension and introduce a speculative mockup of a fifth design concept. Overall, this thesis contributes an initial design space, five contrasting design concepts, and empirical insights for designing context-sensitive reflective GenAI interventions in higher education.

Contents

Abstract	ii
Table of Contents	vii
Dedication	xi
Acknowledgments	xii
1 Introduction	1
1.1 Research Questions	3
1.2 Methodology and Approach	4
1.2.1 Initial Design Space Exploration and Formulation	4
1.2.2 Development of Design Concepts	4
1.2.3 Qualitative Study	5
1.3 Contributions	5
2 Background and Related Work	8
2.1 Challenges of GenAI Use in Higher Education	8
2.2 Design Interventions to Promote Reflective Use of GenAI	11
2.3 Summary	14

3	Design Space Exploration and Concept Development	16
3.1	Research Through Design Approach	17
3.2	Low-Fidelity Sketching for Design Exploration	18
3.3	Emergence of Design Dimensions	21
3.3.1	Degree of Externalization (Implicit - Explicit)	22
3.3.2	Level of Friction (Low - High)	23
3.4	Development of Design Concepts	24
3.4.1	Predict Before Reveal	25
3.4.2	Quick Check	27
3.4.3	Workload Balance Slider	27
3.4.4	Workflow Mapper	29
3.4.5	Video Prototyping	31
3.5	Summary	32
4	Qualitative Study	33
4.1	Study Objectives	33
4.2	Participants	34
4.3	Procedure	35
4.4	Data Analysis	36
4.5	Findings	37
4.5.1	Current Practices with GenAI	38
4.5.2	Overall Preferences for the Design Concepts	38

4.5.3	Perceptions Towards the Design Concepts	40
4.5.3.1	Perceived Metacognitive Skills Shape Receptiveness to Reflection-Oriented Interventions	40
4.5.3.2	Reflection on Ownership Is Often Motivated by Aca- demic Integrity Concerns Rather than Intrinsic Learn- ing Goals	42
4.5.3.3	Reflection Is Often Driven by External Accountabil- ity and Incentives Rather than Intrinsic Motivation .	43
4.5.3.4	Preferences for Degree of Externalization and Level of Friction Varied Across Academic Disciplines . . .	44
4.6	Discussion	47
4.7	Summary	51
5	Reflection on the Design Space	52
5.1	Extending the Design Space	53
5.1.1	Type of Motivation (Intrinsic - Extrinsic)	54
5.1.2	Reflection Credits: A Speculative Design Concept	56
5.1.3	Representation of the Extended Design Space	57
5.2	Summary	59
6	Conclusion	60
6.1	Contributions	61

6.2 Limitations and Future Work	62
Bibliography	85
Appendix A Low-Fidelity Sketches	86
Appendix B Social Media Recruitment Post	87
Appendix C Recruitment Poster	88
Appendix D Pre-Study Questionnaire	89
Appendix E Research Ethics Board Approval	90
Appendix F Baseline Survey	91
Appendix G Context Scenario	92
Appendix H Interview Guide Used After Each Video Demonstration	93
H.1 Predict Before Reveal	93
H.2 Quick Check	94
H.3 Workload Balance Slider	94
H.4 Workflow Mapper	95
Appendix I Interview Guide Used at the End of the Study	96

List of Figures

3.1	Selected examples from the broader set of low-fidelity sketches generated during design exploration phase.	20
3.2	Visual representation of the four design concepts mapped along the two design dimensions: degree of externalization (vertical axis) and level of friction (horizontal axis).	25
3.3	Screenshot of Predict Before Reveal. (A) The AI-generated response is covered by a masked overlay prompting the student to anticipate the answer. The response is revealed after the student clicks on the mask. .	26
3.4	Screenshots of Quick Check (A) The panel has three statements in the form of a checklist prompting students to indicate whether they understand, can evaluate, and can apply the AI's response, with a <i>Generate Weekly Report</i> button that summarizes the student's selections over time. (B) The prompt input field remains disabled until the user interacts with the Quick Check and saves the response.	28

-
- 3.5 Screenshots of Workload Balance Slider. (A) The panel consists of a slider that allows students to indicate their perceived contribution of work between themselves and the AI. After the student positions the slider, a contextual message appears at the bottom. The weekly report visualizes patterns of AI-driven, collaborative, and self-driven interactions. (B) Similar to *Quick Check*, the prompt input field remains disabled until the interaction is completed. 29
- 3.6 Screenshots of Workflow Mapper. (A) The toolbar contains different shaped and color-coded draggable sticky notes representing user contribution and AI contribution respectively, along with arrows for linking steps. (B) The map is where the student can visually organize and connect their contributions with AI-generated content across different stages of a task. The greater number of squared notes compared to rounded notes indicates a heavier AI contribution relative to the student. 30
- 4.1 Participants ranked the four design concepts from most to least preferred separately for three criteria: (A) workflow fit (to what extent the concept might be integrated into their typical GenAI use); (B) reflection on ownership (to what extent the concept would be able to influence on their sense of ownership over AI-assisted work); and (C) reflection on AI response (to what extent the concept would be able to prompt users to pause and reflect on AI-generated responses before using them). . . . 39

- 5.1 Screenshot of Reflection Credits. (A) The side panel shows the student's current point balance, the number of points used for prompting, and the number of points earned through reflection. (B) Students can earn additional points by completing a brief reflection explaining how they have used AI so far in the assignment. (C) Each prompt submitted to the system costs points, with the cost increasing as students submit more continuous prompts. 57
- 5.2 Extended representation of the design space with five concepts mapped across three dimensions: degree of externalization, level of friction, and type of motivation. PBR = Predict Before Reveal, QC = Quick Check, WBS = Workload Balance Slider, WM = Workflow Mapper, and RC = Reflection Credits. 58

*To my parents,
the first believers in every dream I dared to pursue.*

Acknowledgments

First and foremost, I am deeply grateful to Almighty Allah for giving me the strength, patience, and opportunities that made this journey possible.

I would like to express my sincere gratitude to my advisor, Dr. Andrea Bunt, for her guidance, encouragement, and unwavering support throughout my Master's journey. Her mentorship has helped me grow not only as a researcher, but also as a person. Whenever I felt stuck, uncertain, or overwhelmed, she was always there with thoughtful advice, reassurance, and a way forward. I am especially grateful for the trust and confidence she placed in me, the many opportunities she provided for me to learn and grow, and her generous financial support throughout my degree. I feel incredibly fortunate to have had her as my advisor.

I am also grateful to my committee members, Dr. Daniel Rea and Dr. Shaowei Wang, for their time, insightful feedback, and valuable perspectives, which helped strengthen both this thesis and my research.

I would also like to thank all of my HCI Lab mates. Your thoughtful feedback during our many lab presentations and discussions helped me refine my ideas and become a better researcher. Beyond research, I am equally grateful for the conversations, laughter, and moments we shared that made the challenging parts of graduate school much more enjoyable.

To my friends in Winnipeg and back home in Bangladesh, thank you for being there through both the good days and the difficult ones. Whether from nearby or thousands of kilometres away, your encouragement, conversations, and friendship

provided me with a sense of comfort and belonging throughout this journey.

Most importantly, I am endlessly grateful to my parents for their unconditional love, sacrifices, encouragement, and belief in me. Their support has been the foundation behind everything I have been able to accomplish. No words can fully express how much their presence and faith in me have meant throughout this journey.

Finally, I gratefully acknowledge the financial support I received through the University of Manitoba Graduate Fellowship (UMGF), the International Graduate Student Entrance Scholarship (IGSES), and the Faculty of Graduate and Postdoctoral Studies (FGPS) Research Completion Award. These awards provided invaluable support and made my Master's journey considerably smoother than I could have anticipated.

Chapter 1

Introduction

Conversational Generative AI (GenAI), such as ChatGPT and Gemini, has rapidly become embedded in university students' academic work. Students are increasingly using these systems for a range of tasks, including idea generation, summarization, and revision [10; 23; 68; 80; 83]. A 2025 survey indicates that GenAI use in post-secondary education is now ubiquitous, with 92% of students reporting use for at least some aspects of their academic work (up from 66% in 2024) [47].

Conversational GenAI introduces an interaction paradigm in which students can pose questions and issue directives through natural dialogue, receiving near-instant, polished, and often elaborate responses [33; 57; 58]. These characteristics make such tools particularly appealing in competitive, fast-paced academic environments [103; 106]. However, these same affordances can also negatively impact students' cognitive engagement with educational material. When responses are fluently ar-

ticulated and immediately available, students may shift from active sensemaking to passive acceptance, mistaking coherence for comprehension [4; 87]. This dynamic can foster an illusion of understanding, reducing opportunities for students to critically reflect on AI-generated content and to examine their sense of ownership over work produced with its assistance [1].

The use of conversational GenAI in education has introduced a complex tension between institutional guidelines and students’ actual practices. Universities are actively developing and refining policy frameworks to guide the responsible and ethical use of GenAI in academic contexts [74; 104]. However, factors inherent to academic environments (e.g., time pressure, heavy workloads, and performance demands) can lead students to prioritize efficiency over careful adherence to such policies [2; 39]. In parallel, a growing body of work is exploring how interaction design, in the form of interface-level interventions, can foster more reflective use of these systems [11; 61; 72; 93; 97]. Such interventions can help in shaping students’ behavior at the point of interaction and provide immediate scaffolding for reflective engagement [94]. While efforts to design such interventions are underway, complementary research is needed to identify and explore pertinent design dimensions. It is also critical to investigate factors that might influence students’ receptiveness to such interventions in academic settings.

In this thesis, we contribute to the above interaction design challenge by examining how reflective support could be embedded into conversational GenAI interfaces

for academic use. Rather than assuming that a single intervention would be appropriate across students, tasks, or contexts, we explore a broader space of possible interface-level interventions and investigate how university students perceive them in relation to their academic work.

1.1 Research Questions

To guide this exploration, my thesis seeks to answer the following research questions:

1. What potential design dimensions can guide the development and integration of reflection-oriented interventions into conversational GenAI systems for university students' academic use?
2. How can the proposed design dimensions be leveraged to generate reflection-oriented design concepts for conversational GenAI systems?
3. How do university students perceive different reflection-oriented design concepts that represent different combinations of the design dimensions of externalization and friction?
4. What factors shape students' receptiveness to such reflective interventions in conversational GenAI systems?

1.2 Methodology and Approach

To address our research questions, we used a Research Through Design (RtD) approach [115] to: i) develop an initial design space for reflective interventions in conversational GenAI systems, ii) create four contrasting design concepts from this design space, and iii) conduct a qualitative study in which we presented the concepts to university students as video prototypes and identified factors shaping receptiveness to the proposed interventions. The following sections give a summary of each of these thesis components.

1.2.1 Initial Design Space Exploration and Formulation

To explore how reflective support could be integrated into conversational GenAI interactions, I generated a broad set of low-fidelity sketches exploring possible interface-level mechanisms for prompting reflection during GenAI-assisted academic work. These sketches were discussed with my advisor through multiple research meetings, where we compared the ideas, identified recurring characteristics, and considered how they could be organized into a design space. This process led to two design dimensions: degree of externalization and level of friction.

1.2.2 Development of Design Concepts

We used the design space to develop four contrasting design concepts: *Predict Before Reveal*, *Quick Check*, *Workload Balance Slider*, and *Workflow Mapper*. Each

concept represented a different combination of externalization and friction and supported reflection in a distinct way. To present the concepts consistently, we created medium-fidelity prototypes within a simulated ChatGPT environment and developed short video demonstrations situated within an academic scenario. These video prototypes were used as probes to elicit students' perceptions of the proposed interventions.

1.2.3 Qualitative Study

We conducted a qualitative elicitation study with undergraduate students who had prior experience using GenAI for academic work to examine their perceptions of the proposed design concepts and the factors shaping their receptiveness to those concepts. Participants viewed the video prototypes, responded to semi-structured interview questions, and ranked the concepts according to different criteria. We analyzed the interview data using reflexive thematic analysis [18; 19], while treating the design dimensions as sensitizing concepts [15] that informed our interpretation of the data.

1.3 Contributions

This thesis makes the following contributions:

1. We present an initial design space for exploring how reflective support can be

integrated into conversational GenAI interactions in higher education settings.

The design space is organized around three dimensions: *degree of externalization*, *level of friction* and *type of motivation*.

2. We instantiate the design space through five contrasting design concepts: *Predict Before Reveal*, *Quick Check*, *Workload Balance Slider*, *Workflow Mapper* and *Reflection Credits*. These concepts demonstrate different ways of prompting students to reflect on AI-generated responses, their own understanding, and their contributions to AI-assisted work. The first four concepts formed the main design set examined in this thesis, while *Reflection Credits* is presented as a speculative mockup illustrating the extrinsic end of the type of motivation dimension.
3. We present initial insights into how undergraduate students perceive reflection-oriented design concepts through a qualitative elicitation study using the first four design concepts as our design probes. The findings identify perceived preferences, concerns, contextual factors and trade-offs associated with the proposed interventions and highlight the need for reflective GenAI interventions that are sensitive to students' academic contexts and motivations.

The remainder of this thesis is organized as follows: In chapter 2, we review background and prior work related to this thesis. In chapter 3, we present the design space exploration and concept development process, including the low-fidelity sketches, the two design dimensions, and the four final design concepts. In chapter

4, we describe the qualitative study and its findings. In chapter 5, we reflect on the design space by drawing on both study findings and prior literature, and discuss a potential extension to the original design space. In chapter 6, we conclude the thesis, discuss its limitations, and outline directions for future work.

Chapter 2

Background and Related Work

In this chapter, we first review literature on the challenges introduced by conversational GenAI use in higher education, particularly focusing on cognitive offloading and academic integrity, while also situating these concerns alongside related issues of misinformation, hallucination, and bias. In response to these concerns, we then review approaches to designing interventions that promote more reflective use of conversational GenAI across educational and broader HCI contexts.

2.1 Challenges of GenAI Use in Higher Education

Conversational GenAI systems offer instant, fluent, and often elaborate responses to a wide range of academic queries, making them highly appealing to students [43]. In higher education contexts, this appeal is further amplified by factors such as heavy

workloads, time pressure, and performance incentives (e.g., grades) [106]. These conditions can shift students' priorities from meaningful learning toward efficiency and task completion [1].

The extensive use of GenAI in academic settings has introduced or amplified several challenges, one of which is cognitive offloading [84]. Cognitive offloading occurs when students delegate the mental effort required for learning to GenAI systems. This reliance reduces opportunities for critical thinking and can lead to outcomes such as diminished problem-solving abilities [63], more superficial engagement with educational material [7], and a reduced capacity for independent thinking [110]. Longer-term impacts may include poor knowledge retention [9], erosion of higher-order thinking skills [28], and diminished creativity [76]. These findings motivate research on approaches that encourage more critical thinking when students use GenAI systems.

