
Degree Project in the Field of Technology Media Technology and the Main Field of Study Computer Science and Engineering

Second cycle, 30 credits

# WriteFlow: LLM-Assisted Goal Setting for Reflective and Critical Academic Writing

**YUELING FAN**

# WriteFlow: LLM-Assisted Goal Setting for Reflective and Critical Academic Writing

YUELING FAN

Master's Programme, Interactive Media Technology, 120 credits
Date: December 9, 2025

Supervisor: Olga Viberg
Examiner: Anders Hedman
School of Electrical Engineering and Computer Science
Swedish title: WriteFlow: LLM-assisterad målsättning för reflekterande och kritiskt akademiskt skrivande

# WriteFlow: LLM-Assisted Goal Setting for Reflective and Critical Academic Writing

YUELING FAN

Academic writing requires iterative reflection and dynamic goal regulation. However, both prior research and findings from our formative study (N = 17) reveal that students often struggle to formulate and manage evolving writing goals–predictive to their writing success–throughout the writing process. While many commercial LLM-based writing tools emphasize efficiency, they raise concerns about diminishing students' autonomy and critical engagement. We present WriteFlow, an AI voice-based writing assistant that facilitates reflective dialogue and provides goal-setting assistance in the academic writing process. WriteFlow leverages the interactive learning experience as a space for continuous reflection and goal negotiation based on the writer's writing intentions, allowing students to retain autonomy and deepen their involvement with the writing process. In our evaluation study involving six students specializing in human–computer interaction, participants reported that WriteFlow supported reflection-in-action and metacognitive regulation while helping them maintain their writing intentions. We discuss implications for future systems, including hierarchical visualization of writing goals and prompt designs that encourage multi-perspective reflection.

## SAMMANFATTNING

Akademiskt skrivande kräver iterativ reflektion och målreglering. Resultaten från vår bakgrunds- och formativa studie (N = 17) visar dock att studenter har svårt att sätta upp och hantera föränderliga skrivmål under skrivprocessen. Många kommersiella LLM-förbättrade skrivverktyg fokuserar på effektivitet, vilket väcker farhågor om deras inverkan på studenternas skrivautonomi och kritiska tänkande. Vi presenterar WriteFlow, en röstbaserad skrivassistent som syftar till att främja reflekterande dialog och hjälper till med målsättning. WriteFlow använder chattgränssnittet som en arena för kontinuerlig reflektion och målsättning utifrån författarens skrivintentioner, vilket stödjer studenters självständighet och fördjupar deras engagemang i skrivprocessen. I vår utvärderingsstudie med sex studenter som specialiserat sig på människa-datorinteraktion, rapporterade deltagarna att WriteFlow stödde reflektion i handling och metakognitiv reglering samtidigt som det hjälpte dem att behålla sina skrivintentioner. Vi diskuterar implikationer för framtida system, inklusive hierarkisk visualisering av skrivmål och promptdesign som uppmuntrar reflektion ur flera perspektiv.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Interactive systems and tools**.

Keywords: Large Language Models, Intelligent Writing Assistant, Goal Setting, Critical Thinking, Conversational Agent, Human-AI Collaboration

Nyckelord: Stora språkmodeller, intelligent skrivassistent, målsättning, kritiskt tänkande, konversationsagent, samarbete mellan människa och AI

Author's address: Yueling Fan, yueling@kth.se, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, Stockholm, Sweden, SE 100 44.

Document date: December 9, 2025

Swedish title: WriteFlow: LLM-assisterad målsättning för reflekterande och kritiskt akademiskt skrivande

## 1 INTRODUCTION

Academic writing is regarded as a tool for students to consolidate knowledge. Scholars highlight that the process of knowledge transformation occurs when students address the dialectic between rhetorical problems (i.e., how to present it) and content-related problems (i.e., what to present) [9]. Through this process, students transform tacit or experiential knowledge into explicit and transferable understanding [60]. Academic writing also cultivates students' strategic knowledge [29], namely the ability to set manageable goals with high-level procedural knowledge, and monitor and regulate the complex writing process, which can be transferred to their future research and professional work.

While academic writing requires sustained reflection and metacognitive regulation, recent advances in large language models (LLMs) are reshaping how students engage with these tasks. LLMs such as ChatGPT demonstrate notable reasoning and open-ended text generation capabilities [7, 64], and are increasingly integrated into students' learning and research practices [33, 38]. Although they promise personalized support and scalable tutoring, concerns remain that long-term reliance can undermine students' learning abilities [20, 22, 27]. LLMs are particularly effective in producing fluent and general-purpose responses, but often lack sensitivity to the current needs, stage of understanding, and epistemic goals of the learner [33]. Over-reliance risks cognitive offloading, where learners delegate effortful processes to AI rather than exercising active thinking, reflection, and reasoning [22, 26]. This risk is most acute in academic writing, a domain intended not only to demonstrate knowledge but also to cultivate critical thinking and iterative reflection. Writing is a recursive, cognitively demanding process: students must set and revise goals, integrate ideas across topics, and continuously reflect on arguments. Yet, current widely used commercial LLM chat interfaces are optimized for linear, sequential dialogue, effective for single queries, but ill-suited for non-linear tasks like writing. Switching between multiple topics or revisiting earlier reasoning is cumbersome, limiting opportunities for critical thinking, which not only fosters self-regulation but also predicts academic success and deeper understanding [53]. As a result, students may remain in surface-level interactions rather than exercising deeper self-regulation in the writing process.

As LLMs tools are increasingly being integrated into students' academic practices, they are not only changing how students complete writing tasks but also how they employ writing strategies and regulate their learning. This makes it essential to investigate how such tools could be intentionally designed to support students' critical thinking and reflection during writing. Recent empirical studies have begun aligning LLM-based tools with self-regulated learning (SRL) frameworks (e.g., [36, 41, 63]) to promote students' self-evaluation and self-reflection in writing. SRL emphasizes learners' active regulation of goals, strategies, and reflection across forethought, performance, and self-reflection phases [47, 57]. Goal setting is central: it guides planning, supports self-monitoring, and maintains motivation [71]. However, few studies have investigated how AI systems might scaffold reflection and goal adjustment in the writing process, where goals are layered and evolve continuously. A notable example is VISAR [70], which supports iterative planning and content goal management but remains limited in supporting process goals and in tracking or evaluating goal progress. Parallel work in HCI explores alternatives to linear dialogue, such as non-linear conversational systems like Mindalogue [68], which support multi-threaded engagement. However, these designs risk overwhelming users with complexity and disorientation [50, 51, 68]. Similarly, AI writing assistants [13, 32, 49] often prioritize efficiency and textual output, focusing on isolated stages of the writing process (e.g. ideation [59, 70] or revision [45, 67]) with less emphasis

on cultivating critical reflection, promoting learner agency, or addressing the issue of over-reliance on AI assistance.

In this study, we move beyond efficiency-driven AI support for writing to explore how LLMs can be designed to scaffold reflection, goal setting, and critical thinking in the academic writing process. We introduce WriteFlow, a Google Docs-based writing assistant that scaffolds students' goal-setting and reflection throughout the writing process. Prior work identifies conversational interaction as a key design resource for fostering reflection [8, 46]. Voice-based dialogue has been shown to support more spontaneous, personal, and emotionally engaging engagement that promote both intuitive and critical self-reflection [35]. Our system reimagines chat interaction as a space for reflection-in-action [54] and goal negotiation, helping students retain autonomy while deepening engagement with their own ideas and drafts. Through voice-based interaction, writing goal generation, SMART-based evaluation, and alignment tracking, WriteFlow supports the iterative construction and realignment of writing goal networks.

We conducted a formative study (n = 17) to investigate writing challenges students face and their current use of LLMs in academic writing. Based on these insights, we then designed WriteFlow and evaluated it with six users in a Wizard-of-Oz study in Fall 2025. Our results demonstrate that WriteFlow promotes structured thinking, sustained reflection, and greater authorial control. However, we also identify the need for offering multiple perspective suggestions to encourage students'to think from different angles and explore alternative ideas. In summary, our contributions include:

- A formative study demonstrates writing goal-related stress, and how students currently use LLMs to support academic writing.
- WriteFlow, a voice-based writing assistant that scaffolds critical thinking by supporting iterative construction and monitoring of writing goals through reflective dialogue.
- Findings from a Wizard-of-Oz evaluation study showing how WriteFlow scaffolds reflection-in-action, metacognitive regulation, and structured thinking.
- Design implications for designing human-AI writing systems that maintain writer agency and support evolving goal network.

## 2 BACKGROUND

### 2.1 Goal Setting for Self-Regulated Learning and Academic Writing

Goal setting is a fundamental element and is integrated across the different phases of SRL: *forethought* (i.e., set specific goals and plan to use certain strategies, and activate self-motivational beliefs), *performance* (i.e., implement strategies, maintain attention, and monitor progress toward goals), and *self-reflection* (i.e., evaluate performance, attribute outcomes, and adjust future goals or strategies) [71]. Through goal setting, students engage in metacognitive planning, which involves articulating objectives, anticipating cognitive demands, and selecting reasonable strategies that support task execution [47]. Making specific, proximal and appropriately challenging goals [52, 55] is recognized to promote metacognition because it not only fosters self-monitoring and evaluation [55], but also enhances students' self-efficacy, motivation, and achievement [14, 55].

In academic writing, goal setting facilitates students to regulate their writing process including how to address rhetorical problems, time management and progress evaluation. From the cognitive perspective, the writing process is guided by a hierarchically structured network of goals [25, 29]. Students always need to manage multiple abstract goals (e.g., changing audiences' attitudes) and concrete goals (e.g.,

proving a claim by giving examples). Compared to novices, experienced writers tend to generate more complex and elaborated goal networks, including goals, sub-goals, strategies, and corresponding self-evaluation feedback. Moreover, they are good at establishing logical and strategic connections among these goals and can flexibly adjust their goal networks according to task constraints and their goals [29].

Research has shown that devoting more effort to the planning process and setting quality-based goals can improve the quality of written texts [6]. Planning can contribute to building a network of goals and can further stimulate more purposeful autonomous revision behaviors later in the writing process, such as addressing inconsistencies between the written text and the intended plan, or identifying and improving the weak point in the entire logical flow. Thus, the revision behaviors are not limited to the focus from "superficial changes" (e.g., typos) but extend to the higher-level logical structure ((e.g., strengthening the logic flow of arguments across sections).

### 2.2 Effective Goal-Setting Interventions

To approach effective goal setting interventions, Schunk et al. compared the influence of product goals versus process goals on writing performance [56]. The results show that students who received process goals combined with progress feedback performed better than process-only, product goal, and general goal (instructed only to work productively) groups on the post-test, measuring strategy learning, writing skills, and self-efficacy. Product goals, which focus on writing outcomes (e.g., completing a paragraph), appear to lead students to focus on results and neglect strategy use. Process goals, however, highlight strategy use as a means to improve writing, fostering students to retain their skills after training and transfer to other writing tasks [56].

