



Understanding the Role of Large Language Models in Personalizing and Scaffolding Strategies to Combat Academic Procrastination

Ananya Bhattacharjee
Computer Science
University of Toronto
Toronto, Ontario, Canada
ananya@cs.toronto.edu

Yuchen Zeng
Computer Science
University of Toronto
Toronto, Ontario, Canada
yuchen.zeng@mail.utoronto.ca

Sarah Yi Xu
Computer Science
University of Toronto
Toronto, Ontario, Canada
sarahi.xu@mail.utoronto.ca

Dana Kulzhabayeva
Psychology
University of Toronto
Toronto, Ontario, Canada
dana.kulzhabayeva@mail.utoronto.ca

Minyi Ma
Computer Science
University of Toronto
Toronto, Ontario, Canada
minyi.ma@mail.utoronto.ca

Rachel Kornfield
Preventive Medicine
Northwestern University
Chicago, Illinois, USA
rachel.kornfield@northwestern.edu

Syed Ishtiaque Ahmed
Computer Science
University of Toronto
Toronto, Ontario, Canada
ishtiaque@cs.toronto.edu

Alex Mariakakis
Computer Science
University of Toronto
Toronto, Ontario, Canada
mariakakis@cs.toronto.edu

Mary P. Czerwinski
Microsoft Research
Redmond, Washington, USA
marycz@microsoft.com

Anastasia Kuzminykh
Faculty of Information
University of Toronto
Toronto, Ontario, Canada
anastasia.kuzminykh@utoronto.ca

Michael Liut
Mathematical and Computational
Sciences
University of Toronto Mississauga
Mississauga, Ontario, Canada
michael.liut@utoronto.ca

Joseph Jay Williams
Computer Science
University of Toronto
Toronto, Ontario, Canada
williams@cs.toronto.edu

ABSTRACT

Traditional interventions for academic procrastination often fail to capture the nuanced, individual-specific factors that underlie them. Large language models (LLMs) hold immense potential for addressing this gap by permitting open-ended inputs, including the ability to customize interventions to individuals' unique needs. However, user expectations and potential limitations of LLMs in this context remain underexplored. To address this, we conducted interviews and focus group discussions with 15 university students and 6 experts, during which a technology probe for generating personalized advice for managing procrastination was presented. Our results highlight the necessity for LLMs to provide structured, deadline-oriented steps and enhanced user support mechanisms. Additionally, our results surface the need for an adaptive approach to questioning based on factors like busyness. These findings offer

crucial design implications for the development of LLM-based tools for managing procrastination while cautioning the use of LLMs for therapeutic guidance.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Education**.

KEYWORDS

Procrastination, Personalized Reflections, GPT-4, ChatGPT, Large Language Models, Education

ACM Reference Format:

Ananya Bhattacharjee, Yuchen Zeng, Sarah Yi Xu, Dana Kulzhabayeva, Minyi Ma, Rachel Kornfield, Syed Ishtiaque Ahmed, Alex Mariakakis, Mary P. Czerwinski, Anastasia Kuzminykh, Michael Liut, and Joseph Jay Williams. 2024. Understanding the Role of Large Language Models in Personalizing and Scaffolding Strategies to Combat Academic Procrastination. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3613904.3642081>

1 INTRODUCTION

Procrastination, defined as the intentional delay of tasks despite known negative consequences [3, 28, 92], is especially widespread

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0330-0/24/05

<https://doi.org/10.1145/3613904.3642081>

in academic settings. Prior research indicates that at least half of all college students engage in chronic procrastination [37, 101], which can negatively impact their academic performance and overall well-being. Researchers have developed a range of interventions for procrastination management, focusing on task completion [3, 49], time management [49, 59], and personalized motivation [94, 107]. However, they frequently fail to capture the nuanced factors that contribute to procrastination, such as individual learning styles and environmental influences [12, 64]. The challenge lies in these factors being subjective and difficult to quantify, rendering traditional interventions less effective.

Large language models (LLMs) have recently garnered significant attention for their capability to act as personalized intervention tools and supportive scaffolds in numerous settings [1, 3, 66, 72, 109, 114]. With their computational prowess and linguistic sophistication, these models can analyze a vast array of variables via text-based input, potentially offering a level of personalization and support that is difficult to achieve with traditional intervention methods. LLMs also stand out due to their dynamic interaction capabilities, surpassing the constraints of generic surveys and rule-based chatbots that lack conversational depth and follow-up capabilities. In the context of academic procrastination, the potential of LLMs lies not in replacing traditional methods like setting reminders or automatic scheduling, but in their ability to contextualize and intelligently integrate a broad spectrum of procrastination management strategies.

By leveraging extensive datasets and natural language processing techniques, LLMs may possess the ability to craft advice and recommendations that are highly tailored to an individual's unique academic and personal circumstances [71, 79]. For instance, an LLM could ask a student about their upcoming deadlines and work schedule to collaboratively help them generate micro-deadlines that fit into their schedule. Knowing that the student has two term papers and a part-time job, the LLM might advise dedicating the first two hours after work on a specific day to outlining one of the term papers. This sort of targeted advice has the potential to acknowledge students' unique circumstances and constraints, thereby increasing the likelihood of the advice being both practical and effective.

However, LLMs cannot be trivially adapted for such intervention personalization. They currently lack the innate capacity to perceive users' emotional states or to navigate the complexities and subjectivity of intricate social contexts (although this may change with future advancements) [120]. Therefore, users need to provide such nuanced information for LLMs to offer effective personalized recommendations, and the applicability of the recommendations they receive depends substantially on the depth of the information provided [72]. It is also important for researchers to understand which intervention elements users wish to personalize. For instance, in the context of academic procrastination, users may desire an option to specify the type of support they need, such as reminders for deadlines versus tips for time management. Insights into these user expectations can guide the design of prompts to elicit targeted and useful responses through prompt engineering, aligning closely with what users seek in managing procrastination [4, 115]. They may even help adapt the attention mechanisms within LLMs [112] to focus on certain keywords or contextual clues within a user's

query or text input, such as deadlines, emotional states, or specific tasks. Without such insights, there is a risk that developers create LLM-based tools that are misaligned with user needs and expectations [9, 53]. Finally, an understanding of the limitations and guardrails required for LLM-based tools in this space is also important to ensure that intervention adjustments are still consistent with validated strategies or clinical guidelines.