Another widely discussed challenge is maintaining academic integrity in the face of prevalent GenAI use [81]. Violations of academic integrity occur when students use GenAI in unauthorized ways, such as submitting work largely generated by these systems and presenting it as their own [60; 105]. Such behaviour can be intentional, for instance, when students seek to avoid engaging in reflective thinking and rely on GenAI to complete their work [2]. However, violations are not always deliberate, as students may lack clarity about when their use of these tools crosses the boundary of acceptable academic practice [82]. Regardless of intent, these practices

can undermine learning by reducing opportunities to develop core skills such as writing and problem-solving [96]. Beyond individual learning outcomes, they can also affect the credibility of grades and degrees when assessments no longer accurately reflect students' mastery of course material [37]. These concerns underscore the importance of supporting students in reflecting on their own intellectual contributions when working with GenAI.

A further challenge concerns the reliability and fairness of GenAI outputs. Conversational GenAI systems can produce inaccurate, misleading, or fabricated information, often referred to as hallucinations, while presenting such outputs in a fluent and authoritative manner [3; 6; 14; 41]. These systems may also reproduce or amplify biases present in their training data, leading to outputs that reflect stereotypical, exclusionary, or otherwise problematic assumptions [88; 100; 112]. In academic contexts, this is particularly problematic because students may struggle to evaluate generated content, especially when engaging with difficult or unfamiliar subject matter or working under time pressure [35; 48; 113].

Prior work has explored a range of interventions for addressing concerns such as misinformation, hallucination, and bias in GenAI outputs and to support users in recognizing and evaluating such outputs. GenAI literacy and bias-focused tools have been designed to surface model defaults and biases [36; 71]. Other interventions support hallucination awareness through confidence indicators, source links, model comparison, repeated questioning, and output verification [66; 67; 98]. Automated

approaches have also been developed to detect, attribute, and revise unsupported outputs through evidence retrieval and grounding [49; 50]. In contrast, less attention has been given to interface-level interventions that address whether students remain cognitively engaged and maintain a meaningful intellectual contribution when using GenAI, even when the generated output is accurate. In this thesis, we therefore focus on this gap by investigating how interface-level design interventions can encourage university students to think more reflectively with GenAI while preserving their own intellectual contribution.

2.2 Design Interventions to Promote Reflective Use of GenAI

Researchers have begun to explore various ways of supporting reflective engagement with conversational GenAI in educational contexts. One prominent approach involves using text-based prompts, hints, or scaffolded feedback to keep students actively involved in the learning process and support their critical thinking [11; 32; 73; 93]. For example, Chia and Frattarola’s Waai system used pre-engineered prompts tailored to course assignments to guide students toward relevant research topics rather than direct answers, with findings showing that this supported perspective-taking and idea generation during writing tasks [32]. Similarly, Singh et al. used metacognitive prompts that encouraged students to explore more topics, pursue

deeper follow-up inquiry, evaluate AI responses, and consider overlooked perspectives [93]. Other work has designed conversational GenAI systems to provide instructional support during the interaction, for example by providing teacher-designed hints instead of unrestricted answers [11] or sentence-framing response strategies [73] that helped learners construct more substantive responses and engage in critical thinking. Together, this body of work shows how text-based scaffolding can support more active engagement with GenAI. Our work extends this line of research by exploring more interactive forms of reflective support, including checklists and mapping tools, that aim to make students' reasoning and intellectual contributions more explicit.

A second set of interventions explores introducing friction-based mechanisms into learners' interactions with GenAI, particularly by shaping when and how users can access AI-generated answers. To support self-directed learning among children aged 7–13, Xie et al. propose hypothetical design features that scaffold different phases of child-AI interaction, including mechanisms that block direct answers and instead encourages hints, strategies, delayed reveal, or prompts for learners to explain what they have tried before receiving a solution [108]. Related work in programming education introduces friction through structured reflection and cognitive engagement [42; 61]. For example, Fang et al. conducted a quasi-experimental study with senior high school students comparing a reflective GenAI-supported debugging approach with a conventional GenAI-supported debugging approach [42]. The reflective approach prompted students to consider the causes, corrections, and prevention of program-

ming errors after completing a coding task, and was shown to improve programming learning performance compared with the conventional approach. Kazemitabaar et al. explore a design space of cognitive engagement techniques such as asking learners to explain code before using it, trace and predict code execution, or engage in a lead-and-reveal interaction in which the AI guides learners through the problem-solving process before revealing the corresponding code [61]. Building on these friction-based approaches, we explore designs that embed reflective activities directly into the interaction flow, either before students access or after they receive AI-generated responses. We further investigate how students perceive such forms of friction in relation to their everyday academic workflows.

Beyond educational contexts, prior HCI work has explored how conversational GenAI interfaces can be redesigned to elicit greater user reflection and support users' own thinking processes [72; 95; 97]. For example, Tankelevitch et al. propose instructional scaffolds embedded within systems such as ChatGPT, Microsoft Copilot, and GitHub Copilot [97]. These scaffolds take the form of embedded suggestions and guiding questions that support metacognitive processes such as planning, self-monitoring, and self-evaluation (e.g., prompting users to specify aspects of their task, such as content, tone, and length, before generating a response). Maier et al. examine different interaction modes in which the AI prompts users with questions or suggestions for idea generation rather than taking over the creative process [72]. Their experimental findings show that question-based interaction encouraged deeper

elaboration and helped preserve users’ perceived ownership of their ideas. Sensecape takes a different approach by replacing the standard linear conversational format of a GenAI interface with an interactive 3D spatial interface that supports more structured exploration and sensemaking of ideas than conventional chat-based interaction [95]. Drawing on Sensecape’s use of spatial organization, one of our design concepts uses structured mapping to support students in reflecting on ownership of AI-assisted work.

2.3 Summary

In this chapter, we reviewed key challenges associated with conversational GenAI use in higher education, focusing on cognitive offloading and academic integrity while situating them alongside concerns about misinformation, hallucination, and bias. We then examined prior interventions that support more reflective GenAI use through text-based scaffolds, friction-based mechanisms, metacognitive support, and alternative interaction structures. This literature highlights the need to support students’ reflective engagement with conversational GenAI and demonstrates several ways of promoting more reflective use. However, these efforts also point to opportunities for complementary design research that explores a broader range of interface-level interventions and examines how students perceive such interventions in academic contexts. Building on this opportunity, our thesis develops an initial design space for reflection-oriented GenAI interventions, instantiates it through multiple interface-

level design concepts, and examines how university students perceive these concepts in relation to their everyday workflows, motivations, workload, and concerns about ownership and academic integrity.

Chapter 3

Design Space Exploration and Concept Development

In this chapter, we describe how we explored the design space and developed concepts to understand how reflective support could be integrated into conversational GenAI systems for academic use. Rather than beginning with a predefined intervention, our work adopted a Research Through Design (RtD) approach [115] to support open-ended exploration of possible design directions. Through low-fidelity sketching, review of relevant literature, and research team discussions, we generated and examined a range of early interface ideas aimed at encouraging reflection during GenAI-assisted academic work. This process helped us identify recurring design characteristics, which we used for the development of a structured design space. The chapter traces this progression through three stages: early sketch-based exploration,

identification of key design dimensions, and development of four final design concepts.

3.1 Research Through Design Approach

To explore potential ways of encouraging reflective use of conversational GenAI in academic settings, our work adopted a Research Through Design (RtD) approach [115]. RtD uses the creation and refinement of design artifacts as a means of investigating complex problems and generating research insights. Rather than treating design solely as an outcome, RtD considers the design process itself as a method for understanding a problem space and identifying opportunities for intervention. This approach has been widely used in Human-Computer Interaction (HCI) research to explore emerging problems and design spaces where requirements and solutions are not yet well defined [114].

RtD was appropriate for our work because the goal was not to optimize a single interface design or evaluate a predefined intervention. Instead, the objective was to explore how different forms of reflective support could be integrated into students' interactions with conversational GenAI systems. Students' use of conversational GenAI in academic settings is still evolving, and existing work provides limited guidance on how reflective support should be designed for these interactions. In particular, further work is needed to understand what design characteristics may shape students' perceptions of reflective interventions and their willingness to engage with them.

Following an RtD process, we began by generating a broad set of preliminary low-

fidelity sketches to explore possible interface interventions for supporting reflection during GenAI interactions. These sketches were discussed with my advisor through multiple research meetings, where we compared the ideas, examined recurring characteristics across them, and considered how they could be organized into a design space. While this sketching and discussion process was ongoing, we also reviewed prior literature on students' use of GenAI and concerns associated with AI-assisted academic work [1; 2; 43; 80; 81; 82; 84; 105; 106], as well as literature related to reflection [25; 55; 85], metacognition [45; 64; 70; 97; 107], self-regulated learning [65; 78; 92; 109], and reflective interventions [11; 40; 59; 61; 72; 93; 111]. This literature helped us better understand the problem space and identify why reflective support may be relevant for students' use of conversational GenAI.

3.2 Low-Fidelity Sketching for Design Exploration

To begin the design exploration process, we conducted open-ended brainstorming focused on generating a wide range of interface ideas and possible intervention strategies. To support this exploratory process, we used low-fidelity sketching as a way of rapidly generating and communicating ideas. Prior work suggests that sketching plays an important role during early stages of design by enabling ideas to be explored quickly and inexpensively while avoiding premature commitment to specific solutions [24]. Sketches are intended to be plentiful and disposable, serving as a means for generating and exploring ideas rather than refining or validating them. Following

this approach, we generated 16 low-fidelity sketches exploring different possibilities for integrating reflective support into conversational GenAI interactions. The goal at this stage was not to identify a preferred solution or develop polished concepts, but rather to generate a broad set of possibilities and investigate many alternative directions.

We explored substantial variation across the sketches in both form and interaction approach. In some sketches, we embedded lightweight prompts within conversational interactions. In others, we examined more explicit activities such as self-assessment tasks, visual representations of user contributions, and reflective activities placed at different points in the interaction process. We also explored different ways in which reflection might be integrated, including before users accessed AI-generated responses, after they had viewed AI output, or through mechanisms intended to support awareness of patterns of AI use and personal contribution over time.

We include a subset of the generated sketches in Figure 3.1 to illustrate the exploratory phase of the design process (the complete set of sketches is provided in Appendix A). Rather than converging early toward a small number of ideas, the purpose of this process was to encourage breadth of exploration and investigate a diverse range of possible intervention strategies. Through research team discussions, we identified recurring characteristics across the sketches that later helped us to develop the design dimensions discussed in the following section.

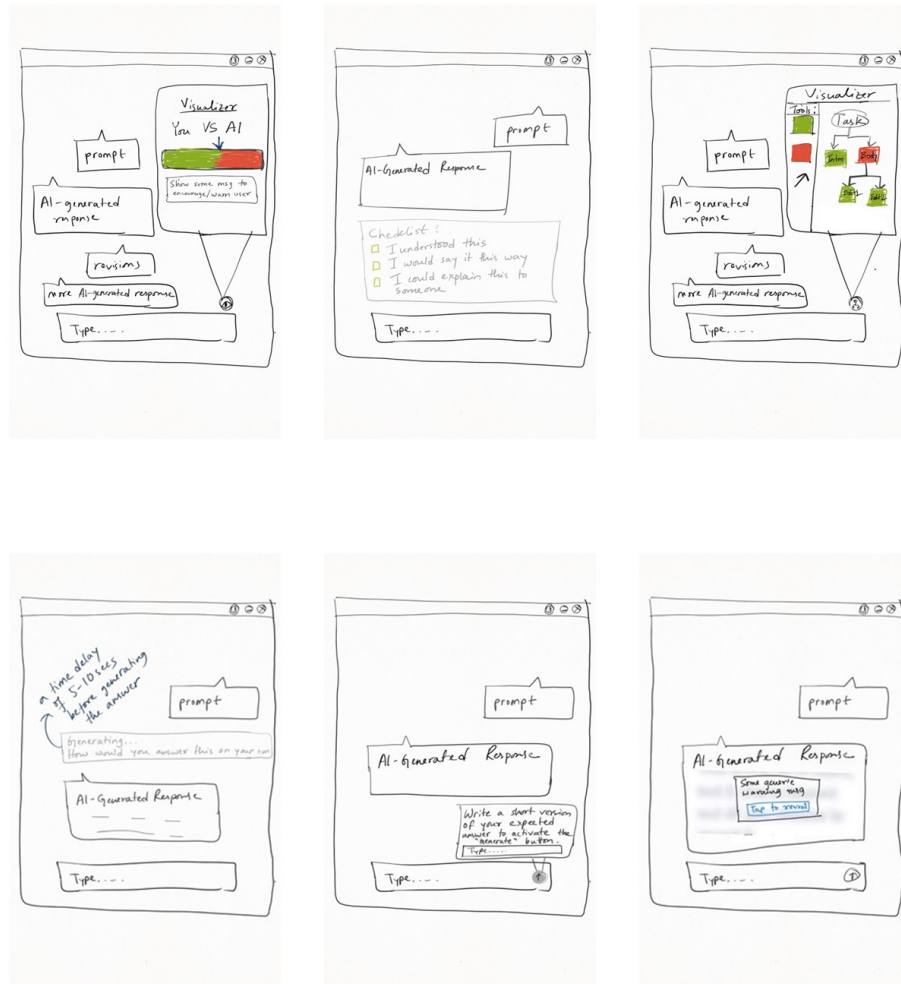


Figure 3.1: Selected examples from the broader set of low-fidelity sketches generated during design exploration phase.

3.3 Emergence of Design Dimensions

As we progressed with our sketching phase, we began identifying broader patterns across the explored concepts. Through research team discussions and comparisons across sketches, we identified recurring characteristics that consistently differentiated concepts from one another. For example, some of the concepts required students to make their thinking process more visible compared to others. To further refine our understanding of these characteristics, we drew on Tankelevitch et al.’s metacognitive demands framework [97], which helped to explain why users may need support in monitoring and regulating their own thinking when working with GenAI. Along with that, we also revisited relevant literature that explored mechanisms such as self-explanation [34], structured prompts [72; 93], visual representations [79], lightweight scaffolding [8; 30], and cognitive forcing mechanisms that interrupt habitual interaction patterns to encourage more deliberate thinking [21; 61]. We iteratively compared the sketches and considered several candidate characteristics, such as, how explicitly students were asked to represent their thinking, when reflection should occur, or how strongly an intervention disrupted the interaction flow. Through research team discussions and repeated comparison against the sketches, we refined these characteristics to two dimensions that most clearly differentiated the concepts: *degree of externalization* and *level of friction*. Together, these perspectives helped us interpret the recurring patterns as broader dimensions for organizing a design space. We view these dimensions as a promising initial set for organizing and exploring

reflection-oriented interventions in conversational GenAI rather than an exhaustive design space.