This study applies the SMART framework [19] to evaluate the appropriateness of goal setting during the writing process in the LLMs-mediated writing context. The SMART framework is not merely an effective benchmark aligned with appropriately challenging goals [55], but also facilitates critical reflection of the process objectives [16, 39]. The SMART criteria were first proposed to emphasize that effective goals should be specific, measurable, attainable, relevant, and time-bound [19]. 'Specific' goals, differing from general goals, promote strategy learning and specific effort in writing, allowing students to compare and analyze which strategies to use [56]. Elaborated goals have been demonstrated to help students write more persuasive essays with more argumentative elements [23]. 'Measurable' goals refer to goals that can be assessed, enabling the formation of goal progress-oriented feedback that informs learners about the usefulness of selected strategies [48]. 'Achievable' goals suggest setting writing subgoals that are realistic, challenging but attainable. According to Schunk, 'proximal' goals result in greater motivation than distant goals [55]. 'Relevant' goals help maintain cognitive and strategic focus on central arguments. Relevance is also essential for ensuring that individual writing goals are logically connected within a coherent goal network. Time-bound goals promote task initiation and self-observation, giving students a sense of priority in writing.

Recent studies applied AWE (Automated Writing Evaluation) to provide individualized goal-setting interventions to address specific weaknesses in students' writing [31, 62]. However, goal setting through AWE is typically based on feedback regarding specific textual features related to writing [17, 58], lacking evaluation of deeper content-level aspects such as the intended purpose behind content choices. This study aims to explore how LLMs can help students to establish and sustain individualized goal networks, encouraging them to reflect on their work and develop new ideas during the writing process.

### 2.3 Critical Thinking and Academic Writing

According to Flower and Hayes' cognitive process model of writing [25, 29], writing is goal-oriented and involves three main processes — planning, translating, and reviewing. Planning involves setting writing goals and organizing thoughts. Translating refers to converting these ideas into written text, while reviewing includes evaluating and revising what has been written.

These processes are not isolated stages but instead recursive and non-linear [25, 29]. Any process can be embedded within another. The writing goals evolves continuously throughout the entire writing process. For example, revising can prompt the writers to reset their goals. This model highlights that students must engage in continuous reflection in academic writing, as they consider how to retrieve knowledge from long-term memory to address rhetorical problems, how to shift from local goals to global goals, and how to set or revise goals more reasonably.

Critical thinking is "reasonable reflective thinking focused on deciding what to believe or do" [21, p. 5]. Within the context of academic writing, this process involves both reflective engagement with disciplinary knowledge and critical self-evaluation of their behaviors throughout the writing process. To better understand the nature of reflective thinking in writing, Fleck and Fitzpatrick [24] proposed a framework consisting of different "levels of reflection," categorizing reflective activity from R0 to R4. At the most basic level, R0 involves non-reflective description, while R1 represents foundational reflection that involves revisiting knowledge with limited reasoning or justification. R2 starts to look for relationships between pieces of knowledge through questioning and the exploration of alternative perspectives. The R3 level represents transformative reflection, where previous assumptions are reframed in light of new insights. The highest level, R4, encompasses critical reflection about wider implications that incorporates social, ethical, and contextual awareness.

In academic writing, these levels manifest as a move from surface-level description of information to a higher-level critique that questions the original assumptions and contextual implications. In terms of writing behavior, students may begin by understanding their writing process for writing tasks, then gradually move toward actively setting and managing goals in the writing process, and eventually reflect on how factors, such as audience, genre, or academic conventions, influence their decisions.

Rapidly emerged AI writing assistants such as Riley [13], Anglekindling [49] can help students think more reflectively about content writing [45]. However, as elaborated in the next section, most existing applications adopt a product-oriented approach [32, 34, 66], focusing primarily on the fulfillment of writing tasks and efficiency. Few studies and commercial products pay attention to designing for supporting users' cognitive and metacognitive processes.

### 2.4 The Impact of LLMs on Motivation and Cognition

Scholars posit that integrating LLMs tools, such as ChatGPT or CoPilot, into educational environments can have a positive impact on students' learning motivation. In a recent study, the authors have observed that students in flipped classrooms who engaged with an LLM-based chatbot demonstrated significantly greater learning motivation and better learning performance, as measured by quiz scores and a questionnaire assessing intrinsic motivation and self-efficacy [69]. According to Wei et al.'s study [61], LLM-mediated language instruction has been shown to enhance language learning outcomes and second language (L2) students' motivation by offering a personalized educational experience. By providing personalized feedback and immediate support, LLMs has been shown to effectively improve students' time management and enhance their sense of control over their learning processes [11, 43]. In addition, these

tools are also effective in helping students reduce social anxiety. That is, LLM tools such as ChatGPT provide a low-pressure environment that makes it easy for students to seek direct assistance [28, 43], which can further enhance students' learning autonomy.

However, a recent study by Fan et al. [22] challenges the positive impacts reported on student motivation and academic performance. The results show that although ChatGPT can improve students' short-term learning performance, it does not significantly enhance intrinsic motivation. Instead, long-term reliance on LLMs may cause cognitive offloading and 'metacognitive laziness' [22, 26], manifesting as a tendency to delegate thinking tasks to external tools and reduce their self-monitoring and regulation processes. They may neglect deeper cognitive engagement, such as questioning, analyzing, or integrating knowledge. Especially when they are looking to utilize LLMs to help complete tasks in a limited time more quickly, they may no longer put effort into thinking and understanding the reasoning process behind the answer. This over-reliance risks undermining the educational purpose of learning tasks.

Yet, we argue that using LLMs tools such as ChatGPT in a balanced and intentional way can help maintain critical thinking skills. For example, leveraging such tools to handle tasks such as memory retention and information retrieval can reduce students' cognitive load and create space for higher-order thinking [26]. Another observation is that promoting students' self-confidence can encourage them to think critically in human–AI interactions, while overly trusting AI systems can lead to the opposite effect [40]. Maiti et al. demonstrated that combining LLMs with pedagogical constraints can effectively foster deeper reasoning. Their LLMs system, Jill Watson [44] has been shown to effectively encourage students to ask more higher-order and critical questions than in a traditional classroom setting. Since the impact of LLMs on critical thinking largely depends on how they are used, we are motivated to design writing support systems that encourage intentional and reflective engagement with LLMs, while enhancing students' self-confidence and capacity for deeper reasoning throughout the writing process.

### 2.5 Intelligent and Interactive Writing Tools

Recently, AI-powered writing tools have been a focus of exploration by the HCI community. These writing assistants address different tasks at different stages of the writing process. Studies focused on supporting writing literature review have explored how to design systems to help with critical reading and defining problems. LitWeaver [13], for example, supports the entire workflow from paper review, topic finding to paragraph writing. It inspires users to think about why to quote certain content by offering examples, and highlights vague expressions to improve clarity. The system was found to deepen users' engagement with the literature and increase their autonomy and self-confidence in writing. PAPER PLAIN [4] has been demonstrated to enhance readers' understanding and confidence when engaging with medical research papers. It provides key questions to guide readers answer passages so that users can read with questions in mind.

Tools designed for ideation in the writing process focus on assisting in brainstorming writing angles and organizing thoughts. AngleKindling helps journalists uncover unconsidered entry points for writing reports by interpreting information from different angles [49]. Luminate enables users to steer the direction of AI-generated brainstorming toward particular dimensions [59]. Kim et al. introduced a personal journal writing assistant, DiaryMate [34], capable of generating the next sentence based on user-provided keywords. Although these tools can effectively inspire users when they are 'stuck' in the writing process, they may also lead to an overreliance on AI-generated examples and ideas. Node-based

interactive systems such as VISAR [70] have also been applied to aid users in organizing the relationship of complex, hierarchical thinking required in the writing planning stage. However, this interactive method was shown to constrain users from actively exploring more complex argumentation structures, and it also introduced unnecessary mental overhead when users elaborated on their ideas [51, 70]. While these systems demonstrate the potential of AI to support idea exploration and organization, they offer limited support for guiding users to continually reflect on their writing goals and evaluate how their ideas align with them. This could undermine students' self-confidence and autonomy during the writing process.

Other tools specifically focus on providing and analyzing feedback to support the revision process. The studies by Karolus et al. [32] and E et al. [66] demonstrated that in-action feedback, also called continuous feedback, was superior to revision-based feedback in enhancing student writing performance and task fulfillment. However, in-action feedback also risks users overly relying on AI-provided feedback instead of engaging in more holistic self-evaluation [66]. In addition to providing feedback, Friction [67] assists users in developing actionable revision goals by clustering and analyzing human feedback. It offers AI evaluations of the quality of each revision and explains why it was or was not effective. Recent systems have also explored ways to leverage AI-driven feedback to support students' self-regulatory engagement. For instance, Neshaei et al. developed ALure [45], which leverages students' errors to provide tailored feedback. Dang et al. applied a reverse outlining strategy by only summarizing users' writing content to help users reflect on their writing content [18].

Building on these developments, advances have also introduced multimodal support into writing assistance systems. Rambler [42] assists in capturing loose thoughts through speech-to-text input, organizing them into modular units that can be merged, split, and revised. NotebookLM [2], an AI-powered research assistant from Google Lab, can support simultaneous uploads of up to 20 articles and generate podcasts based on these sources. Users can ask questions for the uploaded files either in the chatbot or by joining in the podcasts to ask the two 'hosts'. The response in the chatbot will clarify which uploaded source and which part of it the answer comes from. Nonetheless, empirical observations indicated that users engaging with multimodal interfaces experienced difficulties in managing multiple tasks [65].

Many AI writing tools have improved critical engagement by providing examples and enabling users to compare different viewpoints rather than directly offering answers. However, recent research suggests that such designs may also shift the writer's role toward making editorial decisions rather than engaging deeply in content creation [10]. Few studies have examined the long-term effects of these tools on the development of students' independent reflective writing skills and learning performance. ChatTutor [12] attempts to address such concerns by providing adaptive guidance based on course plans and students' individual learning profiles. Similarly, Khanmigo Writing Coach [3] first helps students understand the assignment requirements and goals before offering guidelines for writing a strong paper. Additionally, Hoque et al. propose HaLLMark [30], which visualizes the interaction timeline and the proportion of AI-generated content to foster greater student agency in AI-assisted writing.

Despite these developments, few studies have examined how LLMs can support the iterative management and regulation of writing goals in academic writing. Goal-setting not only directs planning and monitoring but also allows students to compare emerging ideas with their initial plans, connecting their reflective revision with writing intentions. However, most current systems focus on surface-level assistance in an isolated writing process (e.g. planning or revision) without supporting reflection across different writing processes. Building on these insights, we present a writing assistant that can adaptively

support reflection on writing goals, helping students sustain autonomy and critical engagement throughout the academic writing process.

## 3 FORMATIVE STUDY

### 3.1 Data Collection and Procedure

We conducted an online survey with 17 university students to investigate how they use AI tools to handle writing challenges, and what these challenges are. The complete questionnaire can be accessed here . Both qualitative and quantitative data from 17 university students were collected. A requirement for the participation in the study was to have experience with: i) academic writing in English and ii) using ChatGPT to support them in their writing process. The demographics of the study participants are presented in Table 2.