Motivated by both the challenges and opportunities in applying LLMs to intervention personalization, as well as their potential to act as a supportive scaffold, we explore how students can use an LLM-based tool to manage academic procrastination. Specifically, our focus extends to task initiation, deadline management, emotional regulation, and addressing the impact of social and environmental factors on academic performance. We aim to understand how users envision the ways in which LLMs might facilitate these strategies for managing academic procrastination. At the same time, we attempt to identify potential limitations of LLMs and design tensions that should be considered for future tools in this space. Our work is motivated by the following research questions:

- RQ1:** How do users envision the role of LLMs in tailoring strategies for managing academic procrastination?
- RQ2:** What challenges or tensions might arise when using LLMs for tailoring strategies to manage academic procrastination?

In this study, we interacted with 15 university students through individual interviews and focus groups to delve into how they perceive the potential of LLM-based tools in mitigating academic procrastination. Before posing questions, we offered participants an opportunity to engage with a functional technology probe [14, 47], specifically designed to incorporate strategies for managing procrastination. This probe served as a contextual anchor for their subsequent recommendations for the design of LLM-based tools for procrastination management. To bolster our results, we also consulted six experts from the fields of clinical psychology, education, and cognitive science to validate our findings and explore potential design challenges.

Our study illuminates key preferences and concerns regarding the use of LLM-based tools in managing academic procrastination. Participants highlighted a preference for structured, deadline-focused planning coupled with real-world examples for guidance. In contrast, experts advised caution regarding the scope of these tools, emphasizing the need for emotional validation and critical thinking rather than therapeutic intervention. These findings enhance our understanding of human-LLM collaboration, advocating for a balance between methodical task management and adherence to ethical standards in emotional support. The design recommendations derived from our study call for flexible guidance systems and seamless integration into daily routines while setting clear boundaries for the emotional aspects of procrastination management. Overall, our work not only identifies specific user requirements for managing academic procrastination but also contributes to the discourse on the ethical and efficient deployment of LLM-based tools in diverse contexts.

2 RELATED WORK

We first delve into existing literature concerning the application of LLMs in academic environments, focusing specifically on their use in personalization and contextual adaptation, their role as a supportive scaffold, and the ethical challenges that arise from their use. Subsequently, we turn our attention to prior interventions aimed at mitigating academic procrastination.

2.1 Application of LLMs in Academic Settings

The emergence of LLMs has introduced new possibilities in learning and education. Within academic settings, they have the potential to serve multiple roles — from personalizing learning experiences to providing task-oriented support [1, 3, 79, 106, 122]. As a result, researchers have started exploring the role of LLM-based tools in different phases of education, including content creation and summarization, student evaluation, collaborative data analysis, and research [1, 25, 40, 65, 66, 79, 116].

One of the most salient strengths of LLMs in educational settings is their capacity for personalization and contextual adaptation [25, 61, 66, 114]. It has sparked several potential application areas in the academic arena, although much of the current research is exploratory in nature and yet to be validated on a large scale [54]. Educational platforms equipped with LLMs have been used to analyze the historical data of a student's essay submissions, allowing them to suggest areas of improvement [109] and to identify recurring grammatical errors or stylistic issues [25, 79, 80, 119]. In programming courses, LLMs might act as adaptable tutors that provide hints or suggest debugging strategies when students encounter errors in their code [30, 55, 68, 70]. Unlike static FAQs or forums, the generated advice can take into account the specific syntax and logic of the student's work. Furthermore, LLMs have the potential to facilitate personalized learning journeys by recommending courses or materials that correspond with an individual's academic objectives and interests [13, 54, 81]. To illustrate, if a student indicates a desire to achieve proficiency in a specific programming language, an LLM could suggest a series of progressively challenging exercises and relevant readings, thereby offering a bespoke pathway toward mastery [79].

Beyond their capacity for personalization, LLMs also have the potential to excel as supportive scaffolds in academic tasks that require additional guidance or collaborative input [3, 29, 48, 57, 67, 122, 126], although the research is still premature. In writing-intensive courses, LLMs might function as intelligent writing assistants [33, 58, 67, 117]; they can provide more than mere copy-editing, but also real-time support in enhancing the structure and flow of arguments. They may also assist students in brainstorming ideas, organizing their thoughts, and co-creating outlines for essays or research papers [1, 54, 60, 103, 124]. Additionally, LLMs can be helpful for microlearning experiences, particularly in subjects that involve complex scientific concepts [71, 79]. By breaking down complex ideas into more digestible and comprehensible explanations, they might offer students a way to incrementally strengthen their understanding.

While the promises of LLMs in academic settings are manifold, these benefits come with ethical concerns that cannot be ignored

[7, 50, 52, 54, 74, 105]. One significant worry is the potentially detrimental impact LLMs can have on critical thinking and problem-solving skills [50, 54]. The ease and immediacy of AI-generated responses may discourage students from delving deeply into subjects, opting instead for quick answers without a solid understanding of underlying principles [50, 78, 79]. This issue is exacerbated by the possibility that students may rely too heavily on LLMs, diminishing traditional student-teacher interactions that are often essential for learning [78]. There is also the problem of accuracy [24, 85, 87]. Students might not question the validity of the information provided by LLMs under the false assumption that any machine-generated information must be correct. This could lead to the unintentional dissemination of false or misleading information [63, 104]. Such concerns could be heightened in the specific context of academic procrastination management, where erroneous or contextually inappropriate advice could have detrimental effects [52, 100]. These ethical issues bring into focus the need for careful design and implementation strategies when incorporating LLMs into academic environments.