3.3.1 Degree of Externalization (Implicit - Explicit)

This dimension captures the extent an intervention requires students to externalize their thinking. This dimension can be understood as one way of supporting metacognitive monitoring, illustrated in the metacognitive demands framework [97], by making students' judgments and thought processes visible to different degrees. At the implicit end, reflection remains largely internal or is expressed through minimal interface actions. For example, mentally predicting an answer before revealing the AI response involves limited externalization, as students may think through their response without producing any external artifact. Prior work in self-regulated learning highlights that self-explanation, even with limited externalization, can strengthen metacognitive skills as well as critical thinking abilities [16; 31; 101]. At the explicit end, students generate external artifacts, such as workflow maps or written notes, that more directly capture their thought processes. Research on visual mapping suggests that such representations support learners in exploring, analyzing, and synthesizing ideas, thereby strengthening comprehension and critical thinking [79; 90]. Between these extremes lies a moderate range, in which interventions prompt structured but lightweight actions, such as adjusting a slider or completing a short checklist, offering more explicit articulation than internal reflection but less than full

external representations.

3.3.2 Level of Friction (Low - High)

This dimension captures the extent to which the intervention imposes constraints that disrupt a user’s typical interaction flow with a conversational GenAI system, in which the user enters a prompt, receives a response, and proceeds to the next prompt. By interrupting the interaction flow, friction can create opportunities for metacognitive control [97], requiring students to pause, predict, evaluate, or adjust before continuing. At the low-friction end, the intervention preserves this natural workflow. Students can immediately view the AI’s response and continue interacting without restriction, while optionally engaging with the intervention alongside the conversational workflow. Prior work on scaffolding suggests that such minimally disruptive support can encourage reflection without overburdening users [8; 30]. At moderate levels of friction, the system introduces structured interruptions after the AI output is presented. The response remains visible, but students must complete a brief reflective activity, such as a checklist or slider, before proceeding to the next prompt. These forms of structured friction align with prior work on “cognitive forcing functions”, which are designed to interrupt users’ default, heuristic responses at the moment of decision-making and prompt more deliberate, analytical thinking [21]. At the high-friction end, the system intervenes earlier by delaying or withholding the AI’s response until the student engages with the intervention. The distinction

between moderate and high friction lies in the timing of the interruption, whether it occurs before access to the AI’s response or after it has been presented. In our design framing, interventions that restrict access to the AI output are considered higher friction than those that restrict progression to the next prompt.

3.4 Development of Design Concepts

We used the two design dimensions to guide the development and refinement of four contrasting design concepts. We refined the concepts to improve usability, reduce unnecessary complexity, and better align with plausible academic GenAI-use scenarios identified through prior literature and research team discussions. We simplified some concepts from their initial sketches and integrated elements from multiple early ideas into others. Our aim was to explore meaningful combinations along the two design dimensions while avoiding extremes that would have high likelihood of failure either through discouraging use or offering little reflection support.

In academic settings where students often face heavy workloads and time pressure [26; 86; 99], concepts combining high friction and explicit forms of externalization risk being rejected outright. Conversely, designs with very low friction and implicit externalization would resemble standard GenAI interaction and provide limited support for reflection. We therefore selected combinations that are both contrastive and plausible for academic use. In Figure 3.2, we show the approximate placement of the four design concepts within the design space. In the following sections, we describe

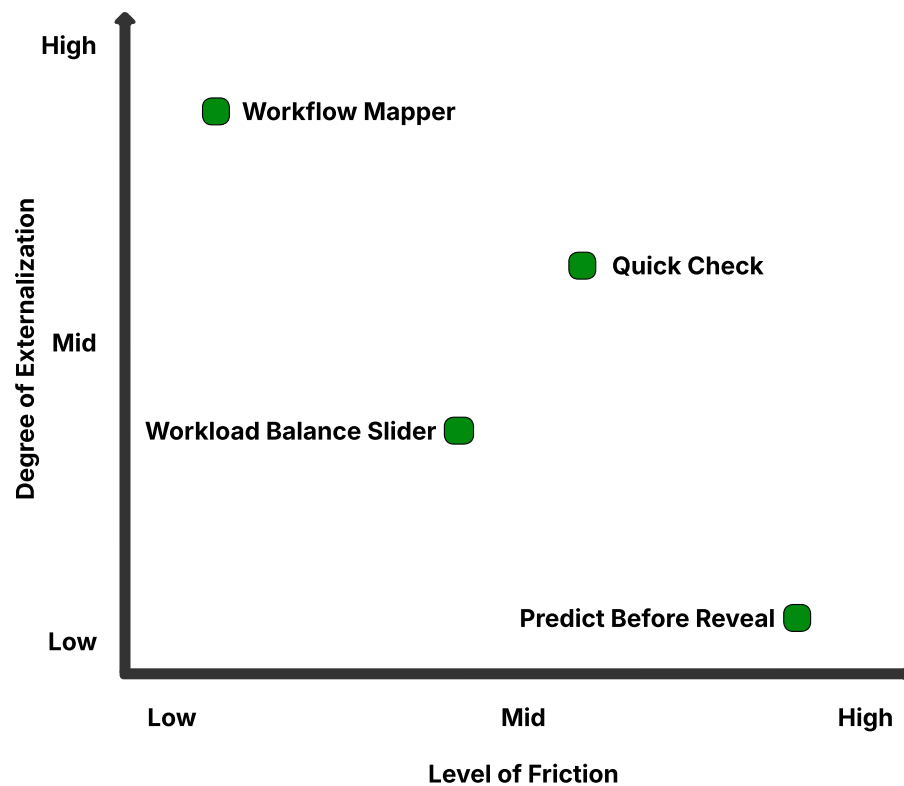


Figure 3.2: Visual representation of the four design concepts mapped along the two design dimensions: degree of externalization (vertical axis) and level of friction (horizontal axis).

each concept in detail and our rationale behind the placement of each concept along the two dimensions.

3.4.1 Predict Before Reveal

This design concept encourages students to reflect on the AI's response before viewing it. When a student enters a new prompt, the AI response remains hidden beneath a masked overlay containing a reflective cue to encourage the student to think

about the AI's response. Clicking the mask reveals the response (Figure 3.3A). This

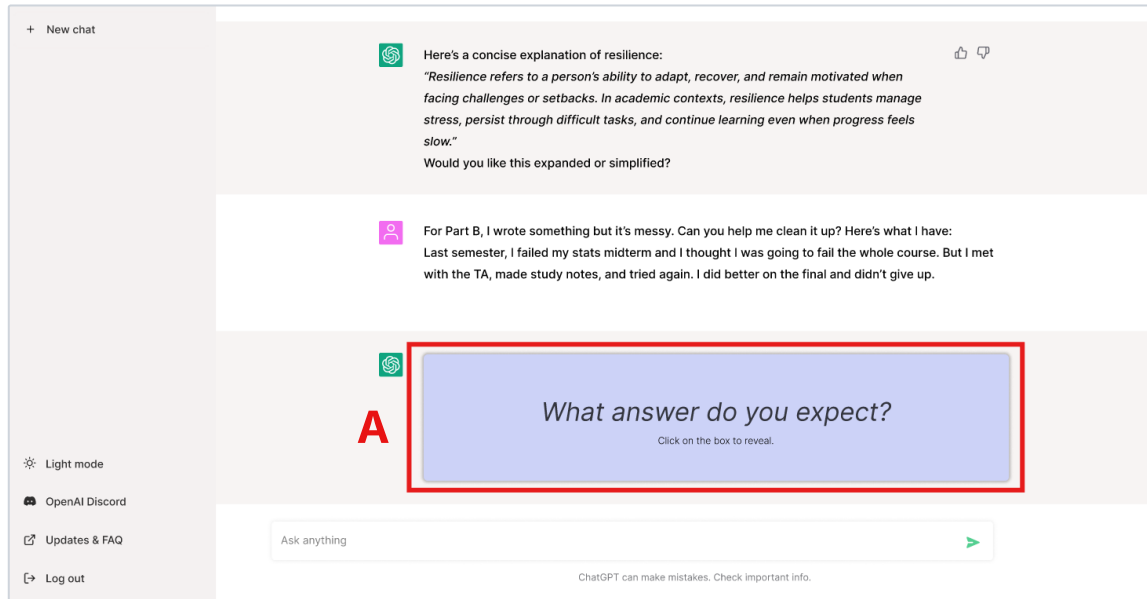


Figure 3.3: Screenshot of Predict Before Reveal. (A) The AI-generated response is covered by a masked overlay prompting the student to anticipate the answer. The response is revealed after the student clicks on the mask.

concept is motivated by prior findings suggesting that temporarily masking system outputs can act as minimal scaffolding by creating a brief interruption that prompts spontaneous self-explanation and as a result promote deeper critical thinking [34]. This reflection remains largely internal, as it does not require students to produce any external artifact beyond a simple click, placing it toward the implicit end of the externalization. On the other hand, the prototype interrupts the typical workflow by withholding the response from the student and thus introduces high friction.

3.4.2 Quick Check

This design concept prompts students to explicitly reflect on their engagement with the AI's response through structured self-assessment. After each AI response, *Quick Check* prompts students to reflect on their understanding of the response via a checklist (Figure 3.4A), whose items were inspired by Bloom's Taxonomy [62]. We selected three of the six cognitive processes outlined in Bloom's Taxonomy for learning to serve as practical indicators of whether the student felt they were meaningfully engaging with the AI's response. The student is blocked from entering another prompt until they complete the checklist (Figure 3.4B). *Quick Check* provides access to weekly summary reports to enable students to reflect more holistically on their learning patterns with AI responses. Within the design space, we position *Quick Check* as involving a mid-to-high degree of externalization, because it requires students to articulate their judgments through structured inputs. The concept creates moderate friction by interrupting progression to the next prompt, but without withholding access to the current AI response.

3.4.3 Workload Balance Slider

This design concept focuses on supporting students in identifying and reflecting on their own contributions to AI-assisted work. After each AI-generated response, students use a slider to indicate their level of contribution relative to that of the AI (Figure 3.5). Based on their selection, the system provides brief feedback, for

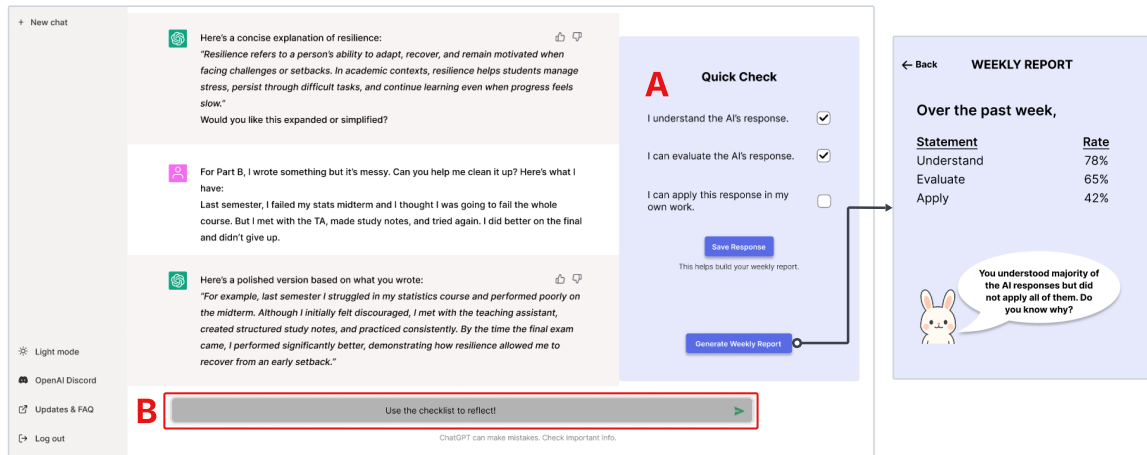


Figure 3.4: Screenshots of Quick Check (A) The panel has three statements in the form of a checklist prompting students to indicate whether they understand, can evaluate, and can apply the AI's response, with a *Generate Weekly Report* button that summarizes the student's selections over time. (B) The prompt input field remains disabled until the user interacts with the Quick Check and saves the response.

example, when a higher AI contribution is indicated, students are prompted to consider how they might contribute more to the task. The concept also includes weekly summary reports that visualize the distribution of contributions between the student and the AI over time. Similar to *Quick Check*, students must complete this self-assessment before entering a new prompt, placing the design within a moderate level of friction. However, compared to *Quick Check*, which requires evaluating and selecting multiple checklist items, *Workload Balance Slider* involves a slightly lower degree of externalization.

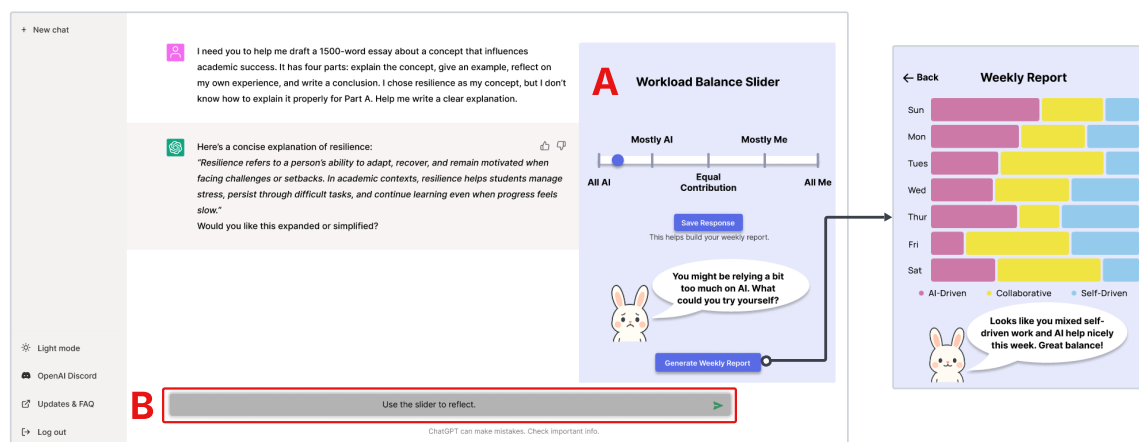


Figure 3.5: Screenshots of Workload Balance Slider. (A) The panel consists of a slider that allows students to indicate their perceived contribution of work between themselves and the AI. After the student positions the slider, a contextual message appears at the bottom. The weekly report visualizes patterns of AI-driven, collaborative, and self-driven interactions. (B) Similar to *Quick Check*, the prompt input field remains disabled until the interaction is completed.