The survey began with a written consent form, followed by factual questions about students writing experience (e.g., frequency usage of ChatGPT for writing, years of academic writing experience, types of academic writing completed, and typical approach to academic writing). The participants were then asked to rate 10 writing challenges (see Table 1) on a 5-point Likert scale, showing how stressful they thought each was (1 - not stressful at all, 5 - extremely stressful). These challenges were informed and conceptually structured based on Flower and Hayes's cognitive process theory of writing [25, 29], reflecting both internal cognitive demands and external task constraints in the writing process. For example, the challenge of *Understanding the writing task* was derived from their discussion of the rhetorical problem, while *Setting writing goals* reflects how writers define and approach the task by organizing knowledge around a problem [29]. The difficulties related to *Losing the big picture* and *Evolving writing goals* reflect the dynamic restructuring of hierarchical goal networks during writing, as writers continuously adjust higher-level and sub-goals [29]. Challenges like *Over-focusing on low-level goals* and *Language proficiency* align with the resource allocation issues highlighted in their description of the cognitive demands involved in translation processes (sentence generation) [25]. Similarly, *Misalignment between content and intent* and *Difficulty interpreting feedback* connect to writers' struggles in evaluating and revising text against their intended meaning [25, 29]. Finally, issues concerning *Academic conventions* and *Not meet assignment or audience expectations* reflect the external task environment [29].

For each rated challenge, participants were asked to describe in an open-ended way (open text box) how they usually do to move forward to address these challenges. And if using AI tools like ChatGPT, how they use them, and whether they are helpful to support them in solving particular writing challenges. The survey concluded with several open-ended questions including any additional writing challenges, their positive experience with writing, concerns about using AI and expected AI support in the writing process. The online survey was conducted via Google Form between May 22 and May 27, 2025 and took approximately 30–40 minutes to complete. All responses were automatically collected and exported to a linked Google spreadsheet for analysis.

Qualitative responses were analyzed using Reflexive Thematic Analysis [15] from a data-centric perspective (Table 3). This method emphasizes flexible, iterative coding of raw responses while remaining close to the participants' own phrasing and lived experiences. In addition, the rating data served as a supplementary reference to help interpret the attitudes of students towards different writing challenges and to contextualize the qualitative themes. In total, 17 students responded to 15 open-ended questions each. During data cleaning, we excluded 12 blank submissions and 38 answers consisting solely of "N/A"

| Writing Challenge | Description |
|---|---|
| Understanding the writing task (Rhetorical Problem) | You found it difficult to understand what a writing task is really asking for-such as its purpose, audience, or constraints. |
| Setting Writing Goals (Content and Process Goals) | You found it difficult to decide what to write (content goals) or how to go about writing it (process goals) when planning your writing task. |
| Academic Conventions | You felt stressed or confused by academic writing requirements such as formatting, referencing style, or formal tone. |
| Losing the Big Picture | As your text gets longer, you find it hard to maintain a global perspective and stay focused on the overall structure and argument. |
| Evolving Goals During the Writing Process | You started writing with a clear goal, but later - after reading more papers or developing your ideas - you realized your original writing goals no longer worked. |
| Over-Focusing on Low-Level Goals | You found yourself spending too much time elaborating on examples or details, which distracted you from your main argument. |
| Language Proficiency & Precision | You felt that limitations in language fluency made it hard to express your ideas clearly, precisely, or in a way that met academic expectations. This may have even distracted you from your main writing goals. |
| Misalignment Between Content and Intent | You felt that what you wrote didn't match what you originally intended to say, and it was difficult to recognize this misalignment. |
| Not Meeting Assignment or Audience Expectations | You realized mid-way through writing that your text didn't meet the expectations of the assignment or the intended reader, and felt unsure how to adjust. |
| Difficulty Interpreting Feedback | You received feedback and found it hard to understand or unclear in terms of what changes were needed, leaving you unsure how to act on it. |

**Table 1: Writing Challenges Adapted from Hayes et al., 1986 [29]**

| | n | % | | n | % | | n | % |
|---|---|---|---|---|---|---|---|---|
| **Gender** | | | **Age** | | | **AI Use Frequency** | | |
| Male | 8 | 47.1 | 22–24 | 11 | 64.7 | Very Often | 7 | 41.2 |
| Female | 8 | 47.1 | 25–27 | 5 | 29.4 | Often | 8 | 47.1 |
| Non-binary | 1 | 5.9 | 34 | 1 | 5.9 | Occasionally | 2 | 11.8 |
| **Program** | | | **Writing Types** | | | Rarely | 0 | 0.0 |
| Civil Engineering | 1 | 5.9 | Essays | 15 | 32.6 | Never | 0 | 0.0 |
| Communication Systems | 1 | 5.9 | Thesis | 14 | 30.4 | **Writing Experience** | | |
| Human-Computer Interaction | 10 | 58.8 | Research papers | 8 | 17.4 | >5 years | 3 | 17.6 |
| Management | 1 | 5.9 | Conference papers | 6 | 13.0 | 3–5 years | 12 | 70.6 |
| Mechanical Engineering | 1 | 5.9 | Lab/project reports | 1 | 2.2 | 1–2 years | 1 | 5.9 |
| Machine Learning | 1 | 5.9 | Gymnasiearbete | 1 | 2.2 | <1 years | 1 | 5.9 |
| Robotics | 2 | 11.8 | Experiment reports | 1 | 2.2 | No Experience | 0 | 0.0 |

**Table 2: Participant Demographics and their writing experience**

or other non-substantive placeholders, leaving 205 valid responses. These responses provided valuable design insights to inform the design of a human-AI-assisted writing system.

## 3.2 Findings

*3.2.1 Managing writing goal-related stress through documentation or outline.* *Evolving writing goals* was identified as the most stressful challenge with the listed issues, followed by *Setting writing*

| Theme | Sub-theme | Codes (# of Participants Mentioned this) |
|---|---|---|
| **Describing the writing experience** | Documentation | Writing down relevant ideas/keywords/sources (6), Listing writing goals based on assignment requirements (2), Coping content from other papers (2), Recording experiment data (1), Writing comments for literature (1) |
| | Well-established Outline | Outlining based on sections and subsections (2), Covering all intended arguments (2), Referring back to outline to stay on track (2), Continuously revise the outline as the project evolves (1) |
| | Quality of academic writing | Originality (8), Well-supported arguments (8), Time management and procrastination (6), Logical consistencies (5), Flexible and unstructured (2), Fluency (2), Accuracy (2), Objective (1) |
| **Coping with challenges using ChatGPT** | Interpreting the writing task | Rephrasing the writing task (9), Checking for missing points in task understanding (1), Clarifying rhetorical situation - purpose and audience (1), Asking for examples of similar tasks (1), Discussing the main purpose of the task (1) |
| | Managing content goals | Exploring writing angles and idea-bouncing (7), Evaluating the relevance of intended writing content (3), Balancing content scope (1), Checking logical flow (1), Generating content directly (1), Building connections between different viewpoints (1) |
| | Managing process goals | Summarizing and clearing up the written text (4), Comparing written content and outlines to check topic drift (2), Prioritizing multiple writing sub-tasks (1), Clarifying high-level writing goals (1), Assessing written content to fit with audience (1), Simulating reviewer feedback (1) |
| | Addressing goal shifting | Asking for new direction and writing goals (5), Reframing outline (3), Assessing the revision workload (2), Saving existing work (2), Comparing different writing goals (2) |
| | Translating idea into text | Rephrasing with more academic tone (11), Translating and polishing (7), Checking for misalignment between written content and intended content goals (4), Checking for factual errors or unclear points (2), Checking content coherence (1), Condensing certain content (1), AI Modifying content directly based on feedback (1) |
| | Addressing external info | Clarifying human feedback and required changes (8), Summarizing and Clarifying information from other sources (4), Providing examples of changes based on feedback (1), Locating feedback-related written content (1) |
| | Concerns with AI | AI altering intended meaning (4), Fabricated content (4), Harming personal thinking and expression abilities (4), Authorship (3), "AI-flavored" words (3), Academic integrity (2), Generic content (1), Distractions from focusing on key arguments (1), Homogenization (1) |

**Table 3: Hierarchy of Themes, Sub-themes and Codes**

*goals* (Figure 1). While both challenges relate to goal management in writing, the former concerns difficulties in articulating initial plans, whereas the latter reflects the need to revise those plans as new ideas or conflicts arise. Changing writing goals is a common and frequent activity during the writing process, as students usually generate new ideas when reading literature or adjust their research focus as research progresses. As P3 mentioned, "With more and more literature read, I can feel lost and question a lot my previous writing and my goal." However, P13 expressed a preference for delaying writing certain sections in detail until obtaining initial results. Other students shared that they prefer to draft after conceptual clarity. For instances, P1 described, "if I'm not feeling inspired, I don't start writing drafts." P12 added: "Once I feel I have a solid grasp of the key concepts and previous work, and a clear logical chain in my mind, I begin outlining the Related Work section in Overleaf." In response to these evolving goals, students prefer to seek assistance from ChatGPT to redirect their writing. Specifically, they used

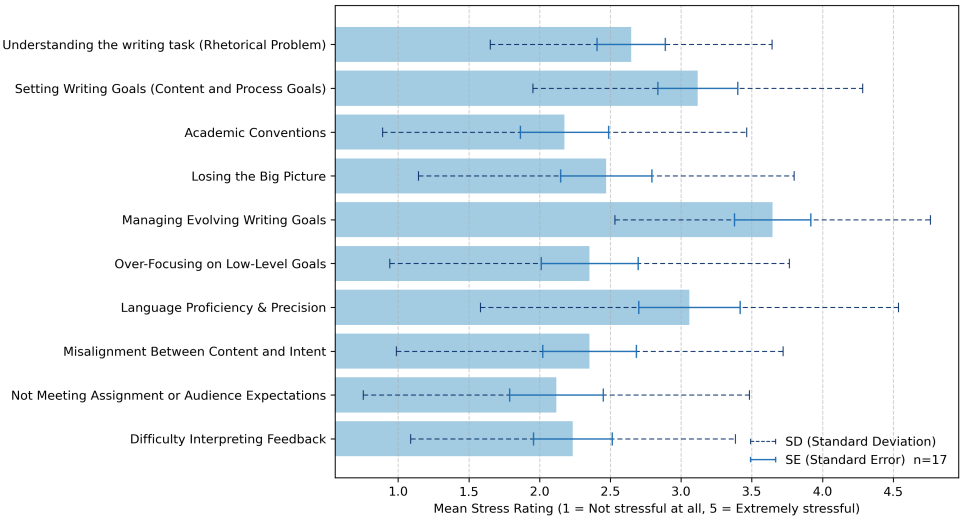


**Figure 1: The result of stress ratings for 10 Writing Challenges (with SD&SE)**

ChatGPT to evaluate and test whether new angles and structural changes would work while attempting to preserve as much of their existing written content as possible (P6, P7, P14, P15).