2.2 Interventions to Manage Procrastination

Efforts to alleviate procrastination typically aim to close the distance between intending to act and actually taking action [35, 99]. These efforts are multifaceted, incorporating tactics that improve self-regulation, bolster self-efficacy, and nurture social support networks [97, 111]. Various modalities have been employed to deliver these strategies, such as self-reflection applications, interactive conversational systems, technology-based interruptions, life management software, structured workshops, and therapeutic sessions [3, 64, 92, 111, 123].

A large body of work on procrastination management has focused on enhancing self-regulatory skills [91], through nurturing positive habits like goal setting and time management [22, 59, 101]. Various tools have been designed to help individuals spend more time on focused tasks, thereby aiding them in meeting their goals and deadlines [16, 26, 36, 44, 64, 89]. For instance, Pereira and Díaz [89] examined the effectiveness of GanttBot, a Telegram chatbot designed to reduce procrastination in single-student capstone projects. Incorporating various features like alerts, advice, automatic rescheduling, motivational messages, and historical project references, the chatbot was successful in reducing the number of overdue days among students compared to a control group. A chatbot developed by Grover et al. [36] that allowed information workers to block out time in calendars and reflect on daily mood also saw similar success in helping them spend more time on focused tasks.

Another prominent line of work has investigated the use of reflection and reminders as mechanisms to promote timely action and adherence to deadlines [5, 31, 46, 76, 121, 123]. For instance, Kim et al. [59] introduced a system that encouraged self-reflection on periods of distraction, noting a moderate decrease in procrastinatory behavior among users. Similarly, Edwards et al. [31] used email reminders to inform students about their prior homework performance, which led to a reduction in late submissions. Further evidence from Ye et al. [121] and Humphrey Jr et al. [46] suggests that both email and text message reminders can offer moderate

benefits in promoting timely submissions and improving academic performance.

Other strategies for combating procrastination included direct technological interventions and the leveraging of social support [3, 22, 45, 59, 108]. For example, Inie and Lungu [49] introduced a browser extension that prompts users with microlearning activities when they attempt to visit time-wasting websites, thereby redirecting their focus. Similarly, Arakawa et al. [3] employed generative AI models to help individuals pick up where they left off on interrupted tasks by providing contextual cues, thus aiding in the transition back to the activity at hand. On the social front, studies have underscored the importance of social support in curbing procrastination [45, 108, 110]. Tuasikal and Patria [108] demonstrated that a higher level of social support could lead to less procrastination in thesis writing, while Hsueh et al. [45] found that being part of a larger group helped reduce distractions and improve goal-setting.

However, a recurring critique of all of these interventions from procrastination management is their frequent inability to address the complex, individualized factors that contribute to procrastination [12, 64]. Factors such as social circumstances, motivational triggers, or unique lifestyles are often challenging to incorporate into these designs. Addressing such nuanced issues could potentially benefit from the open-ended input capabilities of LLMs. However, there is a notable gap in understanding how users might perceive and interact with these LLM systems, how LLMs could enhance existing interventions, and what challenges and ethical considerations may be associated with their use. Our study aims to delve into these issues.

3 DESIGN OF THE PROBE

In order to probe student participants' reactions to how LLMs can be used to help them address procrastination, we developed a web-based technology probe called the Self-generated Personalized Articulations and Reflections Kit (SPARK). SPARK utilizes OpenAI's GPT-4 model [84] to offer context-specific advice for managing procrastination. The tool was iteratively designed by our research team composed of faculty members and graduate students in human-computer interaction, natural language processing, cognitive science, and psychology.

While the design of SPARK was informed by existing literature (as described later in this section), our primary goal was not to propose and validate its design. Instead, our primary goal was to understand the features and affordances users would find valuable in such a tool. We believed that asking participants to generate design ideas in a vacuum could be challenging [12, 69, 104]; therefore, we provided them with SPARK as a probe to elicit more informed responses. SPARK comprises four key components, each of which we elaborate upon below.

3.1 Seed Message

The tool initially presented participants with a message grounded in the existing literature on procrastination management [6, 91, 92, 96, 110]. This message served as a generic blueprint for addressing procrastination and involved a subset of four strategies from prior work [92, 96]:

- **Cognitive Insight:** This component educates users about the complex psychological and cognitive factors contributing to procrastination, such as fear of failure and temporal discounting bias
- **Psychological Flexibility:** This component encourages the adoption of a flexible mental stance, promoting acceptance and mindfulness. It suggests that individuals adopt the role of an observer of their own emotional and cognitive states.
- **Value Alignment:** This component directs users to contemplate how completing the task correlates with their core values and highlights the potential costs of not adhering to these values.
- **Implementation Intention:** This component advocates for the establishment of actionable plans to align behavior with one's values when certain thoughts or emotional triggers arise.

Figure 1 illustrates how seed messages were presented, and Table 1 provides some examples of the seed messages with associated psychological principles.

3.2 Interface Options to Generate Context Specific Messages by LLMs

To gather nuanced insights into the challenges individuals face with procrastination, we provided users with an open-ended text box where they could freely share details they considered pertinent. Users also had the option to specify the tone (formal vs. informal), directedness (direct vs. indirect), and length of the LLM-generated message (50, 100, or 150 words). They could choose whether or not to include specific instructions for addressing their situation, as well as select specific prompts for generating the message (more about this in Section 3.4). These options were informed by prior literature on personalized advice [6, 34, 75]. Figure 2(a) illustrates how these options were presented.

The structure for generating each tailored response was dictated by a specific query format: "Adapt and rewrite [Seed Message] for [User Name], who is facing the following challenge: [User Situation]. The rewritten message should be [Selected Tone], [Selected Directedness], and have fewer than [Selected Length] words. [Include instructions on applying the message in the situation]." In this query, the final sentence would be excluded if the user had no preference for receiving specific application instructions for their particular situation. This query underwent multiple iterations to achieve its final form.

Participants also had the option to make open-ended requests to change the generated message as well as edit the message themselves. Figures 2(b) and 2(c) show how these features were presented, respectively.