3.4.4 Workflow Mapper

This design concept encourages students to visually map out how their contributions and the AI's contributions are distributed across a task. Inspired by prior work on mapping-based interfaces for supporting complex thinking processes [95], this concept adapts the idea of structured visual mapping to support reflection on ownership in AI-assisted work. Students use simple visual elements (e.g., color-coded sticky notes and arrows) to externalize how the task unfolds and how responsibility is distributed (Figure 3.6). Within the design space, we position *Workflow Mapper* toward the explicit end of the externalization, as it requires students to actively con-

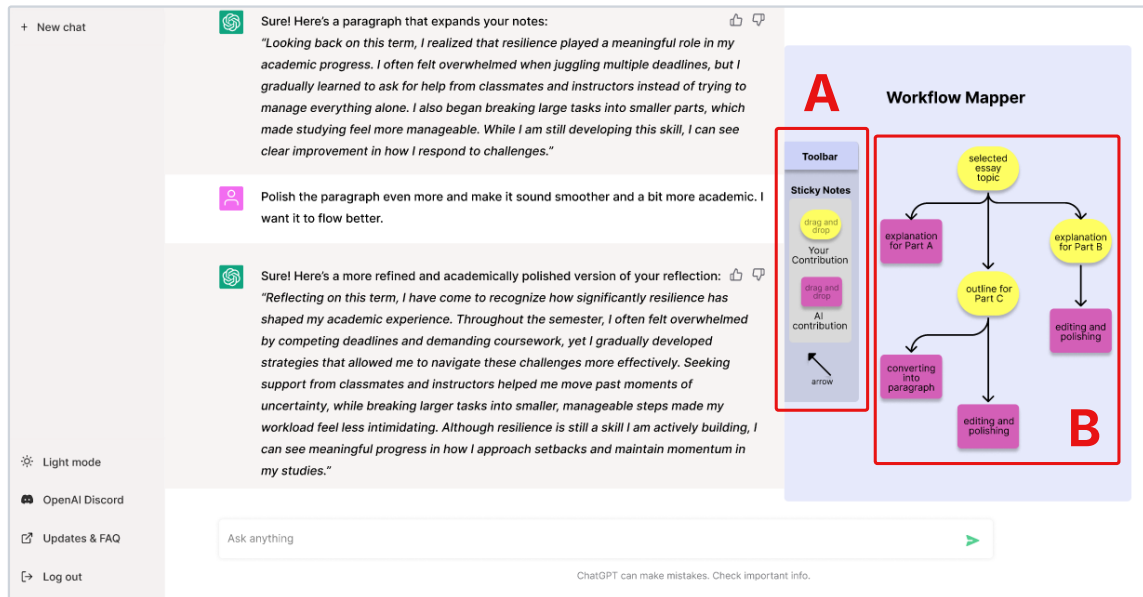


Figure 3.6: Screenshots of Workflow Mapper. (A) The toolbar contains different shaped and color-coded draggable sticky notes representing user contribution and AI contribution respectively, along with arrows for linking steps. (B) The map is where the student can visually organize and connect their contributions with AI-generated content across different stages of a task. The greater number of squared notes compared to rounded notes indicates a heavier AI contribution relative to the student.

struct a representation of their process. Unlike the previous concepts, interaction with the AI is not interrupted, enabling reflection to occur in parallel with the task. Accordingly, we place the concept toward a low level of friction. Similar to *Workload Balance Slider*, this concept focuses on prompting reflection on ownership rather than AI-generated content, but requires a much higher degree of externalization.

3.4.5 Video Prototyping

Given the exploratory nature of our work and the novelty of the design space, we adopted video prototyping as a way of representing and comparing alternative design visions. Video prototypes provide flexibility for exploring different interaction possibilities while avoiding the need to implement fully functional systems at an early stage [116]. For each concept, we created a short video prototype illustrating how the intervention would function within an academic scenario. To provide a consistent context across concepts, all videos followed a common academic scenario in which a student uses a conversational GenAI system to complete a 1500-word essay assignment under time pressure. Using a shared scenario allowed participants in our subsequent study to focus on differences between the reflective interventions rather than differences in task context, while also helping them understand how the concepts would be integrated into a realistic workflow.

The use of video prototypes enabled us to compare multiple concepts in a practical and rigorous manner while reducing the likelihood that participants' perceptions would be influenced by implementation-specific issues or interface limitations [22]. Prior work suggests that video prototyping can effectively communicate envisioned interactions and support the evaluation of proposed experiences before implementation [5; 52; 102].

We developed the prototypes using Figma [46] within a simulated conversational GenAI interface that resembles existing chat-based AI systems. We then created

video demonstrations of each prototype using Microsoft Clipchamp [75], which later served as stimuli for our elicitation study described in the following chapter.

3.5 Summary

In this chapter, we described the design exploration and concept development process used to investigate reflective support within conversational GenAI systems. Using a Research Through Design (RtD) approach, we reviewed relevant literature to understand concerns surrounding students' academic use of GenAI and used low-fidelity sketching to explore diverse intervention ideas. Through research team discussions and analysis of prior literature, we identified two recurring design dimensions: degree of externalization and level of friction. These dimensions helped us to organize the design space and guide the development of four contrasting design concepts. We then created medium-fidelity prototypes and video demonstrations to illustrate these concepts within an academic scenario, which formed the basis for the qualitative study described in the following chapter.

Chapter 4

Qualitative Study

Following the development of the design concepts, we conducted a qualitative elicitation study using video prototypes of the four design concepts. The goal of the study was not to identify a single best design, but to gather in-depth insights into participants' perceptions of the concepts. Specifically, we sought to understand whether participants could envision integrating these designs into their own academic workflows, as well as the factors that would influence their willingness to adopt them.

4.1 Study Objectives

The purpose of the elicitation study was to explore students' perceptions of the proposed reflective design concepts and understand how different forms of reflective support might be received within conversational GenAI interactions. Since the

concepts were intended as exploratory design probes rather than fully implemented systems, the focus of our study was to understand participants' interpretations, reactions, and perceived implications of the interventions rather than evaluating their effectiveness.

Specifically, we designed the study with the following objectives:

1. To understand how students perceive different approaches for integrating reflective support into conversational GenAI interactions;
2. To identify perceived benefits, concerns, and potential barriers associated with each design concept;
3. To explore how variations across the two design dimensions influence participants' preferences and perceptions; and
4. To gather insights that could inform the refinement of the concepts and contribute to a broader understanding of designing reflective interventions for GenAI-supported academic work.

4.2 Participants

We recruited 20 undergraduate students (8 men, 12 women) from our university through social media channels (e.g., Instagram and Slack) (see Appendix B), advertisements posted across campus (see Appendix C), and snowball sampling [53].

Participants represented a range of academic disciplines, including Computer Science/Engineering (6), Science (5), Arts (6), and Business (3). We intentionally recruited students from diverse disciplines to capture varied patterns of GenAI use and to examine whether differences in disciplines influenced receptiveness to the design concepts. To be eligible, participants had to be an undergraduate student at our university and have prior experience using GenAI for academic work. We administered a short pre-screening survey during recruitment to confirm eligibility and ensure diversity in academic discipline (see Appendix D). Each participant received a \$25 honorarium in appreciation of their time. The study was approved by our institutional Research Ethics Board (see Appendix E for approval certificate).

4.3 Procedure

We began each session with a short survey to collect baseline information about the participant's experience with the use of GenAI in academic contexts, along with basic demographic information (see Appendix F). We then introduced the participant to a common example scenario used across all demonstrations (see Appendix G). The scenario described an undergraduate student using ChatGPT to help with an essay assignment that had a pressing deadline. We presented the four video prototypes one at a time in a counterbalanced order using a Latin Square design [17] to mitigate ordering effects. Before each demonstration, we encouraged the participant to consider how they might integrate the design concept into their own academic

workflows. After each video demonstration, we conducted a brief semi-structured interview to capture the participant’s initial reactions, including perceived usefulness, concerns, and overall impressions of the concept (see Appendix H). In the final phase, we asked the participant to rank the four design concepts from most to least preferred along the following criteria:

1. How well the designs fit with their typical workflow;
2. The extent to which the designs would influence their sense of ownership over their work;
3. The extent to which the designs would prompt them to pause and reflect on AI’s response before using it

We concluded the session with a broader discussion about their overall experience across all concepts and their general views on interventions supporting reflective use of GenAI for academic work (see Appendix I). The entire session was audio recorded for subsequent analysis.

4.4 Data Analysis

Our primary data source consisted of audio recordings of the study sessions, which captured participants’ reactions to the video demonstrations, responses during semi-structured interviews, and their rankings of the design concepts. We used Reflexive

Thematic Analysis [18] to identify recurring themes and patterns from our sessions. Our analysis was framework-informed while still allowing themes to be developed inductively from participants' responses [19]. Specifically, the design dimensions of externalization and friction functioned as sensitizing concepts [15] as they informed the interview context and our interpretation of the data, but were not used as fixed coding categories. Sensitizing concepts can provide initial directions for attending to the data without prescribing what the analysis should find [15]. I began the analysis by repeatedly reviewing the transcripts to gain familiarity with the data and conducted iterative open coding [27] to capture meaningful segments related to participants' perceptions of the design concepts. I then identified initial patterns by grouping related codes into candidate themes. Through several rounds of discussion and refinement, my advisor and I collaboratively reviewed and revised these themes, continually cross-referencing the original transcripts to ensure interpretive consistency and grounding in the data.

4.5 Findings

We begin our findings by summarizing the responses from the baseline survey on existing practices with GenAI within academic settings. We then report participants' overall preferences for the proposed design concepts based on ranking data. Finally, we present findings from our thematic analysis of the interview data.

4.5.1 Current Practices with GenAI

Responses from the baseline survey provided insight into participants' prior experience with GenAI for academic work before exposure to the design concepts. Nearly all participants reported using ChatGPT (19/20), followed by Gemini (12/20) and Microsoft Copilot (9/20) for academic work. Most participants had more than one year of experience using GenAI (16/20), and many reported using it regularly, such as a few times per week (13/20). The most common use case was summarizing readings or lecture notes (16/20), often as a way to extract essential information efficiently. For example, P12 described using ChatGPT to get *"the core gist without necessarily reading the whole material."* More than half of the participants (13/20) also reported using GenAI to solve academic problems, particularly under time pressure. As P07 noted, they *"copy pasted the entire handout to ChatGPT"* just to get the answer directly when managing multiple deadlines. Together these findings suggest that our participants often prioritize efficiency and rapid comprehension when using GenAI in academic contexts.

4.5.2 Overall Preferences for the Design Concepts

To get a sense of participants' overall preferences, we asked them to rank the four design concepts based on how well each would fit into their typical use of tools like ChatGPT. As shown in Figure 4.1A (Workflow Fit), rankings were relatively distributed, with no single concept emerging as the clear winner. Clearer prefer-

ences emerged for the other two criteria. For reflection on ownership (Figure 4.1B), *Workload Balance Slider* was most frequently ranked first (13/20), while for reflection on the AI’s response (Figure 4.1B), *Predict Before Reveal* was the dominant top choice (13/20). Notably, both of these concepts fall toward the lower end of the externalization spectrum, suggesting that participants may have been more receptive to interventions that require less explicit articulation. In contrast, *Quick Check* was most frequently ranked second across all the criteria, suggesting that participants might have viewed it as a balance between enabling reflection and maintaining manageable levels of externalization and friction.

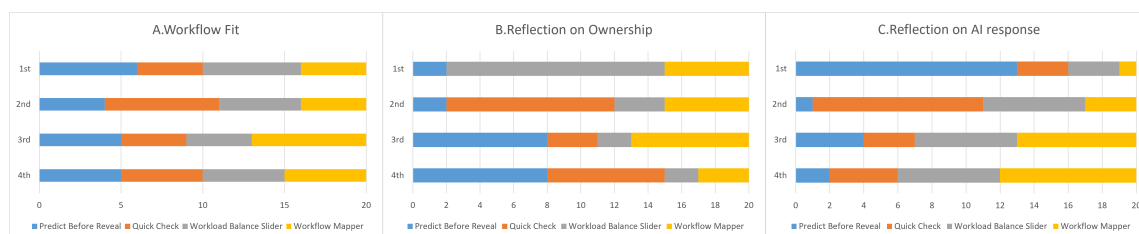


Figure 4.1: Participants ranked the four design concepts from most to least preferred separately for three criteria: (A) workflow fit (to what extent the concept might be integrated into their typical GenAI use); (B) reflection on ownership (to what extent the concept would be able to influence on their sense of ownership over AI-assisted work); and (C) reflection on AI response (to what extent the concept would be able to prompt users to pause and reflect on AI-generated responses before using them).

4.5.3 Perceptions Towards the Design Concepts

In this section, we present findings from our thematic analysis, which revealed that students' receptiveness to reflection-oriented interventions is shaped by a combination of perceived metacognitive skills, external motivations such as accountability and academic integrity, and design characteristics. Importantly, these preferences were not uniform, but rather seemed to depend on disciplinary context and task characteristics.

4.5.3.1 Perceived Metacognitive Skills Shape Receptiveness to Reflection-Oriented Interventions

Participants' receptiveness to the design concepts varied based on how they perceived their own ability to monitor and regulate their use of GenAI. Some participants acknowledged that they may not always engage critically with AI-generated responses or deliberately reflect on their own contributions to the final output. This inconsistency is seen to be influenced by various situational factors such as time pressure or a lack of habitual reflection. For example, in the following, the participant discusses how time pressure leads them to skip reflection, appreciating that *Predict Before Reveal* could help reintroduce a deliberate pause for thinking.

*“under time pressure, sometimes I’m not that clear on the topic but I just want to get going and at that time if **Predict Before Reveal** asks me what do I expect [the answer to be] then I can personally reflect on it and it allows me to pause and think.” - (P03, Engineering)*

Another participant acknowledged that they had not previously considered re-

flecting on their ownership when using GenAI, but having the *Workload Balance Slider* would help to make this typically unmonitored behaviour more visible.

*“Whenever I’m using AI, I usually don’t keep track [of my contribution] like this because whatever is in my mind, I just prompt it to AI and usually I just take the answer, customize it accordingly and submit it. But having the **Workload Balance Slider** will make me aware and cautious of how much the work is done by me and how much I’m using the AI. So it’s good for keeping a track [...] Because I don’t want to be completely dependent on AI and it will also encourage me to do most of the work by myself instead of using AI.” - (P20, Science)*

In contrast, other participants perceived themselves as already capable of self-regulating their use of GenAI and therefore saw little added value to external reflection support. However, these sentiments were typically articulated in broad or implicit terms, with participants rarely describing explicit or systematic strategies for reflection. Some participants relied on a subjective sense of understanding, assuming that using or rejecting an AI response already implied sufficient evaluation.

P06 describes how they view additional reflection prompts as unnecessary:

*“I’m thinking that the **Quick Check** is just a nuisance,[...] because to use the response in my work, I’ll obviously have to understand the response and evaluate it. If I don’t end up using the AI’s response, that means that I did the understanding and evaluating steps already.” - (P06, Computer Science)*

Notably, even among participants who expressed confidence in their ability to self-regulate, occasional moments of uncertainty suggested that their perceived level of metacognition may not always align with actual reflective behaviour.

“I feel like I do enough of the critical thinking myself, but maybe I’m naive and I actually don’t do enough critical thinking. I don’t know.” - (P18, Science)

Overall, the above sentiments suggest that while participants value opportunities for additional reflection, others may reject reflection-oriented design concepts because they are confident in their own reflective abilities.