Some students use documentation or outline as a preparatory strategy to manage evolving writing goals, allowing them to compare new goals with their previous thinking. An obvious pattern observed from the data is that students like to record and note fragmented thoughts during the process of searching for ideas and brainstorming before writing. Six participants described this process as relatively pleasant and less stressful experiences. The notes usually include any interesting thoughts, summary and generated ideas from related literature, relevant reference, listed key points and expectations broken down by the assignment requirements, etc. These materials could serve as a foundation for constructing initial outlines and as a reference point to maintain consistency. Furthermore, participants find that having a well-established outline also improves their interactions with ChatGPT, as it allows the LLMs to better interpret their intended direction and writing plans. As P6 shared, "To stay on track, I often create a detailed outline before writing and refer back to it as I go. If I feel the argument is drifting, I use ChatGPT to help me summarize each section and check if they still support my main thesis. This helps me keep the overall structure clear and consistent."

*3.2.2 'Human-in-the-loop' approach for managing writing stress.* The majority of the students' use of ChatGPT reflects a 'human-in-the-loop' approach, where AI is employed not to replace their writing practice but to provide inspiration and directions, as well as to ensure alignment with task requirements. In addressing writing goal-related stress, students often used AI to clarify task requirements and difficult technical terms in a simpler way, brainstorm ideas together with ChatGPT, evaluate whether their understanding aligned with assignment expectations. For text-related stress, students used ChatGPT to polish language and check the coherence of the content. Some students further utilize ChatGPT to evaluate which ideas deserve emphasis or can be omitted, helping them keep their writing focused on central arguments. In managing revision-related stress, students use ChatGPT to locate specific feedback points, clarify reviewers' comments, and preserve text during major revisions. Besides, LLM supported

summarization shows helpful in reducing the cognitive pressure of processing large amounts of information. As P6 noted, he prefers to use ChatGPT to summarize written sections and verifying alignment between the evolving drafts and original outlines.

While many participants embrace AI assistance for translating ideas into text, they emphasize the importance of retaining their own words and thoughts. The result shows that these students prefer to use ChatGPT to rewrite text in an academic tone without changing the original meaning. Concerns have been raised about AI-generated "AI-feeling" words, and the risk that ChatGPT's direct translation may result in unintentional textual derivation. Several participants (P1, P3, P6, P13, P14) also worried that the extensive use of AI tools could compromise the authenticity of their academic writing, making it unclear whether the content actually originated from their own thinking. This concern highlights the importance of human judgment and discernment when integrating AI outputs into their academic writing practices.

### 3.3 Design Takeaways

Based on the findings from the formative study, we identified the following design goals (**DG**s):

- **DG1. Support Iterative Goal Management and Goal Network Construction.** Students often experience stress not only in setting writing goals, but also in revising goals as their ideas evolve during research. Therefore, the system should support flexible iterative refinement of writing goals throughout the writing process, allowing users to revise, reflect on, and realign goals with their writing directions. To reduce stress, it should also make it easy to preserve and compare evolving goals. The system should help users maintain alignment between evolving drafts and original goals, facilitating reflective refinement and reducing revision stress.
- **DG2. Support Recording and Organizing Emerging Ideas During Writing.** Participants perceive the recording of fragmented ideas, outlines, and literature notes as relatively low-pressure processes during writing. These activities both support initiating writing by helping them explore inspiration and enable them to actively engage in reflection, such as comparing different arguments and structuring their ideas. To support this, the system should provide structured templates, voice interaction or outline prompts to capture their emerging thoughts. It should also help organize and categorize users' previously recorded ideas, and encourage them to revisit relevant thoughts.
- **DG3. Preserve Voice and Ownership in AI-assisted Writing.** Many students voiced concern over the loss of personal tone and academic authenticity. They expressed the need to maintain authorial control and avoid AI overreach. To address this, the system should prioritize preserving the writer's original meaning and voice, offering suggestions that are always traceable, revisable, and tied to the user's writing intentions.

## 4 SYSTEM DESIGN

Figure 2 shows WriteFlow's user workflow and interface. Users start by inputting writing requirements (Figure B1) and uploading Google Docs (Figure A). All information serves as input data, enabling the AI assistant to comprehensively understand the writing task's context. By activating Voice Mode (Figure A), users initiate a conversation with the voice agent to discuss their writing plan. The agent generates writing goals based on their writing intentions (Figure C2). When users accept the goal list, WriteFlow self-evaluates these goals based on the SMART framework to make sure these goals are challenging but reasonable (Figure C4). Accepted goals are added to the My Goals page (Figure D), where users can

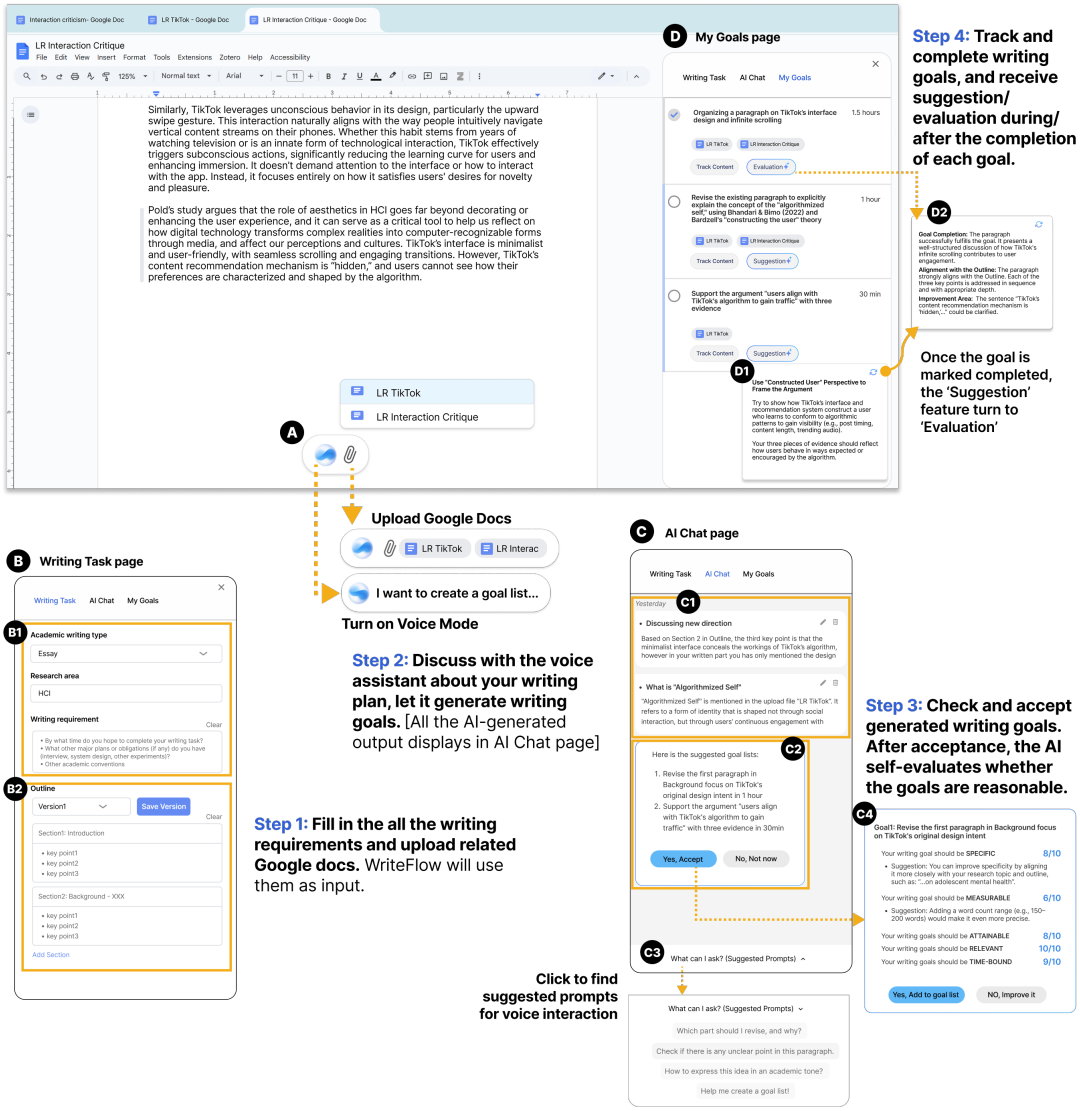


**Figure 2: The system overview of WriteFlow, a Google Docs add-on for goal-oriented academic writing. WriteFlow interface consists of a voice agent (A) and a sidebar panel with three pages: Writing Task (B), AI Chat (C), and My Goals (D). Users can upload Google Docs and communicate with the voice agent at any stage of writing to discuss their writing directions. The agent then generates writing goals to help them plan, track, and monitor their writing process.**

track their progress and receive suggestions (Figure D1) during writing. Upon completion of the writing goal, the system provides evaluations (Figure D2) of goal achievement. The Outline (Figure B2) allows users to create section-based outlines and save different outline versions throughout the writing process.

## 4.1 Goal Setting and Monitoring

WriteFlow supports users in planning and monitoring their writing in two ways. First, when users are discussing with the voice agent based on their inputted information, WriteFlow can sort out their ideas

and generate adaptive writing goals (In line with DG2 and DG3). Users are allowed to check if generated goals are specific, measurable, attainable, relevant and time-bound (SMART) on self-evaluation cards (Figure C4) and decide whether to keep or improve these goals (To meet DG1). My Goals page (Figure D) allows them to track documents and written content associated with each goal, helping them stay focused on the current goal. Suggestion cards (Figure D1) provide actionable advice to complete writing goals, inspiring users to generate subgoals and build a coherent logical structure. After finishing a writing goal, the Evaluation feature (Figure D2) evaluates whether their writing content is aligned with writing goals and their outline, highlighting potential areas for improvement. Second, the Outline view (Figure B2) enables users to create section-based outlines and maintain multiple outline versions throughout the writing process. This supports the flexible management of evolving writing plans, allowing users to revise structure as their ideas develop (Addressing DG1). By assisting users in mapping evolving text to their intended goals, WriteFlow aims to promote reflective revision and structural thinking. By grounding revisions in human-AI co-generated writing goals, WriteFlow is also designed to mitigate revision-related stress and maintain structural coherence.

### 4.2 SMART Goal Evaluation

The SMART goal evaluation feature is designed to help users determine whether their goals are actionable and how they fit within their overall goal network. A specific goal clearly outlines the action steps the writer will take. A measurable goal includes metrics to track progress, such as a word count target or the inclusion of a certain number of examples. Attainable goals are those that can realistically be achieved, taking into account the writer's resources, skills, and time constraints. A goal is relevant if it contributes to the main argument. Lastly, a time-bound goal has a clear and reasonable deadline, which helps writers stay on track and maintain focus. The purpose of this evaluation framework is not to constrain users to set only specific goals, but to help writers assess whether their goals are abstract or already actionable sub-goals, and to encourage them to transform abstract goals into concrete ones.