3.3 Prompting Future Action

To foster proactive behavior and facilitate the translation of intentions into tangible actions, we implemented a feature that requested participants to draft an email to their future selves. Drawing on existing literature in the areas of task management and goal-setting [2, 36, 99], this feature aimed to encourage participants to prioritize their responsibilities and establish clear objectives. Our aim was to facilitate reflection at various temporal stages of their goals, from

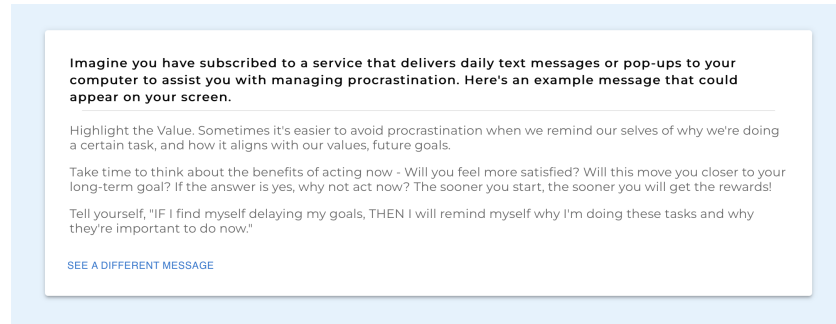


Figure 1: The screen through which participants saw the seed message

Table 1: Example Seed Messages. The psychological principles indicated by arrows were not part of the main text but are included here to annotate the different components of each message.

Number	Complete Message (and Associated Psychological Principles)
1	<p>Highlight the Value. Sometimes it's easier to avoid procrastination when we remind ourselves of why we're doing a certain task, and how it aligns with our values, future goals. → Cognitive Insight</p> <p>Take time to think about the benefits of acting now - Will you feel more satisfied? Will this move you closer to your long-term goal? If the answer is yes, why not act now? The sooner you start, the sooner you will get the rewards! → Value Alignment</p> <p>Tell yourself, "If I find myself delaying my goals, THEN I will remind myself why I'm doing these tasks and why they're important to do now." → Implementation Intention</p>
2	<p>Future Is Real. We can know something without realizing it - without it becoming "real" and relevant to our lives. When you procrastinate on exercising, you are likely aware of the negative consequences on your physical health, but you may not really care about it because the negative consequences can take fairly long to be detrimental. → Cognitive Insight</p> <p>However, that day will come sooner or later. When you are about to procrastinate, think about the negative consequences and how they snowball. The task may not seem pressing now, but would it still be fine if you keep procrastinating on it? How would it change your life trajectory? The future is not far away, it is REAL. Don't leave guilt and regrets until it is too late. → Psychological Flexibility</p> <p>Tell yourself, "If I find myself putting off a task that doesn't seem pressing, THEN I will remind myself that things often don't seem pressing until it's too late. I will get started on the task." → Implementation Intention</p>
3	<p>Own Yourself. Overcoming procrastination is about self-regulation. We know we shouldn't procrastinate, but we are unable or unwilling to regulate ourselves to act because procrastination provides immediate mood repair. → Cognitive Insight</p> <p>When you are about to give in to feel good, always remind yourself - you are the owner of yourself, you can control yourself, and you should control yourself. You wouldn't wish to sacrifice something more important for immediate mood repair. You wouldn't let desires take control over you. → Psychological Flexibility</p> <p>Tell yourself, "If I find myself saying "I feel like it tomorrow", THEN I know it signals giving in, and I will just start on the task that I'm procrastinating." → Implementation Intention</p>

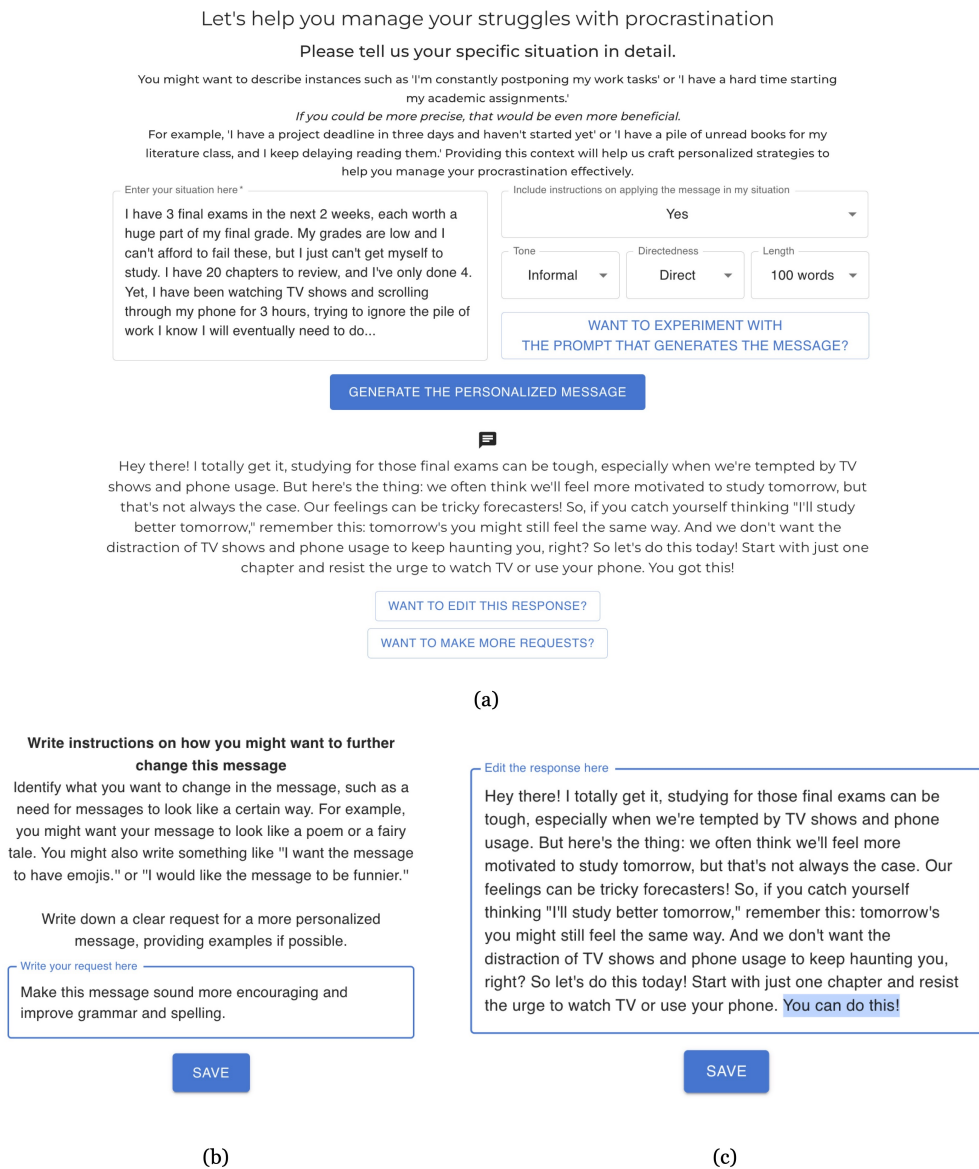
the immediate future to more distant timeframes, thereby enabling incremental action planning. In addition to aiding in the planning process, this feature also prompted participants to consider their underlying motivations and anticipated achievements [89].