4.5.3.2 Reflection on Ownership Is Often Motivated by Academic Integrity Concerns Rather than Intrinsic Learning Goals

Across participants, reflection on the ownership of AI-assisted work was frequently motivated not by a desire to critically examine one’s contribution, but by concerns surrounding academic integrity and the possibility that their work would be interpreted as overly AI-dependent. As an example, one participant described wanting to use the *Workload Balance Slider* as a way to adjust their wording to align with their personal voice. In doing so, their goal was to avoid suspicion by instructors as opposed to ensure sufficient independent contribution or to support their learning.

“When I am positioning the slider, I would think of [...] the type of words I’m using. Teachers, professors and educators can tell a lot of the time if it’s AI or if it’s us because they listen to how we speak and know how we write. So they can see something on the text and know this isn’t my vocabulary so then that’s when it can pretty much expose us.” - (P05, Arts)

We did see contrastive views in our data, where authorship reflection was viewed as an important component for helping to maintain academic integrity, particularly due to participants’ awareness of potential consequences.

*“using **Quick Check** would influence how I engage with AI responses. Whenever I say I’m going to apply it in my own work, it will make me think how many times I used AI’s responses. It’s not good if I use AI’s responses all the*

time when it comes to academic integrity.[...] Suspension is not something you want in your life.” - (P03, Engineering)

There were participants who also reflected on ownership for more intrinsic reasons, such as ensuring that they had made a meaningful contribution to their work or that they were maintaining their own personal standards.

*“If I found that the report in Workload Balance Slider was leaning towards mostly AI or all AI, then I [would] need to change things up because it’s my assignment. Instructors are not expecting an AI assignment. Even for the **Workflow Mapper**, if I see that there’s a lot of purple [sticky notes], which means there’s a lot of AI contribution in my assignment, [...] then I would remove some stuff [from the assignment] and write it myself.” - (P13, Business)*

*“Sometimes I catch myself too lazy to even write grammatically correct sentences because I know ChatGPT is probably going to understand me.[...] But after I see the weekly report in the **Workload Balance Slider** and [realize] that I’m mostly not using my own brain power, I would maybe consider [...] what I want to get back from ChatGPT and engage into the work more.” - (P14, Science)*

Overall, while some participants focused on ensuring meaningful personal contribution, authorship reflection was more often framed as a strategy for avoiding negative consequences than for supporting deeper engagement with the learning process.

4.5.3.3 Reflection Is Often Driven by External Accountability and Incentives Rather than Intrinsic Motivation

Participants, particularly those from Business, often expressed limited intrinsic motivation to engage with reflection-oriented interventions, instead emphasizing the need for external accountability or incentives to justify their use.

*“I think if there was an AI system provided by the university where they actually track these sorts of things, then that I’d probably say is a big factor that would affect how I would use the **Workload Balance Slider**. But if it’s just between me and the AI, then I probably wouldn’t use it.” - (P15, Business)*

*“Maybe it could be an incentive kind of thing where after a certain period of time you get points for using **Quick Check** and that might encourage people to use it and reflect on how they’re relating with the AI responses.” - (P13, Business)*

Similarly, others indicated that when reflection is tied to external structures, such as grading or monitoring, it becomes recontextualized as meaningful and necessary work. For these participants, the value of reflection is not derived from its role in supporting learning, but from its alignment with evaluative or outcome-driven goals.

“I would think that this is a great thing if also the teacher could see it. Because if it’s just for me, then I don’t see any reason to use it. I think it would encourage me to think more critically [if I’m being monitored by a teacher]. Then I wouldn’t just rely on copy-pasting, but I would actually have to write my own input. Because the teacher could see directly how much AI I actually used.” - (P16, Business)

These sentiments highlight how engagement with reflective interventions may be shaped by a cost–benefit logic, where users might be more likely to participate when reflection is externally justified rather than intrinsically motivated.

4.5.3.4 Preferences for Degree of Externalization and Level of Friction Varied Across Academic Disciplines

Participants’ preferences for both design dimensions varied across academic disciplines. Across the design concepts, disciplinary background appeared to shape

how participants perceived meaningful reflection and productive engagement with conversational GenAI.

Science-oriented participants generally preferred more implicit forms of reflection while also expressing greater receptiveness towards high-friction interventions. Participants described mentally predicting or reasoning through a problem before viewing the AI's response as more aligned with their learning practices than explicit forms of reflection that required articulating contributions or expectations. In the quotes below, participants describe how implicit reflection would align with the objective or procedural nature of their tasks.

*"I'm mostly asking [ChatGPT] for a one-sentence answer, so I feel like the **Workflow Mapper** would be kind of a bit too much and has a lot of unnecessary extra [steps] where you have to type in or add your contribution, what you contributed and what the AI contributed." - (P11, Science)*

*"**Predict Before Reveal** is definitely the best. Being able to think critically before [viewing] the AI's response helps you to use the AI as a supporting beam. [...] if you don't get the same answer as the AI's, then you can go back and think where did I go wrong? And you could go find that journey before coming back to the AI [for help]." - (P09, Science)*

Additionally, the participants from Science noted that friction that interrupts access to the answer early on could be beneficial for promoting cognitive engagement. Some participants even related high-friction interventions to their use of flashcards, suggesting that delaying access to the AI's response aligns with their existing practice of active recall.

*"I think **Predict Before Reveal** would work well because as a Science student, I often use flashcards and I often try to use active recall. So this concept is almost like flashcards. It's kind of like you have to think and try your best to remember the definition." - (P11, Science)*

Participants from Computer Science/Engineering backgrounds expressed stronger opinions regarding friction than externalization. These participants generally favored low-friction interventions, often describing highly interruptive mechanisms as disruptive to their workflow, possibly reflecting a more optimization-oriented view of GenAI use.

*“The **Workload Balance Slider** ruins the experience [of using ChatGPT] the most. But the **Workflow Mapper** is the best. It doesn’t lock the chats. It could be there and it’s not interfering [with my workflow].” - (P06, Computer Science)*

Compared to participants from other disciplines, Computer Science/Engineering participants appeared relatively indifferent to whether reflective support required explicit articulation or remained implicit, suggesting that minimizing workflow disruption may have been more important than engaging in reflective articulation itself.

Participants from Arts and Business, on the other hand, were more vocal about their preferences regarding externalization than friction. They often perceived highly implicit interactions as insufficient for meaningful reflection. Out of the concepts shown, participants in these domains frequently preferred concepts such as the *Workload Balance Slider* or *Quick Check* as they required some form of self-expression. However, the participants also suggested modifications that would allow them to demonstrate their understanding via written articulation, indicating a preference for reflection mechanisms that made their understanding visible and comparable to the AI’s output.

*“We can improve **Predict Before Reveal** by prompting the user to write the answer over there to ChatGPT, and then ChatGPT can show the difference*

between our expectation and that of AI's. [...] We'll get used to just clicking on the mask to reveal the answer and ultimately use the answer that ChatGPT is giving instead of brainstorming. " - (P07, Arts)

In contrast, their preferences regarding friction were more mixed, spanning both low- and high-friction interactions depending on the context of use.

Overall, the above findings suggest that preferences along the two dimensions are not likely to be uniform. Some disciplines in our sample exhibited relatively consistent preferences, while others expressed varied preferences that may depend on various situational factors.

4.6 Discussion

Our findings provide insights into how students perceive reflection-oriented design interventions for conversational GenAI that differ according to the amount of externalization elicited and the level of friction introduced. The ranking data suggest that participants did not view any single design concept as universally preferable for integration into their academic workflows. Instead, support for the four concepts depended on how well each aligned with individual reflective goals and contextual needs.

Our interview findings show that receptiveness to reflection-oriented interventions varied based on how participants perceived their own reflective practices. Participants who were aware of gaps in their reflection valued the design concepts, seeing them as helpful for prompting more deliberate engagement. In contrast, those who

expressed confidence in their reflective abilities were less receptive, although the practices they described did not appear to match evidence-based strategies [20; 29]. This inconsistency between perceived competence and actual practice aligns with the concept of cognitive dissonance [44]. A third group of participants expressed uncertainty about whether they reflected adequately at all. Altogether, these patterns suggest that students might lack a clear understanding of what meaningful reflection with GenAI entails and why it could promote better learning outcomes. In this context, the proposed interventions can play a role not only in prompting reflection, but also in shaping a more concrete and shared understanding of what effective reflective use of GenAI can look like in practice.

In terms of reflecting on ownership, we were surprised to hear some participants express a motivation to do so primarily to avoid suspicion, penalties, or accusations. These participants, even when prompted for further details, did not articulate a clear link between reliance on AI-assisted work and learning outcomes. This pattern might be shaped by how GenAI use is positioned within academic environments, where concerns around academic integrity, grading, and potential penalties are often more immediate and salient than long-term learning goals [54; 77]. In such contexts, students may come to interpret questions of ownership primarily through a lens of compliance and risk management. At the same time, a lack of clear guidance around what constitutes meaningful or acceptable AI contribution can leave students relying on surface-level cues, such as writing style or vocabulary, to judge their

work. This ambiguity may further reinforce a defensive stance, where the focus shifts toward avoiding suspicion rather than engaging with authorship as part of the learning process. This pattern of compliance-driven reflection is further echoed in the responses of participants from Business-oriented backgrounds, who appeared more receptive to the design concepts when reflection was tied to monitoring, grading, or incentives, rather than when it was framed as a tool for supporting deeper learning. This consistency across findings suggests that external structures might serve as key triggers for engaging in reflection, particularly in contexts where intrinsic motivation to reflect on GenAI use is limited.

Academic discipline appeared to influence participants' preferences for different points along our two design dimensions. For example, the Computer Science/ Engineering participants preferred low-friction interactions, prioritizing efficiency and uninterrupted workflows. Given their greater familiarity with AI systems and awareness of the rapid pace of advancement in these technologies [23; 91], they may have developed stronger expectations for seamless interactions. This attitude also matches findings from studies of active learning strategies, which students can find too demanding or inefficient, despite their association with deeper understanding and improved learning outcomes [13; 38]. Our participants from Arts and Business backgrounds, on the other hand, were more receptive to higher-friction interventions that matched their typical active recall study practices. We also saw mixed preferences along the degree of externalization, however, how to support those interested in a high degree

of explicit externalization warrants further design investigation. For example, one participant wanted to fully express their understanding with computational support for comparing their understanding to the GenAI's response. Considering these findings collectively, manipulating the two dimensions helped reveal underlying factors shaping students' willingness to engage in reflection, while also suggesting opportunities for designing adaptive interventions that vary the degree of externalization and friction based on disciplinary context, task structure, and user preferences.

Based on our findings, we suggest several provisional design heuristics aimed at supporting students' receptiveness to reflection-oriented GenAI interventions. These heuristics are intended to inform how such interventions might be designed to better align with students' tasks, workflows, and motivations, rather than to prescribe designs that improve learning outcomes. Designers should match friction to task type: brief pre-response friction may be more acceptable by students for active recall or problem-solving tasks, whereas low-friction support may be preferred in contexts where students prioritize speed and workflow continuity. Externalization should likewise be scaled to task complexity, with more explicit reflection potentially better suited to multi-step or articulation-heavy tasks and more implicit forms to procedural or objective tasks. When workflow continuity is central, designers should consider keeping explicit reflection artifacts optional rather than making them mandatory steps in the interaction. Finally, designers should consider both students' perceived need for reflective scaffolding and the motivational framing of reflection, as partici-

pants were more receptive when interventions aligned with their perceived reflective needs or, in some cases, with external accountability or incentives. These heuristics are intended to guide future design exploration and concern the design of interventions that students may be more willing to engage with. Determining whether such interventions ultimately change reflective behavior or improve learning outcomes would require further evaluation studies through functioning systems and behavioral data in authentic academic settings.

4.7 Summary

In this chapter, we described the elicitation study conducted to examine students' perceptions of the proposed reflective design concepts. We presented the study objectives, participant characteristics, study procedures, and methods used for data collection and analysis. Through video-based demonstrations of the design concepts, semi-structured interviews, and ranking activities, we gathered insights regarding students' reactions, perceived usefulness, concerns, and preferences surrounding the interventions. The chapter also presented findings that emerged from the interview and ranking data and discussed their implications for understanding how reflective support may be integrated into conversational GenAI systems.

Chapter 5

Reflection on the Design Space

In this chapter, we reflect on implications arising from the findings presented in the previous chapter and extend the design space introduced in Chapter 3. Drawing on both study findings and prior literature, we consider how students' motivation may shape their willingness to engage with reflective interventions. Building on this discussion, the chapter introduces *type of motivation* as a potential third design dimension and presents *Reflection Credits* as a speculative fifth design concept that explores the extrinsic end of this dimension. We then provide an updated representation of the design space after adding the new design dimension and the fifth design concept.

5.1 Extending the Design Space

The design space presented in Chapter 3 organized concepts according to the degree of externalization required from users and the amount of friction introduced into the interaction process. While these dimensions captured important differences in how reflection can be operationalized, our findings suggest that students' motivation to engage with reflection may also play an important role in shaping the acceptance and perceived usefulness of reflective interventions.

In particular, findings from the elicitation study indicated that participants often framed reflection in terms of external accountability rather than intrinsic learning goals. Rather than valuing reflection solely for supporting understanding or deeper engagement with AI-generated content, several participants emphasized incentives such as grades, monitoring, institutional oversight, or other externally imposed structures. These findings suggest that students may not necessarily engage with reflective activities simply because they support learning outcomes. Instead, participation may depend on whether reflection aligns with motivations that users perceive as meaningful or beneficial. To account for this pattern, the following section introduces *type of motivation* as a potential additional design dimension, ranging from intrinsically motivated reflection to externally motivated reflection.

5.1.1 Type of Motivation (Intrinsic - Extrinsic)

This dimension captures the extent to which students' engagement with reflective interventions is driven by internal versus external sources of motivation. Intrinsic motivation refers to engaging in an activity for its own sake, where the activity is motivated by interest, enjoyment, challenge, or satisfaction derived from the task itself. Extrinsic motivation, in contrast, refers to engaging in an activity in order to attain an external goal or meet an externally imposed constraint [56; 89]. At the intrinsic end of this dimension, reflective interventions may encourage students to engage with reflection because they perceive it as valuable for learning, self-understanding, critical thinking, or skill development. At the extrinsic end, reflective interventions may encourage engagement by connecting reflection to external accountability, such as grades, instructor visibility, institutional expectations, monitoring, or incentives.

Prior literature suggests that academic motivation can be an important factor in the development of critical thinking among students [12]. In particular, intrinsically motivated students are often more likely to engage deeply with learning materials, persist through challenging tasks, and process information in more meaningful ways. In contrast, more extrinsically oriented goals may sometimes be associated with shallower forms of information processing, where students focus on task completion, performance, or external outcomes rather than deeper understanding [51]. However, intrinsic and extrinsic motivation should not be treated as entirely separate or opposing categories. Prior work suggests that extrinsic motivators can sometimes

undermine intrinsic motivation, but they may also support engagement depending on how they are structured and interpreted by the learner [56].