## 5 PRELIMINARY EVALUATION STUDY

The primary objective of this evaluation study is to understand how WriteFlow may support users in planning their writing, and how it influences their thinking and reflection throughout the writing process. Specifically, we aimed to investigate:

- **RQ1:** What kinds of writing strategies do participants employ when using WriteFlow, and how do these strategies support their academic writing process?
- **RQ2:** To what extent does WriteFlow affect students' perceived critical thinking during the writing process?
- **RQ3:** What challenges do users encounter when using WriteFlow to plan and regulate their writing?

By analyzing participants' writing activities and their reflective feedback during and after writing tasks, we seek to understand whether WriteFlow, by encouraging goal-setting, can facilitate more reflective and intentional adacemic writing.

### 5.1 Participants

Six writers with HCI expertise (5 female, 1 male) aged 23 to 26 (M = 24.83, SD = 1.17) participated in our evaluation study. Three of the participants are novice researchers currently in the second year

of their master's programs at Swedish universities, one of them had recently completed the program. One participant is a PhD student at a university in Denmark. The remaining two participants had at least three publications in the field of HCI. As academic writing and reflective reporting are integral to HCI research practice, students in this field regularly engage in reflective academic writing. Their prior experience of designing and evaluating interactive systems also enabled them to critically reflect on WriteFlow's features and design choices. This made them knowledgeable participants for evaluating our system.

Regarding AI usage frequency in academic writing, 1 used very often (heavily rely on it for most or all of their writing assignments), 2 used often (use it for most writing assignments, but not heavily), and 3 used occasionally (use it for certain parts of the writing process, depending on the task). They varied in terms of how they used AI tools during their writing process. 5 of them primarily used AI for language polishing, including sentence refining and grammar check; 2 for idea exploration and inspiration; with 2 having used AI for structural support. Each participant spent approximately 120-150 minutes total in our study, and received a gift voucher as compensation for their time.

### 5.2 Study Design

We employed a Wizard-of-Oz (WoZ) study design to investigate users' planning and reflection process when collaborating with WriteFlow in writing. The study participants interacted with WriteFlow, a high-fidelity prototype built using ProtoPie. This platform was chosen because it enables users to type, supports real-time recognition of voice input, and displays the corresponding transcript, thereby making the system highly authentic. The implemented interface supported multimodal input, including text entry, document upload, and simulated voice-based interaction with the AI assistant. Participants believed they were interacting with an autonomous AI agent, while in fact, all system responses, both text-based and voice, were controlled in real time by a human facilitator, namely the author of this thesis. Since current commercial generative AI models often struggle to produce valid responses in a single attempt, the WoZ approach offers fine-grained control over the content that ensures interaction stability and consistent user experience across sessions. In addition, the WoZ approach substantially reduced the time required for continuous prompt adjustment, enabling us to focus on the design of the goal-setting support during the academic writing process.

*5.2.1 Writing Task.* To provide a realistic scenario, participants were asked to imagine they were completing a real academic assignment: writing an interaction critique of TikTok from a user experience perspective. This assignment originates from the authentic KTH course *DM2630: User Experience Design and Evaluation (HT24)*, served as the contextual basis for all three writing tasks in the study.

Interaction criticism writing is a type of academic writing that is widely adopted in HCI and media studies [5]. It combines analytical reasoning, evidence-based argumentation, and critical reflection, encouraging students to interpret complex systems and evaluate their broader social implications. During the critique writing process, students must negotiate the relationships between purpose, audience, topic, and personal intention. This makes it an ideal task for this study.

*5.2.2 Setup.* We used GPT-4o to generate all AI responses in the study, including voice replies, AI generated writing goals, suggestion cards for each writing goal. All prompts were pre-defined and tested prior to the study to ensure consistency and task relevance.

For general voice replies, the LLM was seeded with the following prompt:

> ***System Prompt:*** *You are now a research expert in the field of '[research field]' and skilled in academic writing. Your role is to assist me by analyzing the information from the literature reviews I provide that are relevant to the writing task, listening to my needs, and guiding me based on the content I have already written. All the related information files are attached.*
> *Writing requirements:'[writing instructions]',*
> *Outline (skip if none exists):'[Outline]',*
> *Current written content (skip if none exists):'[text]',*
> *My question:'[user query]'.*
> *Respond in a conversational tone must be within 600 characters, avoiding bullet points. Your reply must be concise and clear so listeners can grasp the meaning at a glance. Please minimize the use of subordinate clauses to avoid generating excessively long sentences.*

For writing goal generation, the ending of the prompt was modified to:

> *Please create a writing goal list based on the perspective: 1. Each goal needs to meet the SMART (specific, measurable, attainable, relevant, and time-bound) criteria. 2. Please provide the estimated time required for each goal, e.g., minutes/hours. 3. Each goal should be summarized in a single sentence that is concise.*

The researcher only modified or filtered the output when GPT-4o provided repeated content or inappropriate formatting (e.g., bullet points, excessively long outputs exceeding 600 characters).

Voice output was played using ElevenLabs' *"text-to-speech"* feature[1], with playback speed adjusted to ensure clarity and a comfortable listening pace.

To help participants quickly familiarize themselves with the writing task and ease into the writing process, we pre-prepared literature review notes: *"LR Interaction Critique"*, introducing theoretical frameworks for writing interaction criticism; *"LR TikTok"*, providing real empirical case studies of TikTok (e.g. interface design, algorithmic mechanism, user behavior and its social impact). Participants were encouraged to draw upon these materials, but were also free to bring in their own insights or conduct additional online research during the study. We agree that these reading materials probably influenced participants' choice of writing content. However, this study did not focus on investigating their final written output. Instead, we aimed to observe how writers approached the writing process using WriteFlow, particularly in terms of their writing strategy use and thinking process.

To ensure that participants' interactions felt as authentic as possible, we took several steps to "hide the wizard". We invited another researcher, a Master's student at KTH to assist with noting down participants' questions and their written output on the interface during the study. This was explained to participants in the consent form. We also conducted several pilot studies to ensure that participant's question could be answered and played back within 1-2 minutes.

### 5.3 Procedure

The study involved three writing tasks followed by a post-study interview. All sessions were conducted remotely via Zoom and were both screen- and audio-recorded for accurate transcription and thematic analysis [15].

Prior to the study, participants accessed a Doodle booking link where they provided demographic information and were informed of the study's purpose and general expectations. They were also required to watch a 3-minute video to familiarize themselves with WriteFlow's functionalities and understand its design intentions. The video demonstrated the entire goal-setting workflow through a case example. It

showed how users can input their writing requirements, set goals and review the SMART evaluation, as well as track their progress.

This video presented WriteFlow in an idealized manner, emphasizing that it would generate tailored and reasonable writing goals based on the user's inputted information. While this portrayal aimed to help participants understand the system's design intentions, it may also have shaped their initial perceptions and expectations of its effectiveness in supporting the writing process. However, the complexity of the task (required to engage in a real writing assignment), and the sufficient duration of the study (over 2 hours) enabled deeper, more hands-on interaction with the system. This encouraged participants to move beyond the expectations set by the video and engage with WriteFlow in practice. Therefore, our analysis focuses on the ways in which participants interacted with the system and their actual use of WriteFlow, rather than on the initial expectations formed by the video.

*5.3.1 Introduction (~15 minutes).* Participants were welcomed and informed in detail about the study's objectives, procedures, and data handling through the consent form. After verbally confirming their consent, the facilitator responded to questions regarding the video demo to ensure that the participants understood how to use this tool. Following this, the facilitator introduced the writing assignment and asked participants to imagine they were completing a real academic writing task. The researcher clarified that the purpose of the study was to observe their writing processes rather than evaluate the quality of their final written output. In particular, we were interested in how participants interacted with WriteFlow by asking questions, setting goals, and making decisions. Participants were then briefly introduced to the two literature review notes and given time to familiarize themselves with the provided information.

*5.3.2 Writing tasks (~90 minutes).* Participants were introduced to three writing tasks and asked to use think aloud during the study. In Task 1 (Goal Generation Before Writing, ~45 minutes), participants were instructed to discuss their initial ideas with the AI, which then generated writing goals based on the conversation. After that, participants were asked to evaluate and choose to accept or reject the suggested goals. They were required to generate as many writing goals as possible (at least three writing goals) in Task 1. This was followed by Task 2 (Goal Revision, ~20 minutes), participants were provided with a pre-written paragraph and a revised outline, and asked to identify the misalignments between them. They were encouraged to engage in dialogue with the AI to generate goals that would help align their writing with their outline. In Task 3 (Goal Reflection, ~15 minutes), participants were asked to review the goals they previously generated in task 1, which were displayed on the "My Goal" page. They were guided to share their reflections on the overall goal-setting process with WriteFlow. A five-minute break was scheduled between each task to allow the researcher to prepare materials for the next phase. These three writing tasks aimed to examine how WriteFlow influenced participants' planning and reflection before and during the writing process.

*5.3.3 Interview (~25 minutes).* Upon completion of the writing tasks, an in-depth semi-structured interview was conducted. The interview was designed to understand participants' overall perception and experiences with WriteFlow. Participants were encouraged to reflect on whether and how the system influenced their critical thinking. For instance, in evaluating writing angles, comparing arguments, or judging the relevance of content. They were also guided to discuss how WriteFlow might impact their future writing, such as the way they review literature and utilise outlines. Other questions investigated participants' perceptions of AI-generated goals, and whether they supported their writing intentions.

They were asked to compare their experiences with WriteFlow and ChatGPT, discussing moments when WriteFlow felt more or less helpful in supporting academic writing.

### 5.4 Data Analysis

Six participants provided six recordings of their interactions with WriteFlow, along with their post-study interviews. The thematic analysis was conducted by the author, with the assistance of another student from KTH, in order to minimize potential bias. The assisting researcher's involvement was noted in the consent form and confirmed by all participants. All recordings were first transcribed and translated into English by two researchers with the assistance of ChatGPT. ChatGPT was limited used for initial translation purposes, and all outputs were carefully reviewed and verified by researchers to ensure consistency and human oversight of automatically-generated outputs. We applied thematic analysis to analyze the qualitative data. The transcripts were coded in three iterative rounds by two researchers, driven by the three research questions. In each round, we refined the codes by analyzing newly emerging patterns and assessing how well they aligned with the focus of each research question. Once initial themes had been identified, we collaboratively examined and compared whether they were coherent, consistent, and distinctive across different participants.

### 5.5 Results

*5.5.1 Users found WriteFlow to be encouraging for sustaining human-AI dialogue and fostering reflection-in-action in the writing process.* Some participants viewed WriteFlow as a partner that was actively engaged throughout all phases of their writing process (U2, U3, U5, U6). While some participants noted that traditional AI tools such as ChatGPT often only facilitate one-sided interactions limited to isolated stages of writing (U3, U5), they perceived WriteFlow as fostering a more continuous and integrated engagement that supported idea development and decision-making throughout the writing process. One of the participants has explicitly stated:

> "Right, it's not just something for an initial summary, or a final summary. That kind of usage is more linear, made up of isolated steps, separate processes or assistance. But when using WriteFlow, this tool can actually involve itself in the writing process, like a partner... It's kind of like acting as a focus group, where there are a few people with different roles or perspectives giving you feedback, and then it becomes integrated into a kind of qualitative research process." (U3)

When U5 noticed that the AI's early suggestions (e.g., on how to write a conclusion) were overly general, she attributed this to the lack of sufficient discussion with the agent. As she explained, "I think what it's done is still a bit general, probably because we haven't discussed it in depth yet, so it ends up giving a rather vague summary." She expressed a desire to continue discussing with the voice agent to obtain more specific and tailored feedback that could support her next steps in the writing task.