While participants were given the autonomy to draft their future-oriented email independently, the LLM was available to assist in structuring and refining their thoughts. Participants had the flexibility to make open-ended requests for editing their drafts. They could also utilize specific keywords to conceptualize their future actions more effectively. These keywords were provided to serve as

semantic anchors that can help individuals better align their actions with their core values and goals, thereby enhancing the potential for successful implementation [35]. These features are illustrated in Figure 3.

3.4 Instructions and Examples

Throughout the probe, participants were offered instructions and sample inputs for open-ended response boxes. The aim of providing these guidelines was to steer the participants toward relevant and effective suggestions [41, 88]. These examples and instructions were



Write a message about your future self

How do you see yourself one week from now, having made progress with your procrastination? What about in three months? And in ten years? Envision your future self at these distinct stages. Think about the personality traits you may have strengthened, the goals you could have achieved, the improved relationships due to better time management, and your academic or professional accomplishments that you have made possible by overcoming procrastination.

My future self is... *

The next two weeks are the final weeks of my school year. They will be hectic and stressful, filled with final exams, final projects, and the release of my final report card. At the end of these two weeks, I want to know that I have tried my best in studying for all 4 exams and walk out of the exams with a feeling of relief and satisfaction. For my final project, I want to have finished it 2 days before the deadline, unlike how I usually leave things until the last second. Finally, I wish that when I receive my report card in three weeks, my GPA is 3.5 or above to maintain on the Dean's List.

NEED HELP IN DRAFTING THE MESSAGE?

NEED HELP IN IMAGINING THE FUTURE?

Email this to my future self

Email address *

Date selector

09/05/2023

Basic time picker

07:58 PM

SEND

(a)

Consider providing detailed instructions about how you'd like your message to be improved. These could include changes in its style, format, tone, or the inclusion of specific elements such as emojis. For instance, you may want your text to have a more formal or conversational tone, require a grammar and typo check, or need it to resemble a certain writing style.

How can we help you? *

Help me sound more excited about the future.

SAVE

(b)

Select the following term that best describes your desired future:

☐ Successful: achievement, prosperity, impact, growth

☐ Happy: fulfillment, joy, gratitude, serenity, meaning

☐ Free: independence, autonomy, exploration, authenticity

☐ Exciting: adventure, novelty, growth, spontaneity

☐ Balanced: well-being, contentment, health, harmony

☒ Choose my own keywords

Insert list of keywords here *

Independent: self-sufficient, self-made, liberated, bold

Our system will tailor your message according to the future you describe.

MODIFY MESSAGE

(c)

Figure 3: Features for drafting an email to future self: (a) an open-ended text box for drafting the message, (b) a feature enabling open-ended modifications to the draft, (c) a list of suggested keywords for envisioning the future.

could use as a starting point. These example prompts are displayed in Figure 4.

4 METHODS

We involved 15 university students in our study through interviews and focus group discussions and subsequently corroborated our findings through interviews with six experts in clinical psychology, learning and education, and cognitive psychology. The following section provides details on the methodology and logistics that shaped our study.

4.1 Participants

We recruited 15 individuals for our study through email invitations and social media calls. Participants were required to be between

the ages of 18 and 25, currently enrolled at a university, and residing in North America. Additionally, they had to self-identify as at least average procrastinators, scoring 24 or higher according to the Irrational Procrastination Scale (IPS) [102]. The mean age of our participants is 21.6 ± 0.6 years old. They identified with two genders (10 women, 5 men; other options were offered) and several racial groups (8 Asian, 4 White, 2 African American, 1 Mixed-Race). We refer to these participants as P1–P15.

To obtain a professional perspective on participant feedback, we also conducted interviews with six experts. We refer to these experts as E1–E6. These participants were also recruited from email invitations and social media calls. All of them had post-graduate degrees from diverse fields including clinical psychology (E1 and E2), learning and education (E3–E6), and cognitive psychology (E4 and E6). They identified with two genders (3 women and 3 men;

The screenshot shows a web interface titled "Select a prompt:". It contains five radio button options for rewriting a message:

- ☐ Rewrite this message but make it in the tone of a coach trying to motivate a person.
- ☐ Rewrite this message but include a relevant quote.
- ☐ Rewrite this message but elaborate and add a bit more psychological theory and examples.
- ☐ Rewrite this message but keep it as short and to the point as possible, without any unnecessary information.
- ☒ Create my own prompt

Below the "Create my own prompt" option is a text input field with the placeholder "Insert your prompt here*". The field contains the text: "Rewrite the message in the tone of a mother trying to comfort her child." Below the input field is a blue "SAVE" button.