Including this dimension in our design space was motivated by findings from our elicitation study, where several participants framed reflection less as an intrinsically valuable learning activity and more as something they would engage with if it were externally justified. For example, some participants described reflective interventions as more useful if instructors could view their activity, if the intervention was embedded in a university-provided and monitored system, or if it was connected to incentives. These comments were especially visible among participants from Business-oriented backgrounds, suggesting a possible relationship between motivation and disciplinary context. However, this pattern should not be interpreted as evidence that disciplinary background directly causes a particular motivational orientation. Disciplinary norms, assessment cultures, professional expectations, or students' pre-existing motivational orientations may all shape how students interpret reflection and accountability.

This dimension may also interact with the existing dimensions of externalization and friction. For instance, students may be more willing to tolerate higher-friction interventions or more explicit forms of externalization if the reflective activity is connected to external accountability or perceived academic benefit. Conversely, if reflection is framed only as optional self-monitoring, students with limited intrinsic motivation may be less likely to engage, particularly when the intervention requires

additional effort. Future design work could explore how different motivational framings shape students' acceptance of reflective interventions. Since the four design concepts developed in Chapter 3 primarily rely on students recognizing the intrinsic value of reflection for learning and self-monitoring, the following section introduces a speculative fifth design concept that explores the more extrinsic end of the motivation dimension.

5.1.2 Reflection Credits: A Speculative Design Concept

This speculative design concept (Figure 5.1) explores how reflective engagement might be encouraged through an extrinsic point-based mechanism. In this concept, each prompt submitted to the GenAI system costs reflection credits, with prompt costs increasing as students continue to rely on the system. To earn additional credits, students complete a brief reflection activity in which they explain how they have used AI so far in the assignment. The reflection is saved as part of an AI-use record that could be submitted to an instructor, connecting the activity to external accountability rather than only self-monitoring. Within the extended design space, we position *Reflection Credits* toward the extrinsic end of the motivation dimension, while involving a moderate level of externalization and friction.

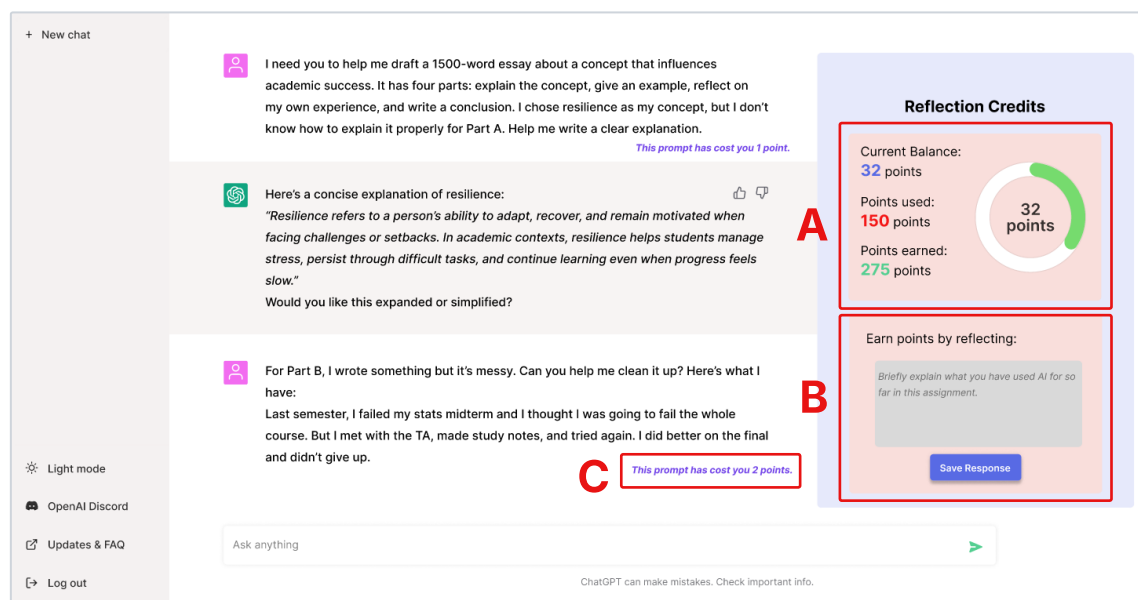


Figure 5.1: Screenshot of Reflection Credits. (A) The side panel shows the student’s current point balance, the number of points used for prompting, and the number of points earned through reflection. (B) Students can earn additional points by completing a brief reflection explaining how they have used AI so far in the assignment. (C) Each prompt submitted to the system costs points, with the cost increasing as students submit more continuous prompts.

5.1.3 Representation of the Extended Design Space

In this section, we present the extended design space after introducing type of motivation as a third dimension and *Reflection Credits* as a speculative fifth concept (Figure 5.2). Unlike the original representation, which positioned the four concepts only in relation to externalization and friction, this revised representation shows how the concepts can also be interpreted through their motivational orientation. Along this dimension, *Predict Before Reveal* is positioned closest to the intrinsic

end because it relies primarily on students' own willingness to pause and mentally anticipate the AI response before viewing it. *Workflow Mapper* is positioned near the intrinsic end of the motivation dimension, as it supports self-directed reflection by asking students to create a visible representation of their process while they work. *Quick Check* and *Workload Balance Slider* are positioned closer to the middle of the motivation dimension because both provide structured self-assessment and weekly reports that can encourage students to monitor their engagement over time. Of these two, *Workload Balance Slider* is positioned slightly closer to the extrinsic end because its contextual feedback and summary reports more directly prompt students to adjust their AI-use behavior and reflect on the balance between their own contribution and the AI's contribution. Finally, *Reflection Credits* is positioned closest to the extrinsic end because it connects reflection to a point-based incentive mechanism and an instructor-submittable AI-use record.

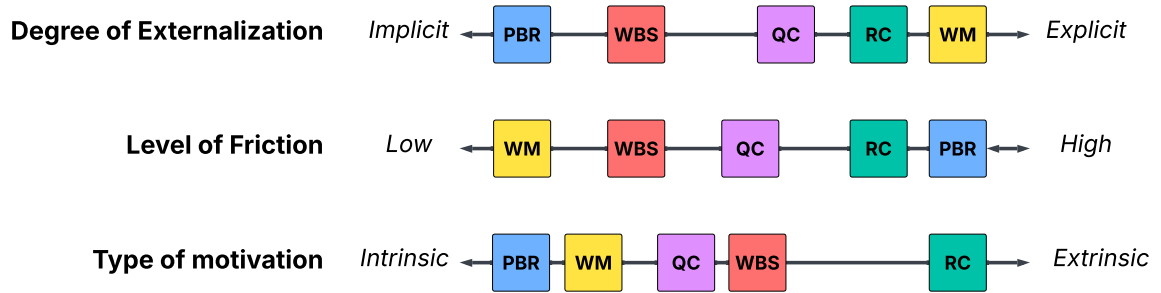


Figure 5.2: Extended representation of the design space with five concepts mapped across three dimensions: degree of externalization, level of friction, and type of motivation. PBR = Predict Before Reveal, QC = Quick Check, WBS = Workload Balance Slider, WM = Workflow Mapper, and RC = Reflection Credits.

5.2 Summary

In this chapter we extend the original design space by identifying motivation as an additional dimension in the design space. Our discussion highlights that students may not always engage with reflection because they value it as a learning activity but rather reflection may become more meaningful when connected to external accountability, incentives, or institutional expectations. To illustrate this extension, we introduce *Reflection Credits*, a speculative design concept that links reflective activity to a point-based prompting system. We also provide an updated design space representation that shows how the new concept extends beyond the four original concepts by exploring a more extrinsically motivated form of reflective engagement.

Chapter 6

Conclusion

In this thesis, we explored how undergraduate students perceive interface-level interventions designed to promote reflective use of conversational GenAI. We addressed this question by developing an initial design space using externalization and friction as guiding design dimensions for reflection-oriented interface interventions. We then created four contrasting design concepts, and examined how undergraduate students perceived these concepts through a qualitative elicitation study. Our findings show that receptiveness to such interventions is not uniform, but shaped by students' perceived metacognition, external accountability, academic integrity concerns, and disciplinary context. Building on our findings, we further reflect on the type of motivation as an additional design dimension consideration for designing reflective GenAI interventions and introduce a new design concept that is a speculative extension to the original design space.

6.1 Contributions

This thesis contributes an initial design space for reflective interventions in conversational GenAI systems for higher education. We initially organized the design space around two dimensions: degree of externalization and level of friction. These dimensions characterize how much an intervention asks students to externalize their thinking and how strongly it interrupts the typical prompt-response workflow. Building on the study findings and prior literature, we extend this design space by introducing type of motivation, ranging from intrinsic to extrinsic, as a potential third dimension. This extended design space considers that reflective engagement may depend not only on how an interface structures reflection, but also on whether students perceive reflection as personally valuable, externally accountable, or connected to academic incentives. To illustrate this extension, we present *Reflection Credits* as a speculative fifth design concept.

This thesis also contributes five reflection-oriented design concepts: *Predict Before Reveal*, *Quick Check*, *Workload Balance Slider*, *Workflow Mapper* and *Reflection Credits*. These concepts demonstrate different ways of embedding reflective support into a simulated conversational GenAI environment, including delaying access to AI responses, prompting structured self-assessment, encouraging students to consider their own contribution, supporting visual mapping of AI-assisted work, and linking reflection to point-based incentives and external accountability. The first four concepts formed the design set examined in our user study, whereas we developed

Reflection Credits afterward as a speculative concept informed by our study findings as well as prior literature.

Finally, this thesis contributes empirical findings from a qualitative elicitation study with undergraduate students. The findings show that students' receptiveness to reflective GenAI interventions varies across concepts and contexts. Participants' responses depended on how they perceived their own reflective abilities, whether they connected reflection to academic integrity concerns or external accountability, and how well the interventions fit disciplinary practices, task types, and workflow constraints. Together, these findings highlight the need for context-sensitive reflective GenAI interventions that account for students' academic settings, motivations, and everyday practices.

6.2 Limitations and Future Work

While our findings offer useful implications for the design of reflection-oriented GenAI interventions, they should be interpreted in light of several limitations. Given the sensitive nature of GenAI use in academic contexts, some participants may have been hesitant to fully disclose their actual practices, a challenge that has been documented in prior work [2]. Although we made efforts to reassure participants about confidentiality and create a comfortable interview environment, this hesitation may have resulted in partial or socially desirable responses [69].

Second, our sample size within each academic discipline was relatively small,

limiting our ability to draw strong conclusions about disciplinary differences. The patterns that we report across discipline should be interpreted as exploratory rather than definitive. Perspectives from other disciplines, particularly those emphasizing creative production (e.g., Fine Arts and Architecture) might yield different insights. In addition, all participants were recruited from a single institution, which may limit the transferability of our findings to other university contexts.

Third, we used video prototypes to elicit students' perceptions of the design concepts. This approach allowed us to present multiple concepts consistently and avoid implementation-specific issues, but it did not provide behavioral data. Participants' stated preferences and anticipated responses may therefore not reflect how they would actually engage with these interventions in authentic academic settings, particularly under workload and deadline pressures. Future work should investigate how such promising interventions function in real-world deployments over time, particularly how they influence actual learning behaviors and outcomes, and how they can be tailored to align with diverse academic contexts and motivations.

Building on our initial insights, future work should continue to refine and extend the design space by identifying additional dimensions of reflective GenAI interaction and developing contrasting design concepts that instantiate them. One potential dimension could be intervention timing. Although timing did not emerge as a distinct theme in our interview findings, our concepts placed reflective support at different points in the interaction, suggesting a useful direction for future studies. Future work

could systematically explore reflective support that appears before response reveal, after each prompt, after several prompts, or at the end of a task or session. Future work should also further investigate the motivation dimension and the speculative *Reflection Credits* concept. Studies could examine whether point-based incentives and external accountability encourage meaningful reflection or instead produce superficial compliance, strategic completion, or resistance. Future studies could also compare intrinsically framed and extrinsically framed interventions to understand how the type of motivation interacts with externalization, friction, disciplinary context, and students' academic goals.

Future work should also examine how reflective GenAI interventions are shaped by the institutional and assessment contexts in which they are deployed. Rather than treating these interventions as standalone interface features, future studies should consider how university policies, course expectations, and assessment designs influence how students interpret and respond to them. Institutional policies may shape what students understand as acceptable GenAI assistance, meaningful student contribution, and appropriate evidence of authorship. Assessment practices may also affect the value of specific reflection artifacts. For example, weekly AI-use reports or workflow maps may be more relevant in process-oriented assessments than in contexts where evaluation focuses primarily on the final output. Therefore, future studies should examine reflective GenAI interventions across institutions, courses, and assessment cultures with different expectations around acceptable AI use.

Bibliography

- [1] ABBAS, M., JAM, F. A., AND KHAN, T. I. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education* 21, 1 (2024), 10.
- [2] ADNIN, R., PANDKAR, A., YAO, B., WANG, D., AND DAS, M. Examining student and teacher perspectives on undisclosed use of generative AI in academic work. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2025), Association for Computing Machinery.
- [3] AL-KFAIRY, M., MUSTAFA, D., KSHETRI, N., INSIEW, M., AND ALFANDI, O. Ethical challenges and solutions of generative ai: An interdisciplinary perspective. In *Informatics* (2024), vol. 11, MDPI, p. 58.
- [4] ALTER, A. L., AND OPPENHEIMER, D. M. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* 13, 3

- (2009), 219–235.
- [5] ALVINA, J., BUNT, A., CHILANA, P. K., MALACRIA, S., AND MCGRENERE, J. Where is that feature? Designing for cross-device software learnability. In *Proceedings of the ACM Designing Interactive Systems Conference* (New York, NY, USA, 2020), Association for Computing Machinery, p. 1103–1115.
- [6] AMOOZADEH, M., DANIELS, D., NAM, D., KUMAR, A., CHEN, S., HILTON, M., SRINIVASA RAGAVAN, S., AND ALIPOUR, M. A. Trust in generative ai among students: An exploratory study. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (2024), pp. 67–73.
- [7] BAI, L., LIU, X., AND SU, J. ChatGPT: The cognitive effects on learning and memory. *Brain-X* 1, 3 (2023), e30.
- [8] BANNERT, M., SONNENBERG, C., MENGELKAMP, C., AND PIEGER, E. Short- and long-term effects of students’ self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior* 52 (2015), 293–306.
- [9] BARCAUI, A. ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. *Social Sciences & Humanities Open* 12 (2025), 102287.