Participants mentioned that through sustained, reflective dialogue, WriteFlow supported the clarification of unclear concepts and ideas (U1, U2, U4), and in some cases, inspired new perspectives and insights that extended beyond their uploaded notes while remaining aligned with their writing intentions during the mid-writing process (U2, U4). All six participants noted that WriteFlow consistently prompted them to re-engage with their prior knowledge, identify overlooked perspectives, and re-evaluate the linkage between their arguments during the writing process. As U3 expressed that AI suggestions led him to return to his notes, re-assess their alignment with emerging ideas.

> "Because honestly my mind is pretty messy, and then it gives me some suggestions that tell me to go back and take another look... that lead me to revise something I had not thought of before. It pushes me to find more relevant content." (U3)

Participants expressed that the system's use of sufficient and concrete evidence and clear reasoning based on their materials makes its suggestions feel more trustworthy and reliable (U1, U2, U3, U4, U5, U6). It helped them determine whether a particular point or argument is worth developing further, thereby enhancing their writing confidence and making their writing become more focused. However, U3 raised concerned about whether the system would still work reliably when he uploaded incomplete or lower-quality notes.

Participants also underlined that WriteFlow helps them evaluate their written content, identify missing content by checking their outline, and suggest improvements aligned with their original goals (U1, U2, U3, U4). When U1 was revising a previously written paragraph based on her newly revised outline, she found the system's feedback particularly helpful, "Because the voice assistant say that you're missing what, but you're not missing what... It helps me to see from more broader angle, to know that what I've missed in my paragraph". Such feedback helped participants take a more holistic view of their drafts, making it easier to reflect on whether each part served their intended goals.

*5.5.2 The goal-setting feature was viewed as an effective scaffold to guide critical thinking.* All the participants pointed out that WriteFlow encourages articulating writing goals, sub-goals, and key points in advance, prompting them to write and think in a more structured way. Most participants expressed that by setting goals with WriteFlow, they were encouraged to clarify the overall structure before writing, which helped them think more systematically and stay focused on their intended direction throughout the writing process (U2, U3, U4, U5, U6).

> "The way it encourages me to list out writing goals, sub-goals, and key points really helps me grasp the blueprint of the whole essay before I start writing. With that blueprint, I feel like during the writing process, I'll be more conscious about whether what I'm writing actually aligns with my writing goals, and what role it plays in the whole essay." (U4)

Furthermore, U3 noted that using WriteFlow helped him connect earlier ideas with his current writing goals and avoid tunnel vision.

Other study participants also appreciated that WriteFlow made it easier to revise or adjust their writing plans during the writing process (U2, U5, U6). For example, U2 shared: "I can easily to modify my writing plan at any stage of the writing process. And I don't need to be scared of losing the directions". U5 expressed the flexibility to refine goals: "They (AI-generated writing goals) summarize a few things you need to do, and then you can go through and filter or refine those goals yourself." During the user test, several participants (U2, U3, U5) initially set a few broad writing goals, then continued discussing with the AI, they said when they had new thoughts, they would turn back to refine those goals or generated new sub-goals based on emerging ideas.

Goal-setting also appeared to encourage one participant to evaluate whether her writing fulfilled the intended goals, promoting her metacognitive awareness during the writing process. As U5 noted, while she might not typically initiate such evaluations on her own, WriteFlow "actively prompts you to do evaluations," making her more aware of monitoring the completion of each goal.

*5.5.3 Concerns about hindering active thinking and further exploration.* Although WriteFlow supports structured thinking, some participants cautioned that its relatively complete and single-perspective

responses could constrain divergent thinking and reduce opportunities for deeper exploration. U2, U3, U4, U6 felt that they would rely on WriteFlow to summarize literature, which might reduce their deep analysis and synthesis during the literature review. As U3 stated, "It could also cut off your further exploration, like, it might stop right there. Or, it might also encourage you to reflect, like maybe you didn't explore a certain topic enough, and then you'll go find more papers or dig further. So yeah, both things can happen."

U4 further noted that when the system delivers a single-perspective and seemingly complete response, she is inclined to copy and paste it without further elaboration, "So I was hoping it could give me several possible ways to think about it, instead of just giving me one answer that I could copy and paste." To mitigate this, U3 and U4 suggested that the system could provide multiple perspectives or alternative options to prompt further reflection and support user agency. This could encourage users to expand beyond their current materials, seek out new information more aligned with their writing intentions, and keep refining their literature review notes.

*5.5.4 Perceived impact of voice-based interaction on reflection and writing.* Several participants reflected on how the voice modality influenced their thinking process during writing. While voice interaction is designed to afford immediacy and a more natural flow of expression, both U5 and U6 felt that typing provided more space for deliberate reasoning when formulating questions. They expressed a desire for a text input option to support the formulation of complex questions. As U6 noted "when you're typing things down, there's more of a thinking process involved." Similarly, U5 was concerned that relying solely on voice input made it harder to organize her thoughts and maintain logical flow while speaking.

> "Because I feel like when I'm speaking, if the logic gets cut off in the middle, I can't pick it back up. But if I can think while typing, I can still make changes anytime — like finish writing a paragraph and then send it to it." (U5)

Some participants also noted challenges in keeping up with what the AI voice assistant said (U1, U2, U4, U6). As the AI responses grew longer, U2 found it difficult to recall earlier utterances. To mitigate this, U1, U2, U4 and U6 emphasized the importance of having both audio and textual output, enabling them to easily revisit and compare AI's suggestions with their own writing plans. U3 also suggested that the system could improve the differentiation of the voice agent's states, such as listening, replying or thinking, by using cues (e.g., a notification sound, icon or animation change). He expressed concern that unexpected voice responses might distract him during the writing process.

*5.5.5 Compared to ChatGPT, WriteFlow is perceived as enhancing authorial control.* Participants described WriteFlow as offering a greater sense of authorial control, referring to the extent to which users retain agency over the content and direction of their writing when interacting with AI.

Several participants commented that WriteFlow helped them make more informed decisions by providing explanations behind each response (U4, U5). For example, as U4 highlighted, the SMART framework [16, 19] tells you clearly why a goal is not reasonable, the evaluation card does not just give you an overall score, but clarify multiple dimensions such as specific, measurable and relevant. U5 also noted:

> "This tool allows decision-making at every decision point. When the AI provides something that's incorrect or off-track, it lets the user make a direct and convenient choice, like accept or reject... It (SMART framework) gives you a comparison and a framework for evaluation. After evaluating, then I can decide whether to add that goal to the list." (U5)

WriteFlow was perceived to discourage fully delegating content creation to the AI tool. In the user study, one participant sought to initiate content creation by asking the AI to generate an outline from scratch. The system responded by redirecting attention back to the user's uploaded notes and asking, "You already have a lot of ideas here, where would you like to start?" This interaction was seen to reinforce the user's role as the primary author and thinker in the writing process.

Participants (U2, U3, U4) also shared that WriteFlow supported ongoing evaluation of whether their writing stayed aligned with their original intentions, as shared by U2 "I think I would prefer to use this system to help me to maintain my intention in the writing". U2 and U3 perceived WriteFlow as a supportive writing partner, one that encouraged them to propose confusion, disagreement, and alternative statements. This seemed to shift them from passively receiving AI-generated content (as with ChatGPT) to actively negotiating meaning and structure, thereby reinforcing their agency and voice throughout writing (U2, U3).

*5.5.6 Design needs for structuring and visualizing writing goals.* Participants expressed specific expectations around how writing goals should be presented to better support the organization of their writing.

U1 and U2 were particularly attentive to goal sequencing during the user test. Both U1 and U2 both reflected on how the sequence of goals might affect the logical flow of their writing. U1 focused on how goal order could affect the logical flow within paragraphs. While U2 developed her outline through dialogue with the agent, she asked the agent to generate writing goals based on her outline and then requested the AI to propose a more logical sequencing to reorder them. She found the AI's suggested ordering reasonable and subsequently reorganized her own outline to match it.

Several participants expressed the need to distinguish different writing goals (U3, U4, U5, U6). They expressed a desire for writing goals to be represented beyond a simple to-do list format, with visualizations that reveal the hierarchical relationships among goals. Participants suggested that visualizing these relationships could help clarify how goals contribute to the overall structure of the academic.

> "The other one could be a kind of hierarchical structure, where it tries to group similar goals or show which ones are sub-goals under others. That would make things clearer, because writing is a big task. If you just list them all together, you might lose track of which goal belongs where. And it'll get visually overwhelming. It can have a detailed way of presenting the goals, and also a more high-level, summary-based ordering. It targets both the to-do list, and also the overall structure of the article, the layered arrangement of my writing content and suggestions." (U3)

Such support was also seen as useful for managing cognitive load, especially when navigating complex writing tasks.

In summary, participants mentioned that WriteFlow not only supported them in organizng their writing more effectively, but also encouraged reflection-in-action [54] and a greater sense of ownership over their decision-making in the writing process. Through continuous dialogue and goal-setting, some participants felt the system guided them to think and write in a more structured and organized way. Several participants also suggested providing multimodal input and output options, as well as clearer indicators of the voice agent's states, to support smoother interaction and minimize disruptions to their thinking. At the same time, they expressed a desire for providing multiple perspectives feedback and the necessity to visualize the hierarchical relationships of writing goals, to better support reflection and planning.

These insights offer concrete directions for improving WriteFlow's design in the future, ensuring it better supports goal-setting and critical reflection throughout the writing process.

## 6 DISCUSSION

Drawing on the study's findings, we revisit the three research questions to analyze the ways in which WriteFlow can support participants during their writing processes. In the following sections, we discuss what writing strategies participants used differently when using WriteFlow. (RQ1), how the interactions with WriteFlow influenced their critical thinking (RQ2), and what design opportunities emerge for guiding users in setting structured writing goal networks (RQ3).

### 6.1 RQ1: Different Writing Strategies with WriteFlow

To answer RQ1, our initial findings demonstrate that unlike existing commercial writing assistants which focus on product goals, WriteFlow appeared to emphasize the *process goals* and may enhance several writing strategies such as self-talk, self-questioning, goal setting, and outlining [56]. Participants were observed to clearly clarify their writing intentions, frequently question different arguments and effectively leverage the outline and writing goals to monitor their writing when interacting with WriteFlow.