Figure 4: Prompts to generate personalized messages

other options were offered) and belonged to two racial groups (5 White and 1 Asian).

4.2 Procedure

Our research protocol was approved by the Research Ethics Board at the first author's institution. Recognizing the sensitive nature of our research topic, we anticipated that some participants might be more comfortable sharing their experiences and preferences in a one-on-one setting, while others might find a group discussion more conducive for collaborative idea generation. Therefore, we collected feedback through both individual interviews and focus group discussions. Ten individuals gave individual interviews (P1–P10) and five took part in a focus group discussion session (P11–P15). Two members of the research team organized the interviews and focus groups using the Google Meet videoconferencing platform. Interview durations ranged from 30 to 50 minutes, while the focus group discussion lasted 90 minutes.

The interviews and focus groups employed a semi-structured approach to facilitate flexible investigation. After being given access to SPARK through a publicly hosted link, participants were given a brief overview of the study's objectives and a walkthrough of the tool according to the four main components as detailed in Section 3. Participants were then instructed to share their screens as they explored SPARK on their own for 10 minutes; some participants requested and were granted additional time for this part of the protocol (P1, P3, P4, P6, P9, and P10). While this exploration was self-directed, the typical user interaction proceeded as follows:

- (1) Participants initially experimented with various versions of the seed messages (Figure 1). This initial exploration typically lasted between 30 seconds and 2 minutes.
- (2) They then explored different interface options as shown in Figures 2(a) and 4. Participants often sought clarification on the extent of detail required for their inputs, to which we clarified that they were free to provide as little or as much as they wanted. This phase of interaction took 6 to 8 minutes.
- (3) The next step involved participants generating personalized messages using SPARK. They either requested further

customization from the LLM or modified the messages themselves, as depicted in Figures 2(b) and (c) respectively. This step usually spanned 2 to 4 minutes.

- (4) Finally, participants engaged with the email drafting activity for future actions (Figure 3). Six participants composed short emails to be sent within the session's duration, while others experimented with different features. Participants spent 2 to 4 minutes on this activity.

It is important to note that participants sometimes navigated non-linearly between the features of SPARK. Therefore, the times reported here represent the approximate duration for each stage of interaction.

Next, participants were invited to give their thoughts on the utility of the tool and potential features they would find beneficial. The focus group session adhered to a similar structure, with the primary difference being that participants in the focus group were given turns to share their opinions. Questions posed in interviews and the focus group included, but were not limited to:

- What kinds of interface options or customization features would you like to see in a tool for managing academic procrastination?
- How much effort are you willing to invest in detailing your procrastination issues in an open-ended text box?
- What kinds of resources or guidance might help users effectively personalize messages using LLMs for managing academic procrastination?
- How do you see such tools fitting into your daily routine for managing academic procrastination?

All participants were compensated with \$20 USD per hour for their involvement in the study.

Following the completion of participant interviews, we applied thematic analysis to extract themes from the interview and focus group transcripts (Section 4.3). We then interviewed the experts to solicit feedback on these findings. These discussions also began with an introduction to the goals of our study and a walkthrough of SPARK, after which we presented a summary of the findings

from our thematic analysis. For each theme, our questions for the experts centered around two broad questions:

- What are the potential benefits and risks of implementing the features related to this theme?
- What ethical challenges or considerations should be anticipated when implementing these features?

The expert interviews were conducted via Google Meet and had durations ranging from 45 to 60 minutes.

4.3 Data Analysis

The interviews and focus group discussions yielded approximately 9.5 hours of audio recordings, which translated into 172 pages of transcribed data. All personal identifiers were removed from the transcripts to ensure anonymity, and the data was then thoroughly cleaned for analysis. We followed a thematic analysis approach [15, 23, 77] to analyze the collected data. Two team members took on the role of coders and reviewed all data to familiarize themselves. The coders then began the process of assigning specific codes to three interview transcripts using an open-coding approach [56]. Each coder individually created an initial codebook.

We subsequently employed a consensus coding approach akin to the process described by Braun and Clarke [15]. The coders engaged in multiple collaborative sessions where they compared and merged their individual codebooks. This process of consensus coding with iterative discussions and refinement aimed to enhance consistency and reduce individual biases [42, 77]. Coders identified overlapping codes, clarified code definitions, and excluded codes not closely aligned with the research questions. The refined codebook was then applied iteratively to three additional transcripts, allowing for further adjustments based on emerging insights. Once a final version of the codebook¹ was agreed upon, each coder independently applied the established codes to half of the remaining data.

Our analysis generated 19 codes, such as “need for direct solutions”, “step-by-step guidelines”, “use of calendars”, and “clear disclaimers”. The coders arranged multiple meetings to group the codes into broader themes through axial coding [118]. We organized our findings around the themes that are presented in Section 5.

4.4 Ethical Considerations

Given the nature of our research, which touches upon aspects of procrastination, stress management, and goal setting, we were mindful of the ethical considerations involved. Throughout our work, we took measures to address these concerns. To ensure participant comfort and safety, we informed them at the outset of the interviews that they had the option to skip any questions or discontinue the conversation if they felt uncomfortable at any point. Our interviewers were trained in the Columbia-Suicide Risk Assessment protocol [17] and prepared to respond appropriately if participants expressed thoughts of suicide or self-harm. They were equipped to provide safety planning or make referrals to crisis services as necessary. However, none of these risks emerged during the study.

¹The definitions of each code from the final codebook are provided in Appendix A.

5 FINDINGS

Participants explored various strategies for engaging with LLM-based tools to tackle their procrastination challenges during their engagement with the probe. For instance, P1 and P12 envisioned asking the probe to divide a large writing task into smaller, actionable steps. P6 contemplated using LLM-based tools to generate motivational messages tailored to different situations. Additionally, P2 and P10 requested help in creating a daily schedule to better manage their tasks and responsibilities.