-
- [10] BARRETT, A., AND PACK, A. Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 59.
- [11] BASTANI, H., BASTANI, O., SUNGU, A., GE, H., ÖZGE KABAKCI, AND MARIMAN, R. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences* 122, 26 (2025), e2422633122.
- [12] BERESTOVA, A., KOLOSOV, S., TSVETKOVA, M., AND GRIB, E. Academic motivation as a predictor of the development of critical thinking in students. *Journal of Applied Research in Higher Education* 14, 3 (2022), 1041–1054.
- [13] BJORK, E. L., BJORK, R. A., ET AL. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* 2, 59-68 (2011), 56–64.
- [14] BORJI, A. A categorical archive of chatgpt failures. *arXiv preprint arXiv:2302.03494* (2023).
- [15] BOWEN, G. A. Grounded theory and sensitizing concepts. *International journal of qualitative methods* 5, 3 (2006), 12–23.

-
- [16] BRAAD, E., DEGENS, N., BARENDREGT, W., AND IJSSELSTEIJN, W. Improving metacognition through self-explication in a digital self-regulated learning tool. *Educational Technology Research and Development* 70, 6 (2022), 2063–2090.
- [17] BRADLEY, J. V. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association* 53, 282 (1958), 525–528.
- [18] BRAUN, V., AND CLARKE, V. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.
- [19] BRAUN, V., AND CLARKE, V. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International journal of transgender health* 24, 1 (2023), 1–6.
- [20] BROCKBANK, A., AND MCGILL, I. *Facilitating reflective learning in higher education*. McGraw-Hill Education (UK), 2007.
- [21] BUÇINCA, Z., MALAYA, M. B., AND GAJOS, K. Z. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021).

-
- [22] BUCHENAU, M., AND SURI, J. F. Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (2000), pp. 424–433.
- [23] BUDHIRAJA, R., JOSHI, I., CHALLA, J. S., AKOLEKAR, H. D., AND KUMAR, D. “It’s not like Jarvis, but it’s pretty close!” - Examining ChatGPT’s usage among undergraduate students in computer science. In *Proceedings of the Australasian Computing Education Conference* (New York, NY, USA, 2024), Association for Computing Machinery, p. 124–133.
- [24] BUXTON, B. *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann, 2010.
- [25] CAO, X., XU, S., CHEN, X., SONG, Y., LUO, Q., AND HE, T. Leveraging generative ai agent to promote teaching reflection in a k-12 ai course: Effects on teachers’ reflection self-efficacy, instructional design, and reflective thinking. *IEEE Transactions on Learning Technologies* (2026).
- [26] CARSON, L., BARTNECK, C., AND VOGES, K. Over-competitiveness in academia: A literature review. *Disruptive Science and Technology* 1, 4 (2013), 183–190.
- [27] CASCIO, M. A., LEE, E., VAUDRIN, N., AND FREEDMAN, D. A. A team-based approach to open coding: Considerations for creating intercoder consensus. *Field Methods* 31, 2 (2019), 116–130.

-
- [28] CHAN, C. K. Y., AND HU, W. Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 43.
- [29] CHANG, B. Reflection in learning. *Online learning* 23, 1 (2019), 95–110.
- [30] CHEN, X., TANKELEVITCH, L., VANUKURU, R., SCOTT, A. E., PANDA, P., AND RINTEL, S. Are we on track? AI-assisted active and passive goal reflection during meetings. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2025), Association for Computing Machinery.
- [31] CHI, M. T., DE LEEUW, N., CHIU, M.-H., AND LAVANCHER, C. Eliciting self-explanations improves understanding. *Cognitive Science* 18, 3 (1994), 439–477.
- [32] CHIA, J., AND FRATTAROLA, A. A design-based approach to analysing student engagement with a genai-enabled brainstorming app. *Computers and Education: Artificial Intelligence* (2025), 100468.
- [33] COHN, M., PUSHKARNA, M., OLANUBI, G. O., MORAN, J. M., PADGETT, D., MENGESHA, Z., AND HELDRETH, C. Believing anthropomorphism: Examining the role of anthropomorphic cues on trust in large language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024), Association for Computing Machinery.

-
- [34] CONATI, C., AND VANLEHN, K. Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education* 11 (2000), 389–415.
- [35] DAHLKEMPER, M. N., LAHME, S. Z., AND KLEIN, P. How do physics students evaluate artificial intelligence responses on comprehension questions? a study on the perceived scientific accuracy and linguistic quality of chatgpt. *Physical Review Physics Education Research* 19, 1 (2023), 010142.
- [36] DARABIPOURSHIRAZ, H., SMYSLOVA, D., XU, D., JIANG, S., AND LONG, D. Biasviz: A project-based, narrative-centered learning tool for engaging middle school students in critical thinking about ai biases. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (2026), pp. 1–18.
- [37] DAWSON, P., BEARMAN, M., DOLLINGER, M., AND BOUD, D. Validity matters more than cheating. *Assessment & Evaluation in Higher Education* 49, 7 (2024), 1005–1016.
- [38] DESLAURIERS, L., MCCARTY, L. S., MILLER, K., CALLAGHAN, K., AND KESTIN, G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences* 116, 39 (2019), 19251–19257.

-
- [39] DEVLIN, M., AND GRAY, K. In their own words: A qualitative study of the reasons Australian university students plagiarize. *Higher Education Research & Development* 26, 2 (2007), 181–198.
- [40] DROSOS, I., SARKAR, A., TORONTO, N., ET AL. ” it makes you think”: Provocations help restore critical thinking to ai-assisted knowledge work. *arXiv preprint arXiv:2501.17247* (2025).
- [41] ECKHARDT, S., KNAEBLE, M., BUCHER, A., STAEHELIN, D., DOLATA, M., AGOTAI, D., AND SCHWABE, G. “garbage in, garbage out”: Mitigating human biases in data entry by means of artificial intelligence. In *IFIP Conference on Human-Computer Interaction* (2023), Springer, pp. 27–48.
- [42] FANG, J.-W., CHEN, J., WENG, Q.-L., TU, Y.-F., HWANG, G.-J., AND XIA, Y.-C. Effects of a genai-based debugging approach integrating the reflective strategy on senior high school students’ learning performance and computational thinking. *Educational Technology & Society* 28, 3 (2025), 66–81.
- [43] FAUZI, F., TUHUTERU, L., SAMPE, F., AUSAT, A. M. A., AND HATTA, H. R. Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education* 5, 4 (2023), 14886–14891.
- [44] FESTINGER, L. Cognitive dissonance. *Scientific American* 207, 4 (1962), 93–106.

-
- [45] FIEDLER, K., ACKERMAN, R., AND SCARAMPI, C. Metacognition: Monitoring and controlling one’s own knowledge, reasoning and decisions. *The psychology of human thought: An introduction* (2019), 89–111.
- [46] FIGMA, INC. Figma: The collaborative interface design tool. <https://www.figma.com/>, 2026. Accessed: 2026-05-01.
- [47] FREEMAN, J. Student generative AI survey 2025. *Higher Education Policy Institute: London, UK* (2025).
- [48] FULSHER, A., PAGKRATIDOU, M., AND KENDEOU, P. Genai and misinformation in education: a systematic scoping review of opportunities and challenges. *AI & SOCIETY* 41, 2 (2026), 1373–1385.
- [49] GALITSKY, B. Truth-o-meter: Collaborating with llm in fighting its hallucinations. In *Interdependent human-machine teams*. Elsevier, 2025, pp. 175–210.
- [50] GAO, L., DAI, Z., PASUPAT, P., CHEN, A., CHAGANTY, A. T., FAN, Y., ZHAO, V., LAO, N., LEE, H., JUAN, D.-C., ET AL. Rarr: Researching and revising what language models say, using language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2023), pp. 16477–16508.
- [51] GARCIA, T., AND PINTRICH, P. R. Critical thinking and its relationship to motivation, learning strategies, and classroom experience.

-
- [52] GIANNISAKIS, E., ALVINA, J., BUNT, A., CHILANA, P., AND MCGRENERE, J. Promoting feature awareness by leveraging collaborators' usage habits in collaborative editors. In *Graphics Interface* (2022).
- [53] GOODMAN, L. A. Snowball sampling. *The Annals of Mathematical Statistics* 32, 1 (1961), 148–170.
- [54] GRUENHAGEN, J. H., SINCLAIR, P. M., CARROLL, J.-A., BAKER, P. R., WILSON, A., AND DEMANT, D. The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence* 7 (2024), 100273.
- [55] HALLNÄS, L., AND REDSTRÖM, J. Slow technology—designing for reflection. *Personal and ubiquitous computing* 5, 3 (2001), 201–212.
- [56] HENNESSEY, B., MORAN, S., ALTRINGER, B., AND AMABILE, T. M. Extrinsic and intrinsic motivation. *Wiley encyclopedia of management* (2015), 1–4.
- [57] JI, J. J. Demystify ChatGPT: Anthropomorphism around generative AI. *GRACE: Global review of AI community Ethics* 2, 1 (2024).
- [58] JOSE, B., AND THOMAS, A. Digital anthropomorphism and the psychology of trust in generative AI tutors: An opinion-based thematic synthesis. *Frontiers in Computer Science* 7 (2025).

- [59] JOSHI, N., AND VOGEL, D. Writing with ai lowers psychological ownership, but longer prompts can help. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces* (2025), pp. 1–17.
- [60] KASNECI, E., SESSLER, K., KÜCHEMANN, S., BANNERT, M., DEMENTIEVA, D., FISCHER, F., GASSER, U., GROH, G., GÜNNEMANN, S., HÜLLERMEIER, E., KRUSCHE, S., KUTYNIOK, G., MICHAELI, T., NERDEL, C., PFEFFER, J., POQUET, O., SAILER, M., SCHMIDT, A., SEIDEL, T., STADLER, M., WELLER, J., KUHN, J., AND KASNECI, G. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences 103* (2023), 102274.
- [61] KAZEMITABAAR, M., HUANG, O., SUH, S., HENLEY, A. Z., AND GROSSMAN, T. Exploring the design space of cognitive engagement techniques with AI-Generated code for enhanced learning. In *Proceedings of the 30th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2025), Association for Computing Machinery, p. 695–714.
- [62] KRATHWOHL, D. R. A revision of bloom’s taxonomy: An overview. *Theory Into Practice 41*, 4 (2002), 212–218.
- [63] KRUPP, L., STEINERT, S., KIEFER-EMMANOULIDIS, M., AVILA, K. E., LUKOWICZ, P., KUHN, J., KÜCHEMANN, S., AND KAROLUS, J. Unreflected acceptance—Investigating the negative consequences of ChatGPT-

- assisted problem solving in physics education. *arXiv preprint arXiv:2309.03087* (2023).
- [64] KU, K. Y., AND HO, I. T. Metacognitive strategies that enhance critical thinking. *Metacognition and learning* 5, 3 (2010), 251–267.
- [65] KUIPER, R. Enhancing metacognition through the reflective use of self-regulated learning strategies, 2002.
- [66] LEISER, F., ECKHARDT, S., KNAEBLE, M., MAEDCHE, A., SCHWABE, G., AND SUNYAEV, A. From chatgpt to factgpt: A participatory design study to mitigate the effects of large language model hallucinations on users. In *Proceedings of Mensch und Computer 2023*. 2023, pp. 81–90.
- [67] LEISER, F., ECKHARDT, S., LEUTHE, V., KNAEBLE, M., MÄDCHE, A., SCHWABE, G., AND SUNYAEV, A. Hill: A hallucination identifier for large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024), pp. 1–13.
- [68] LIN, X., HUANG, H., SHI, S., AND VINES, J. Relying on LLMs: Student practices and instructor norms are changing in computer science education. *arXiv preprint arXiv:2602.05506* (2026).
- [69] LING, Y., KALE, A., AND IMAS, A. Underreporting of AI use: The role of social desirability bias. In *Proceedings of the CHI Conference on Human*

- Factors in Computing Systems* (New York, NY, USA, 2026), Association for Computing Machinery.
- [70] MA, B., LI, H., LI, G., CHEN, L., TANG, C., XIE, Y., GU, C., SHIMADA, A., AND KONOMI, S. Scaffolding metacognition in programming education: Understanding student-ai interactions and design implications, 2025.
- [71] MA, Q., CHAI, M., TAN, Y., CHOI, J., KIM, J., HARPSTEAD, E., KAUFFMAN, G., AND WU, T. From prompts to reflection: Designing reflective play for genai literacy. *arXiv preprint arXiv:2509.13679* (2025).
- [72] MAIER, S., SCHNEIDER, M., AND FEUERRIEGEL, S. Partnering with generative AI: Experimental evaluation of model-led and human-led interaction in Human-AI co-creation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2026), Association for Computing Machinery.
- [73] MAO, L., DONG, L., LI, W., HU, X., LEE, K.-P., AND ZHOU, Z. Designing scaffolding cards to facilitate llm-based socratic instruction: An exploratory study of response strategies to support learning. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (2026), pp. 1–23.
- [74] McDONALD, N., JOHRI, A., ALI, A., AND COLLIER, A. H. Generative artificial intelligence in higher education: Evidence from an analysis of insti-

- tutional policies and guidelines. *Computers in Human Behavior: Artificial Humans 3* (2025), 100121.
- [75] MICROSOFT CLIPCHAMP. Clipchamp: Online video editor. <https://clipchamp.com/en/>, 2026. Accessed: 2026-05-01.
- [76] NILOY, A. C., AKTER, S., SULTANA, N., SULTANA, J., AND RAHMAN, S. I. U. Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning 40*, 2 (2024), 919–930.
- [77] ORTIZ-BONNIN, S., AND BLAHOPOULOU, J. Chat or cheat? Academic dishonesty, risk perceptions, and ChatGPT usage in higher education students. *Social Psychology of Education 28*, 1 (2025), 113.
- [78] PANADERO, E. A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology Volume 8 - 2017* (2017).
- [79] PANGANDAMAN, H. K., DATUMANONG, N. T., MUKATTIL, N. P., HAYUDINI, M. A. A., ABDULHAN, M. S., JILAH, A. J., ELAM, K. S., ABDURASUL, J. N. A., NAJAR, A. A., SARADI, M. A., ET AL. Effectiveness of mind mapping in the improvement of students academic performance: A systematic review. *Cuestiones de Fisioterapia 53*, 03 (2024), 1363–1375.
- [80] PARK, H., AND AHN, D. The promise and peril of ChatGPT in higher education: Opportunities, challenges, and design implications. In *Proceedings*

- of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024), Association for Computing Machinery.
- [81] PERKINS, M. Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice* 20, 2 (2023), 1–24.
- [82] PHAN, B. V. H., TONG, H. A., TRAN, D. Q., AND VAN NGUYEN, T. Investigating university students’ perceptions of plagiarism in the use of ChatGPT: An explanatory mixed-method study. *Education and Information Technologies* (2026), 1–33.
- [83] REITER, L., JÖRLING, M., FUCHS, C., , AND BÖHM, R. Student (mis)use of generative AI tools for university-related tasks. *International Journal of Human–Computer Interaction* 41, 19 (2025), 12390–12403.
- [84] RISKÓ, E. F., AND GILBERT, S. J. Cognitive offloading. *Trends in Cognitive Sciences* 20, 9 (2016), 676–688.
- [85] RODGERS, C. Defining reflection: Another look at john dewey and reflective thinking. *Teachers college record* 104, 4 (2002), 842–866.
- [86] ROSS, P., SCANES, E., AND LOCKE, W. Stress adaptation and resilience of academics in higher education. *Asia Pacific Education Review* 25, 4 (2024), 829–849.