WriteFlow was perceived by participants as encouraging reflection on their writing intentions through self-questioning and self-talk. Compared to ideation-focused systems such as AngleKindling [49] and Luminate [59], which automatically generate ideas for users, WriteFlow was experienced to inspire learners to explore writing angles grounded in their own writing intentions. This approach has potential to enhance their autonomy in the writing process. By ongoing AI-mediaeted voice-based dialogue, the participants were able to ask WriteFlow to navigate concepts and examples within their notes, based on their writing interests. The system organized and categorized related concepts and topics, helping participants stay focused and preventing them from becoming overwhelmed by extensive literature and notes. Its integrated, voice-based interaction allowed users to frequently ask questions while writing, which appeared to enhance the fluency of human-AI collaboration. However, some participants noted that formulating questions through voice-based interaction could be stressful. U5 worried that if her train of thought cut off while speaking, it would be difficult to return to the previous topic or reorganize her throughts. In the study, she still wrote down her questions first and then engaged in voice conversations with the AI. Additionally, long voice outputs were also observed to potentially frustrate writers and hinder them from continuously questioning and comparing different arguments.

Consistent with Flower and Hayes' cognitive process model of writing [25, 29], which conceptualizes writing as a recursive process involving planning, translating, and reviewing, participants generally found that WriteFlow supports goal setting and management across these interrelated processes. Unlike VISAR [70], which supports content goal management only during the planning process, WriteFlow was observed to enhance users' goal-setting strategies by encouraging them to consider process goals (*when* and *how* to present the content), track goal progress, and considering subgoals through suggestions.

The SMART [16, 19] evaluation for goal setting appeared to be beneficial for some participants in terms of goal refinement, time management and self-evaluation. Based on their expression, the SMART criteria encouraged them to articulate why certain goals were insufficient, and how to adjust them. Through goal evaluation, participants were observed to frequently compare their writing with outline, checke misalignments by treating their outlines as reference points. This process allowed them to re-evaluate their plan and move fluidly to the next planning process. This integration of goal setting, goal

monitoring, and goal progress feedback differentiates WriteFlow from prior writing assistants focus on isolate stages (as discussed in Section 2.5). Our initial insights suggest that this integration further supports learners' goal-oriented writing through iterative cycles of planning, translating, and reviewing.

### 6.2 RQ2: Opportunities and Risks in Supporting Critical Thinking

Most participants believed that WriteFlow encouraged *reflection-in-action* [54], enabling them to actively assess and adjust their writing plans throughout the process, in contrast to systems such as Friction [67] and ALure [45], which primarily foster reflection only at the final revision process. By articulating goals and refining outlines iteratively, participants expressed that they became more aware of their intentions, progress, and gaps in their writing process.

Aligned with Fleck and Fitzpatrick's framework of reflection levels [24], Participants appeared to experience a shift from descriptive and dialogic reflection (R1–R2) - in which participants revisited prior ideas, compared alternative goals, and analyzed their relationships - to transformative reflection (R3–R4), where they re-evaluated their plans and reframed their writing goals. WriteFlow's goal setting feature appeared to serve not only as a planning scaffold but also as a basis for continuous self-monitoring, aligning with the self-observation, self-evaluation, and self-reaction processes described in the Zimmerman's SRL model[71].

WriteFlow was designed to facilitate self-observation by allowing users to visually compare their outlines or goals with their written content through the interface, and through frequent voice-based dialogue. It appeared to support self-evaluation as users judged the relevance, specificity, and attainability of their goals and engaged in nonlinear, goal-oriented assessment of their progress. Finally, it was observed to encourage self-reaction, as participants continuously adapted or reformulated their goals in response to their own evaluations.

Based on participants' report, WriteFlow consolidates writers' uploaded notes, task requirements, and outlines, thereby providing context-aware suggestions that are well tailored to their goals and the study context. However, this also introduced a risk of over-reliance. As mentioned by some participants, they may become depended on the system to summarize information, it would reduce opportunities for deeper reading, comprehension, and autonomous reasoning. This highlights the need to carefully constrain the system's responses, avoiding presenting overly complete answers that may discourage active thinking.

### 6.3 RQ3: Design Opportunities for Building Goal Networks

Regarding RQ3, we identified several design opportunities to better guide and encourage users in building a solid goal network (see Section 2.1). One design implication is the need to present hierarchical visualizations of writing goals. Showing the structural relationships between goals and sub-goals offers a clearer overview of the writing plans, both short- and long-term. This could help writers understand how each writing goal connects to broader writing intentions and the overall objective, thereby reducing cognitive overload for planning and goal tracking. Additionally, system prompts should be refined to present multiple perspectives or guide users to explore alternative angles, helping them evaluate and refine ideas rather than follow a single suggestion. Lastly, AI-driven voice interaction should be supported by clear visual feedback that distinguishes system states, such as speaking, generating, or listening. This will help writers stay aware of the system's status, reducing unnecessary attention shifts and preventing interrupting users' reflective thinking.

### 6.4 Wizard of Oz approach

In this study, we employed the Wizard of Oz method to simulate the interaction with a fully autonomous AI system. This approach allowed us, at this stage, to effectively gather users' authentic experiences and behaviors during their interaction with the system. This was crucial to gain an initial understanding of how WriteFlow may affect writers using writing strategies and reflection processes. More importantly, it enabled us to gain a deeper understanding of why WriteFlow may foster or hinder writers' critical reflection in writing. The used method provided the opportunity to tailor the AI's responses in real time, ensuring that users could experience the system's capabilities without the current limitations of fully autonomous AI models. As part of our formative study, it provided valuable insights, helping us refine the system to enhance the fluency of human-AI collaboration. However, the Wizard of Oz approach limits scalability. It requires researchers to manually control each participant's interaction. Meanwhile, this approach is also difficult to apply in studies involving long-term, real-world writing scenarios. Therefore, we emphasize the need to develop a fully functional application to evaluate the system's generalizability across various writing contexts.

### 6.5 Ethical Considerations: Autonomy and Over-Reliance on AI in Writing

Ethical considerations concerning writing autonomy and the potential for over-reliance on AI have been outlined in sections 2.4 and section 2.5. Whilst existing systems aim to assist users by providing personalized suggestions, concerns arise that they may diminish users' autonomy within the writing process. As these tools directly generate ideas for users, long-term reliance upon them may impair writers' independent critical thinking and expression abilities. According to Kosmyna et al. [37], students' early dependence on LLM tools appeared to have impaired long-term semantic retention and contextual memory. To improve writers' authorial control in writing, it's crucial to ensure transparency in the system's decision-making process and carefully design such system to reduce AI's over assistance.

## 7 LIMITATIONS AND FUTURE RESEARCH

This study is not without limitations to be considered.

*Not fully replicate an authentic writing experience.* While our study aimed to simulate realistic academic writing scenarios, several factors limited the authenticity of the writing context. In real-world academic contexts, writing is often a self-initiated process instrumented by the instructor. Students are typically introduced to task-related knowledge through lectures or assigned readings, and, following task instructions, they conduct their own literature reviews to develop ideas. However, due to time constraints, participants in our study were provided researcher-prepared literature notes. They were required to skim through and use them to plan their writing. This may have made them feel less connected to the material, which could have reduced how deeply they engaged in writing and how smoothly they interacted with the system. Furthermore, the extensive background materials may introduce additional cognitive pressure. Writing under researcher observation may also have hindered immersion in the writing task. Future evaluations would benefit from allowing participants to bring their own academic writing assignments and literature notes into the study. However, despite these limitations, we observed that most participants carefully went through the literature notes and collaborated with the voice agent to explore possible structures. By thoughtfully engaging with the provided materials and interacting with WriteFlow, participants provided rich qualitative data, which offered valuable insights into our research questions.

*Limited control over prompt structure and response length.* As noted in Section 5.2, we intentionally fixed the prompt structure and controlled the length of voice response (600 words). However, several participants expressed a desire for more diverse and multi-perspective responses. Participants' feedback suggests that the prompt formulation had a strong impact on their user experience and cognitive engagement. For example, prompts that raised questions or challenged existing arguments were seen as effective in encouraging their reflection. We also observed that longer voice responses sometimes made it difficult for participants to retain information. Future work should investigate different prompt engineering strategies and response lengths to determine how WriteFlow can more effectively facilitate students' reflective thinking and writing planning processes.

*Lack of testing with diverse participants.* Our evaluation study involved six writers with HCI expertise who had prior experience in using AI for their writing. To evaluate the system's usability and generalizability across different writing contexts, recruiting a more diverse participant pool across disciplines and writing expertise is needed. In addition, as the participants' primary language was Chinese, their language background and educational background may affect their experience with the voice agent, with some reporting difficulty recalling the responses. Testing with users from different linguistic backgrounds and cultures is needed to more effectively evaluate the voice interaction and the effectiveness of the AI-mediated writing support.

## 8 CONCLUSION

We designed *WriteFlow*, an LLM-enhanced writing assistant that supports iterative goal-network construction and progress monitoring for academic writing. With *WriteFlow*, students can discuss their writing plans with an AI-mediated voice-based reflective agent, define writing goals, and track their progress. These features were informed by insights from a formative study with 17 writers. In a Wizard-of-Oz evaluation with six novice HCI researchers, participants reported that the system encouraged more deliberate reflection and supported aspects of critical thinking during the writing process. At the same time, several participants noted challenges in interacting with the voice agent in certain situations. We see this work as offering a promising direction for designing LLM-based tools that support goal-oriented writing processes and may help foster critical thinking in the academic writing process. We outline potential design improvements and directions for future research, including opportunities to further leverage reflective dialogue and goal-setting as scaffolds for developing critical thinking skills.

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Professor Olga Viberg, for her patience and insightful guidance throughout this research project. I am also grateful to my examiner, Professor Anders Hedman, for his insightful feedback. Special thanks to Runyu Yang, a Master's student at KTH, for his assistance with the Wizard-of-Oz study. Finally, I would like to extend my appreciation to all the student participants who took part in the study. Their time and valuable insights made this research possible.

## REFERENCES

[1] [n. d.]. ElevenLabs — Text to Speech. https://elevenlabs.io/app/speech-synthesis/text-to-speech. Accessed: 2025-10-30.
[2] [n. d.]. NotebookLM. https://notebooklm.google.com/. Accessed: 2025-04-26.
[3] Khan Academy. 2024. Khanmigo Writing Coach. https://www.youtube.com/watch?v=u2l0ujUiUv8 Accessed: 2025-04-26.

[4] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A Hearst, Andrew Head, and Kyle Lo. 2023. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–38.

[5] Jeffrey Bardzell. 2011. Interaction criticism: An introduction to the practice. *Interacting with computers* 23, 6 (2011), 604–621.

[6] Caroline Beauvais, Thierry Olive, and Jean-Michel Passerault. 2011. Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology* 103, 2 (2011), 415.

[7] Jonas Becker, Jan Philip Wahle, Bela Gipp, and Terry Ruas. 2024. Text generation: A systematic literature review of tasks, evaluation, and challenges. *arXiv preprint arXiv:2405.15604* (2024).

[8] Marit Bentvelzen, Paweł W Woźniak, Pia SF Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting reflection in hci: Four design resources for technologies that support reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.

[9] Carl Bereiter and Marlene Scardamalia. 1987. An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. *Curriculum inquiry* 17, 1 (1987), 9–30.