In this section, we delineate various themes that emerge concerning the ways in which users perceive the utility of LLM-based tools for managing academic procrastination. Additionally, we discuss how expert opinions either corroborate or diverge from these user perspectives. Table 2 shows the correspondence between themes and codes as well as the number of participants and experts who mentioned them.

5.1 Aspirations for Structured Action Steps

Participants generally appreciated the flexibility offered by the tool (e.g., the ability to select a preferred tone or directedness) for message customization. They noted that customization not only enhanced the utility of the tool but also gave them a greater sense of control over their interaction. P1 commented:

“I think the tone and the length I’d love to play around with, I’m trying to imagine myself as a user. ChatGPT has a tendency to maintain the same tone in all of its messages. I don’t think I’m going to the effort of putting this information there. But here people can just directly choose what they want to click in to use.”

Participants also raised potential ideas for additional customization features. For example, some vouched for the ability to switch between action-oriented and suggestive tones, and others suggested the ability to enhance instructions with pertinent resources (e.g., links to studies or relevant websites). P12 stated that sometimes users may express a desire to understand the reasoning behind a step or to explore alternative approaches, so making such resources available could satiate this curiosity and augment user involvement. However, despite these design opportunities identified by participants, some of them also expressed the concern that an excessive number of options might be overwhelming. To address this, P2 suggested a two-tiered approach: presenting basic customization features on the main page and more advanced options on a separate page.

Extending this need for customization, participants in our study wished to receive structured guidance from the system to carry out their tasks. Many of them emphasized the challenges they face with unstructured or expansive tasks. The widespread view was that tasks appear more manageable when presented in a clear, step-by-step sequence. P8 drew upon a cooking analogy, referencing their preference for following a recipe versus freestyle cooking without any guidance. Beyond ease of use, participants also identified psychological benefits to seeing a concrete action plan. For instance, P2 noted how users could feel fulfillment upon completing individual steps for larger tasks.

Expert insights also underscored the value of offering users assistance in breaking down larger tasks. E1, however, cautioned

Table 2: Breakdown of themes from participant (P) and expert (E) responses.

Themes	Codes	Number of Individuals Mentioned
Aspirations for structured action steps	Need for direct solutions	13P + 2E
	Step-by-step guidelines	10P + 4E
	Structuring recommendations	10P + 3E
	Overreliance	2P + 5E
	Adaptive breakdown	3P + 2E
Deadline-driven instructions	Reliance on deadlines	12P + 4E
	Use of calendars	7P + 2E
	Documenting goals	8P + 5E
	Need for reminders	7P + 4E
	Diversity in motivation	2P + 4E
Guided versus unguided questions	Varied levels of guidance	6P + 5E
	Providing multiple layers of features	7P + 3E
	Flexible engagement	3P + 2E
Concerns and boundaries in providing emotional support	Need for emotional support	9P
	Limitations of LLMs in addressing emotions	1P + 6E
	Clear disclaimers	5E
Providing support on the use of LLM-based tools	Examples and instructions	5P + 4E
	Collaborative elements	4P + 2E
	Feedback to LLMs	5P + 1E

that while a comprehensive plan can be constructive, there is a risk of overwhelming students. Their recommendation was to detail the imminent steps while providing a broader outline for later tasks. As users navigate through the initial steps, future steps and details can be introduced. E3 and E6 also raised concerns about fostering over-reliance on the tool. E3 expressed:

“While structured guidance is valuable, we must ensure students retain their ability for independent thought and action. The danger lies in spoon-feeding them precise answers or solutions. It’s crucial for students to develop their own thinking and information-gathering skills. Relying too heavily on a step-by-step model, especially one presented by an LLM, might foster a false sense of accomplishment. Students might follow every instruction and still wonder why their outcomes don’t match their expectations.”

In summary, while structured guidance was seen as vital for task management, experts recommended a balance should be struck to ensure learners’ independence and critical thinking remain intact.

5.2 Deadline-Driven Interactions

Participants appreciated the tool’s ability to allow them to draft future emails to themselves. This feature was seen as especially beneficial for managing important deadlines. For instance, P10 indicated they would leverage this capability as a way to document their goals, with the periodic emails serving as reminders to review and refocus on these objectives. P2 and P4 further elaborated that

the act of drafting these emails would also assist them in gaining better clarity regarding their goals and plans.

For better management of deadlines, however, several participants expressed an interest in integrating a tool like ours into their daily workflow for on-the-spot, personalized advice and actionable plans. Participants like P2 and P3 suggested integrations with platforms they already use for professional purposes, such as Google Calendar, Microsoft Teams, and Slack. In this context, a recurrent suggestion was the addition of a feature allowing participants to specify task deadlines. They believed that by doing so, the tool would be better positioned to deliver timely and pertinent recommendations. P2 shared an illustrative perspective:

“I have a friend who will be like, ‘Yeah, you have this coming up, you have that coming up. So you better do this first and do that first and prioritize this because this is more work than others.’ So the AI can somewhat do something similar where you can be like, ‘You have this coming up. I think you should prioritize this.’ It’s not like it is telling me to do it. ...It’s just that, like a gentle reminder or something of that fashion.”

Some participants offered thoughts regarding the optimal frequency and timing of reminders from the tool. For instance, P6 emphasized the importance of receiving motivational cues significantly ahead of an impending deadline to prevent last-minute rushes. They went on to suggest incorporating features that allow users to set task priorities and provide estimates for completion times, tailoring the tool’s reminders to their individual needs.

Experts largely agreed with the participants but emphasized the diverse motivational needs people require at different times. E2 suggested multiple strategies to this end: for those who feel regret reflecting on past procrastination, a message like “The past is in the past, focus on today”; for specific barriers, a response inspired by WOOP (wish, outcome, obstacle, plan) goal-setting model [82]; and for a lack of perceived purpose, affirmations connecting tasks to future goals or values. Furthermore, E1 posited that presenting an array of motivational prompts could enable users to select the one that resonates most with them.