-
- [87] ROZENBLIT, L., AND KEIL, F. The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science* 26, 5 (2002), 521–562.
- [88] RUTINOWSKI, J., FRANKE, S., ENDENDYK, J., DORMUTH, I., ROIDL, M., AND PAULY, M. The self-perception and political biases of chatgpt. *Human Behavior and Emerging Technologies* 2024, 1 (2024), 7115633.
- [89] RYAN, R. M., AND DECI, E. L. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology* 25, 1 (2000), 54–67.
- [90] SANTIAGO, H. C. Visual mapping to enhance learning and critical thinking skills. *Optometric Education* 36, 3 (2011).
- [91] SCHEFER-WENZL, S., VOGL, C., PEIRIS, S., AND MILADINOVIC, I. Exploring the adoption of generative AI tools in computer science education: A student survey. In *Proceedings of the 16th International Conference on Education Technology and Computers* (New York, NY, USA, 2025), Association for Computing Machinery, p. 173–178.
- [92] SCHUNK, D. H., AND ZIMMERMAN, B. J. *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press, 1998.
- [93] SINGH, A., GUAN, Z., AND RIEH, S. Y. Enhancing critical thinking in generative AI search with metacognitive prompts. *Proceedings of the Association for Information Science and Technology* 62, 1 (2025), 672–684.

-
- [94] SINGH, A., TANEJA, K., GUAN, Z., AND GHOSH, A. Protecting human cognition in the age of AI. *arXiv preprint arXiv:2502.12447* (2025).
- [95] SUH, S., MIN, B., PALANI, S., AND XIA, H. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2023), Association for Computing Machinery.
- [96] SULLIVAN, M., KELLY, A., AND McLAUGHLAN, P. ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching* 6, 1 (2023), 31–40.
- [97] TANKELEVITCH, L., KEWENIG, V., SIMKUTE, A., SCOTT, A. E., SARKAR, A., SELLEN, A., AND RINTEL, S. The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024), Association for Computing Machinery.
- [98] TIAN, X., OZTURK, D., EDULA, S., ADIL, J., JIN, Q., SHI, Y., AND BARNES, T. When ai gets it wrong: Scaffolding ai hallucination detection for children through chatbot creation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (2026), pp. 1–20.

-
- [99] TUCKMAN, B. W. Relations of academic procrastination, rationalizations, and performance in a web course with deadlines. *Psychological Reports* 96, 3 (2005), 1015–1021.
- [100] URCHS, S., THURNER, V., ASSENMACHER, M., HEUMANN, C., AND THIEMICHEN, S. How prevalent is gender bias in chatgpt?-exploring german and english chatgpt responses. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (2023), Springer, pp. 293–309.
- [101] VANLEHN, K., JONES, R. M., AND CHI, M. T. A model of the self-explanation effect. *Journal of the Learning Sciences* 2, 1 (1992), 1–59.
- [102] VITALE, F., ODOM, W., AND MCGRENERE, J. Keeping and discarding personal data: Exploring a design space. In *Proceedings of the ACM Designing Interactive Systems Conference* (New York, NY, USA, 2019), Association for Computing Machinery, p. 1463–1477.
- [103] VÁZQUEZ-CANO, E., RAMÍREZ-HURTADO, J. M., SÁEZ-LÓPEZ, J. M., AND LÓPEZ-MENESES, E. ChatGPT: The brightest student in the class. *Thinking Skills and Creativity* 49 (2023), 101380.
- [104] WANG, H., DANG, A., WU, Z., AND MAC, S. Generative AI in higher education: Seeing ChatGPT through universities’ policies, resources, and guidelines. *Computers and Education: Artificial Intelligence* 7 (2024), 100326.

-
- [105] WANG, J., AND CORNELLY, P.-R. Addressing academic misconduct in the age of ChatGPT: Strategies and solutions. In *Proceedings of the 7th International Conference on Education and E-Learning* (New York, NY, USA, 2024), Association for Computing Machinery, p. 19–25.
- [106] WANG, J., HU, H., WANG, Z., YAN, S., SHENG, Y., AND HE, D. Evaluating large language models on academic literature understanding and review: An empirical study among early-stage scholars. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024), Association for Computing Machinery.
- [107] WU, J., WANG, J., LEI, S., WU, F., AND GAO, X. The impact of metacognitive scaffolding on deep learning in a genai-supported learning environment. *Interactive Learning Environments* 33, 9 (2025), 5166–5183.
- [108] XIE, J., WU, C., WANG, G., YU, R., ZHANG, H., METOYER, R., AND CHEN, S. Understanding parents’ perspectives on responsible ai for children’s self-directed learning. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (2026), pp. 1–20.
- [109] XU, X., QIAO, L., CHENG, N., LIU, H., AND ZHAO, W. Enhancing self-regulated learning and learning experience in generative ai environments: The critical role of metacognitive support. *British Journal of Educational Technology* 56, 5 (2025), 1842–1863.

-
- [110] YAVICH, R. Will the use of AI undermine students independent thinking? *Education Sciences* 15, 6 (2025).
- [111] YEO, S., LIM, G., GAO, J., ZHANG, W., AND PERRAULT, S. T. Help me reflect: Leveraging self-reflection interface nudges to enhance deliberativeness on online deliberation platforms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024), pp. 1–32.
- [112] ZHANG, Y., TAN, L., AND O’HALLORAN, K. L. Bias in genai: Exploring cultural variations between large language models. *Linguistics and Education* 92 (2026), 101515.
- [113] ZHI, J., KUMAR, H., AND LEE, M. Investigating the effects of llm use on critical thinking under time constraints: Access timing and time availability. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (2026), pp. 1–21.
- [114] ZIMMERMAN, J., AND FORLIZZI, J. *Research through design in HCI*. Springer New York, New York, NY, 2014, pp. 167–189.
- [115] ZIMMERMAN, J., FORLIZZI, J., AND EVENSON, S. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), Association for Computing Machinery, p. 493–502.

-
- [116] ZWINDERMAN, M., LEENHEER, R., SHIRZAD, A., CHUPRIYANOV, N., VEUGEN, G., ZHANG, B., AND MARKOPOULOS, P. Using video prototypes for evaluating design concepts with users: a comparison to usability testing. In *IFIP Conference on Human-Computer Interaction* (2013), Springer, pp. 774–781.

Appendix A

Low-Fidelity Sketches



Appendix B

Social Media Recruitment Post

Social Media Recruitment Post

Are you an undergraduate student at the University of Manitoba who has used tools like ChatGPT, Gemini, or Copilot for your academic work?

We'd love to hear your thoughts!

We're looking for people to share their thoughts on how they use Generative AI tools and how certain features might help use these tools more thoughtfully and reflectively.

Participation will take approximately 90 minutes. The session will include completing a short survey, watching short video demonstrations, a card-sorting activity and an interview.

You will receive \$25 (CAD) as compensation for your time!

To get more details about the study, please visit the link below.

[study-details.pdf](#)

📍 Location: University of Manitoba, Fort Garry Campus (EITC Building, Room: E2-486)

If you are interested or have questions, please contact:

Fairuz Shaiara

Graduate Student

shaiaraf@myumanitoba.ca

University of Manitoba

The faculty advisor of the research study, Dr. Andrea Bunt, can be contacted at

Andrea.Bunt@umanitoba.ca.

Appendix C

Recruitment Poster

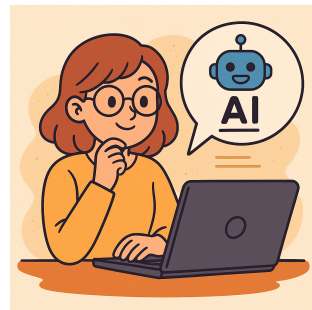


Are you an undergraduate student at the University of Manitoba? Have you ever used ChatGPT, Gemini, or similar Generative AI tools for your academic work?

We're looking for people to share their thoughts on how they use Generative AI tools and how certain features might help use these tools more thoughtfully and reflectively.

Participation involves:

- Complete a short survey
- Watch short video demonstrations
- Take part in a card-sorting activity
- Share your opinions in an interview



Participation will take approximately 90 minutes. You will receive \$25 cash for your time.

If interested, please contact:
Fairuz Shaiara
shaiaraf@myumanitoba.ca

OR



Principal Investigator
Dr. Andrea Bunt
Andrea.Bunt@umanitoba.ca
University of Manitoba

This research has been approved by the Research Ethics Board at the University of Manitoba, Fort Garry campus. The Research Ethics Board can be reached by phone 204-474-7122 or email humanethics@umanitoba.ca

Appendix D

Pre-Study Questionnaire

1. Email Address:
2. What is your current major or field of study? (*e.g., Computer Science, Psychology, Business, Fine Arts, etc.*)
3. Which year of your undergraduate studies are you currently in?
4. Have you used Generative AI tools (e.g., ChatGPT, Gemini, Copilot, or similar) for your academic work before?

Appendix E

Research Ethics Board Approval



University of Manitoba | Research Ethics and Compliance

Human Research Ethics - Fort Garry
66 Chancellors Circle
Winnipeg, MB R3T 2N2
humanethics@umanitoba.ca

PROTOCOL APPROVAL

Effective: December 16, 2025

Expiry: December 15, 2026

Principal Investigator: Andrea Bunt
Protocol Number: HE2025-0373
Protocol Title: *Exploring Interfaces to Encourage Reflective Use of Generative AI*

Liz Millward, Chair, REB2

Research Ethics Board 2 has reviewed and approved the above research. The Office of Human Research Ethics (OHRE) is constituted and operates in accordance with the current *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans- TCPS 2 (2022)*.

Please note the following important information about your protocol approval:

- i. Approval is granted for the research and purposes described in the protocol only.
- ii. Any changes to the protocol or research materials must be approved by the OHRE **before implementation**.
- iii. Any **deviations** to the research or **adverse events** must be reported to the OHRE immediately through an **REB Event**.
- iv. This approval is valid **for one year only**. A Renewal Request must be submitted and approved prior to the above expiry date.
- v. A **Protocol Closure** must be submitted to the OHRE when the research is complete or if the research is terminated.
- vi. The University of Manitoba may request to audit your research documentation to confirm compliance with this approved protocol, and with the UM [Ethics of Research Involving Humans](#) policies and procedures.

Appendix F

Baseline Survey

1. Section 1: Demographics

- (a) Age:
- (b) What is your gender identity?

2. Section 2: Prior Experience with Generative AI Tools

- (a) Which Generative AI tools have you used for academic purposes?
- (b) How frequently do you use Generative AI tools for academic tasks?
- (c) How long have you been using Generative AI tools?
- (d) For which of the following purposes do you typically use Generative AI tools?

3. Section 3: Your thoughts

- (a) Please add any comments about the use of Generative AI tools (including AI chatbots) that you would like to share with the researchers. *[Optional]*

Appendix G

Context Scenario



Computer Science
University of Manitoba
Winnipeg, Manitoba
Canada R3T 2N2

Context Sheet

Please read this short scenario before watching the video demonstrations.
All videos you will see are based on this same situation.

- Taylor is a **first-year undergraduate student**.
- They have a **1500-word essay** due **tomorrow**.
- The assignment carries a **significant portion of their course grade**.
- The essay has multiple parts, including:
 - explaining a concept,
 - giving an example,
 - reflecting on a personal experience,
 - and writing a conclusion.
- Taylor has already done **some work on the assignment**, but not all parts are complete.
- Because of **time pressure**, Taylor decides to use **ChatGPT** to help them work on the assignment.
- Taylor uses ChatGPT for **different kinds of support**, such as:
 - asking for explanations,
 - revising or polishing text they have written,
 - and developing parts of the essay further.
- The following videos that will be demonstrated will show **different versions of a ChatGPT-like interface** that Taylor uses while working on their assignment.

Please keep this scenario in mind while watching the videos and answering the questions.

Appendix H

Interview Guide Used After Each Video Demonstration

H.1 Predict Before Reveal

1. What are your initial thoughts about this prototype? What do you think about the interface in general?
2. If this feature were available while you were using ChatGPT, how do you think it would affect the way you work?
3. Are there any situations where you think this prototype might not work well?
4. How did you feel about having the AI output temporarily hidden before you could see it?
5. How did you feel about the AI response being masked every time you entered a prompt?

6. Do you think this brief pause would change how you engage with AI responses?
If so, how?
7. Is there anything else you would like to add about this interface before we move on to the next one?

H.2 Quick Check

1. What are your initial thoughts about this prototype? What do you think about the interface in general?
2. If this feature were available while you were using ChatGPT, how do you think it would affect the way you work?
3. How useful do you think the weekly report would be to you, if at all?
4. Are there any situations where you think this prototype might not work well?
5. Do you think answering these three statements would influence how you engage with AI responses? If so, how?
6. How did you feel about having to complete the Quick Check before being able to continue interacting with ChatGPT?
7. Is there anything else you would like to add about this interface before we move on to the next one?

H.3 Workload Balance Slider

1. What are your initial thoughts about this prototype? What do you think about the interface in general?

2. If this feature were available while you were using ChatGPT, how do you think it would affect the way you work?
3. Are there any situations where you think this prototype might not work well?
4. What factors would you consider when deciding where to place the slider?
5. How did you feel about having to use the slider after every AI-generated response?
6. What did you think about having the weekly report?
7. Is there anything else you would like to add about this interface before we move on to the next one?

H.4 Workflow Mapper

1. What are your initial thoughts about this prototype? What do you think about the interface in general?
2. If this feature were available while you were using ChatGPT, how do you think it would affect the way you work?
3. Are there any situations where you think this prototype might not work well?
4. How did you feel about distinguishing between your own contributions and AI contributions using different types of notes?
5. Do you think creating this kind of visual map would influence how you plan or structure your work? If so, how?
6. Is there anything else you would like to add about this interface before we move on to the next one?

Appendix I

Interview Guide Used at the End of the Study

1. Which of the factors felt more important to you when ranking?
2. Can you walk me through how you decided on this ranking?
3. Looking across all the designs, do you think any of them would change how you use AI tools for your work?
4. Overall, how do you feel about the idea of integrating such scaffolding features into AI tools like ChatGPT?
5. If you could improve or modify any of the designs, what changes would you suggest?
6. Based on your experiences today, what do you think makes an AI interface “reflective” or helpful for deeper learning?

7. Is there anything else you would like to share about your experience, thoughts, or suggestions related to these designs or the study?