[10] Daniel Buschek. 2024. Collage is the New Writing: Exploring the Fragmentation of Text and User Interfaces in AI Tools. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 2719–2737.

[11] Cecilia Ka Yuk Chan and Wenjie Hu. 2023. Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 43.

[12] Yulin Chen, Ning Ding, Hai-Tao Zheng, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2024. Empowering private tutoring by chaining large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 354–364.

[13] Kiroong Choe, Seokhyeon Park, Seokweon Jung, Hyeok Kim, Ji Won Yang, Hwajung Hong, and Jinwook Seo. 2024. Supporting novice researchers to write literature review using language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

[14] Huy Q Chung, Vicky Chen, and Carol Booth Olson. 2021. The impact of self-assessment, planning and goal setting, and reflection before and after revision on student self-efficacy and writing performance. *Reading and Writing* 34, 7 (2021), 1885–1913.

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

[16] Anne Conzemius and Jan O'Neill. 2009. *The power of SMART goals: Using goals to improve student learning*. Solution Tree Press.

[17] Scott A Crossley, Perpetual Baffour, Yu Tian, Aigner Picou, Meg Benner, and Ulrich Boser. 2022. The persuasive essays for rating, selecting, and understanding argumentative and discourse elements (PERSUADE) corpus 1.0. *Assessing Writing* 54 (2022), 100667.

[18] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond text generation: Supporting writers with continuous automatic text summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[19] George T Doran. 1981. There's a SMART way to write managements's goals and objectives. *Management review* 70, 11 (1981).

[20] Tom Duenas and Diana Ruiz. 2024. The risks of human overreliance on large language models for critical thinking. *Research Gate, 2024e. URL http://dx. doi. org/10.13140/RG* 2, 26002.06082 (2024).

[21] Robert Ennis. 2011. Critical thinking: Reflection and perspective Part II. *Inquiry: Critical thinking across the Disciplines* 26, 2 (2011), 5–19.

[22] Yizhou Fan, Luzhen Tang, Huixiao Le, Kejie Shen, Shufang Tan, Yueying Zhao, Yuan Shen, Xinyu Li, and Dragan Gašević. 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology* 56, 2 (2025), 489–530.

[23] Ralph P Ferretti, Charles A MacArthur, and Nancy S Dowdy. 2000. The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology* 92, 4 (2000), 694.

[24] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd conference of the computer-human interaction special interest group of australia on computer-human interaction*. 216–223.

[25] Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College Composition & Communication* 32, 4 (1981), 365–387.

[26] Michael Gerlich. 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* 15, 1 (2025), 6.

[27] Emma Harvey, Allison Koenecke, and Rene F Kizilcec. 2025. " Don't Forget the Teachers": Towards an Educator-Centered Understanding of Harms from Large Language Models in Education. *arXiv preprint arXiv:2502.14592* (2025).

[28] Ahmed M Hasanein and Abu Elnasr E Sobaih. 2023. Drivers and consequences of ChatGPT use in higher education: Key stakeholder perspectives. *European journal of investigation in health, psychology and education* 13, 11 (2023), 2599–2614.

[29] John R Hayes and Linda S Flower. 1986. Writing research and the writer. *American psychologist* 41, 10 (1986), 1106.

[30] Md Naimul Hoque, Tasfia Mashiat, Bhavya Ghai, Cecilia D Shelton, Fanny Chevalier, Kari Kraus, and Niklas Elmqvist. 2024. The HaLLMark Effect: Supporting Provenance and Transparent Use of Large Language Models in Writing with Interactive Visualization. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[31] Thorben Jansen, Jennifer Meyer, Johanna Fleckenstein, Andrea Horbach, Stefan Keller, and Jens Möller. 2024. Individualizing goal-setting interventions using automated writing evaluation to support secondary school students' text revisions. *Learning and Instruction* 89 (2024), 101847.

[32] Jakob Karolus, Sebastian S Feger, Albrecht Schmidt, and Paweł W Woźniak. 2023. Your text is hard to read: facilitating readability awareness to support writing proficiency in text production. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 147–160.

[33] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.

[34] Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. 2024. DiaryMate: Understanding User Perceptions and Experience in Human-AI Collaboration for Personal Journaling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[35] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for workplace reflection: a chat and voice-based conversational agent. In *Proceedings of the 2018 designing interactive systems conference*. 881–894.

[36] Siu-Cheung Kong and Yin Yang. 2024. A human-centred learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K–12 settings. *IEEE Transactions on Learning Technologies* (2024).

[37] Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. 2025. Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872* (2025).

[38] Samuli Laato, Benedikt Morschheuser, Juho Hamari, and Jari Björne. 2023. AI-assisted learning with ChatGPT and large language models: Implications for higher education. In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 226–230.

[39] K Blaine Lawlor. 2012. Smart goals: How the application of smart goals can contribute to achievement of student learning outcomes. In *Developments in business simulation and experiential learning: Proceedings of the annual ABSEL conference*, Vol. 39.

[40] Hao-Ping Hank Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. (2025).

[41] Tongguang Li, Debarshi Nath, Yixin Cheng, Yizhou Fan, Xinyu Li, Mladen Raković, Hassan Khosravi, Zachari Swiecki, Yi-Shan Tsai, and Dragan Gašević. 2025. Turning real-time analytics into adaptive scaffolds for self-regulated learning using generative artificial intelligence. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*. 667–679.

[42] Susan Lin, Jeremy Warner, JD Zamfirescu-Pereira, Matthew G Lee, Sauhard Jain, Shanqing Cai, Piyawat Lertvittayakumjorn, Michael Xuelin Huang, Shumin Zhai, Björn Hartmann, et al. 2024. Rambler: Supporting Writing With Speech via LLM-Assisted Gist Manipulation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[43] Chung Kwan Lo, Khe Foon Hew, and Morris Siu-yung Jong. 2024. The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education* (2024), 105100.

[44] Pratyusha Maiti and Ashok Goel. 2025. Can an AI Partner Empower Learners to Ask Critical Questions?. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 314–324.

[45] SEYED PARSA NESHAEI, ANTONIA TOLZIN, YVONNE BERKLE, MIRIAM LEUCHTER, JAN MARCO LEIMEISTER, ANDREAS JANSON, and THIEMO WAMBSGANSS. 2025. Leveraging Learner Errors in Digital Argumentation Learning: How ALure Helps Students Learn from their Mistakes and Write Better Arguments. (2025).

[46] Pablo Ortiz and D Fox Harrell. 2018. Enabling critical self-reflection through roleplay with chimeria: Grayscale. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 353–364.

[47] Ernesto Panadero. 2017. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology* 8 (2017), 422.

[48] Scott G Paris, Marjorie Y Lipson, and Karen K Wixson. 1983. Becoming a strategic reader. *Contemporary educational psychology* 8, 3 (1983), 293–316.

[49] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. Anglekindling: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–16.

[50] Hua Xuan Qin, Guangzhi Zhu, Mingming Fan, and Pan Hui. 2025. Toward Personalizable AI Node Graph Creative Writing Support: Insights on Preferences for Generative AI Features and Information Presentation Across Story Writing Processes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–30.

[51] Ahmed Y Radwan, Khaled M Alasmari, Omar A Abdulbagi, and Emad A Alghamdi. 2024. SARD: A Human-AI Collaborative Story Generation. In *International Conference on Human-Computer Interaction*. Springer, 94–105.

[52] Robert Reid, Torri Ortiz Lienemann, and Jessica L Hagaman. 2013. *Strategy instruction for students with learning disabilities*. Guilford Publications.

[53] Xuezhu Ren, Yan Tong, Peng Peng, and Tengfei Wang. 2020. Critical thinking predicts academic performance beyond general cognitive ability: Evidence from adults and children. *Intelligence* 82 (2020), 101487.

[54] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[55] Dale H Schunk. 1990. Goal setting and self-efficacy during self-regulated learning. *Educational psychologist* 25, 1 (1990), 71–86.

[56] Dale H Schunk and Carl W Swartz. 1993. Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary educational psychology* 18, 3 (1993), 337–354.

[57] Dale H Schunk and Barry Zimmerman. 2011. *Handbook of self-regulation of learning and performance*. Taylor & Francis.

[58] Mark D Shermis. 2014. State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing* 20 (2014), 53–76.

[59] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured generation and exploration of design space with large language models for human-ai co-creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26.

[60] Päivi Tynjälä. 2001. Writing, learning and the development of expertise in higher education. In *Writing as a learning tool: Integrating theory and practice*. Springer, 37–56.

[61] Ling Wei. 2023. Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in psychology* 14 (2023), 1261955.

[62] Joshua Wilson, Andrew Potter, Tania Cruz Cordero, and Matthew C Myers. 2023. Integrating goal-setting and automated feedback to improve writing outcomes: a pilot study. *Innovation in Language Learning and Teaching* 17, 3 (2023), 518–534.

[63] Jacqueline Wong and Olga Viberg. 2024. Supporting Self-Regulated Learning with Generative AI: A Case of Two Empirical Studies.. In *LAK Workshops*. 223–229.

[64] Fangzhi Xu, Qika Lin, Jiawei Han, Tianzhe Zhao, Jun Liu, and Erik Cambria. 2025. Are large language models really good logical reasoners? a comprehensive evaluation and beyond. *IEEE Transactions on Knowledge and Data Engineering* (2025).

[65] Xuyu Yang, Wengxi Li, Matthew G Lee, Zhuoyang Li, JD Zamfirescu-Pereira, and Can Liu. 2025. Rambler in the Wild: A Diary Study of LLM-Assisted Writing With Speech. *arXiv preprint arXiv:2502.05612* (2025).

[66] Yu-Chun Grace Yen, Isabelle Yan Pan, Grace Lin, Mingyi Li, Hyoungwook Jin, Mengyi Chen, Haijun Xia, Steven P Dow, et al. 2024. When to Give Feedback: Exploring Tradeoffs in the Timing of Design Feedback. (2024).

[67] Chao Zhang, Kexin Ju, Peter Bidoshi, Yu-Chun Grace Yen, and Jeffrey M. Rzeszotarski. 2025. Friction: Deciphering Writing Feedback into Writing Revisions through LLM-Assisted Reflection. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM.

[68] Rui Zhang, Ziyao Zhang, Fengliang Zhu, Jiajie Zhou, and Anyi Rao. 2024. Mindalogue: LLM-Powered Nonlinear Interaction for Effective Learning and Task Exploration. *arXiv preprint arXiv:2410.10570* (2024).

[69] Yuhao Zhang, Eng Lieh Ouh, Adam Ho, Siaw Ling Lo, Kar Way Tan, and Feng Lin. 2025. PromptTutor: Effects of an LLM-Based Chatbot on Learning Outcomes and Motivation in Flipped Classrooms. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*. 445–451.

[70] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–30.

[71] Barry J Zimmerman. 2002. Becoming a self-regulated learner: An overview. *Theory into practice* 41, 2 (2002), 64–70.

TRITA-EECS-EX-2025:962
Stockholm, Sweden 2025