5.3 Guided Versus Unguided Questions

In our prototype, participants had the flexibility to elaborate on their situations to whatever extent they desired. Participants embraced this feature as an opportunity for them to mold canned messages to fit their unique context. However, opinions varied when discussing the nature and quantity of questions the tool used to probe users for more information. Some participants advocated for a guided approach featuring a sequence of open-ended questions, believing it would help frame their responses more effectively. P7 illustrated this preference by stating,

“The first problem is ‘What is something you are putting off for today?’ And then you would write, ‘I’m putting off x and y’. And then after you type and press [the continue button], it might prompt you to say, ‘How has that made you feel?’ And then it might prompt you to say, ‘What made me feel this way?’”

Conversely, other participants were concerned that a sequence of questions could serve as a new avenue for procrastination. P2 emphasized the potential to overwhelm users with numerous prompts, especially during initial interactions with the platform. Additionally, building on past experiences with self-management tools, Participant P4 noted that a barrage of questions in the midst of a busy schedule could also heighten stress levels. In scenarios like these, they predicted that users would prefer to receive personalized guidance that is both immediate and easy to access, without the need to navigate through a multitude of questions.

To mitigate these concerns, E1, E3, and E6 suggested the inclusion of adaptive contextual factors to adjust both the quantity and type of questions posed to the user. They recommended employing contextual variables like the time of day, day of the week, and the user’s schedule to fine-tune how the LLM elicits more information about the user’s situation. While these experts generally supported providing longer interactions informed by psychological literature [92, 97], they also recognized that users should have the option for shorter interactions when experiencing higher levels of activity or stress.

5.4 Concerns and Boundaries in Providing Emotional Support

The recognition and acknowledgment of user emotions emerged as a significant theme, with diverse opinions shared by both participants and experts. They recognized that discussions of procrastination could often lead to revelations about mental states, including symptoms of anxiety and depression. In such cases, some participants expressed the need for emotional support. P4 said:

“Assuming I already signed up for it, then I might like more emotional support. Like actual links [saying] ‘Here are resources that you can use’. Maybe there will be options [for showing] pep talks or something like that.”

Participants like P7 felt that the tool could be proactive in inquiring users about their emotions and the possible culprits behind negative feelings. P11 went further, expressing an interest in having access to resources for managing more serious and vulnerable issues such as depression.

Nevertheless, experts advised caution regarding the prospect of LLMs taking on roles akin to therapists for providing support in the management of emotions. Their concern stemmed from the possibility that these models could dispense advice that, while perhaps appealing to users, may lack clinical validity. Instead, experts like E1, E4, and E5 emphasized that the tools should only simply recognize users’ emotions while clearly communicating their limitations in providing therapeutic support. E3 said:

“First, acknowledge that the individual is struggling. You know, thank them for the description kind of thing. ... Just say this, ‘This kind of falls out of the scope of what I am able to assist you with.’”

E5 elaborated on the need for transparency regarding the tool’s limitations, emphasizing that users should be clearly informed from the outset about the tool’s capabilities and constraints. This includes an explicit declaration at the introduction page outlining the tool’s capacities and the message that any unintentional suggestion regarding emotion management should not be acted upon by users. They further stated that the system should contain contact information for university support services or crisis management centers for further assistance and support.

5.5 Providing Support on the Use of LLM-based Tools

Although LLMs are meant to be conversational and intuitive, participants in our study still desired guidance on how to best phrase their responses to get the tool to behave as they anticipated. Some felt that video tutorials and demonstrations might be useful, but there were stronger calls for concrete examples of inputs and outputs. P6 noted that the example prompts in Figure 4 helped them learn about the kinds of requests they could make and showed the range of possible outputs they could expect. Extending on this idea further, participants proposed the idea of adding a collaborative element so that users could help one another optimally use the tool. P2 expanded on this concept by saying:

“Maybe share with me what other people are saying as well. So yeah, make it more like a social thing. I would also want to see what other people are asking for in their messages.”

Participants also expressed an interest in providing feedback on the advice generated by the tool so that it would work better for them in the future. Again, they were unsure about how this could be accomplished, so they sought instructions and concrete examples that could assist them in this aspect as well. P2 and P9 went on to suggest that such tools should offer users ways of providing feedback beyond inputs to the LLM. For instance, they envisioned

a star-rating system that allows users to evaluate tailored advice on perceived qualities such as relevance, applicability, and level of detail.

Experts concurred with the participants' suggestions, adding an extra layer of caution. E5 stressed that tools need to explicitly distinguish the contexts in which they can be helpful from those in which they cannot. Accordingly, the examples provided to users should be designed to make this distinction abundantly clear so that users are fully aware of the tool's capabilities and limitations.

6 DISCUSSION

In our discussion, we reflect on how our findings answer our research questions and then use those insights to generate design recommendations for LLM-based tools for managing academic procrastination. We then highlight our works' contributions to the broader literature on human-LLM collaboration. Finally, we enumerate the limitations of our study along with avenues for future research.

6.1 RQ1: Users' Envisioning of LLMs in Tailoring Strategies for Managing Academic Procrastination

6.1.1 Structured Guidance with Adaptability. In alignment with existing literature on procrastination management and structured goal-setting [16, 26, 36, 44, 64, 89], our study reiterates the importance of breaking down larger tasks into more manageable components. LLMs emerge as particularly adept for this purpose, offering the capability to provide detailed, actionable steps for immediate tasks while also delivering overarching, abstract guidelines for long-term objectives [71, 79]. This dual approach of LLMs caters to both immediate task execution and longer-term goal planning.

Design Recommendations: Future procrastination management tools should strive to achieve a balance between providing structured guidance and maintaining adaptability. Ensuring that tools are both systematic in their approach to task management and flexible enough to adapt to individual user requirements could support users in a wide range of contexts [12, 30, 55, 68, 70]. For LLM-based tools specifically, an important feature could be the development of a *dynamic structured guidance system*, which would generate hierarchical task lists dynamically based on user input [71, 79, 83]. Including functionality to automatically break down large tasks into manageable micro-tasks with clear deadlines would further enhance this system. Such a feature would not only provide structured task management but also adapt to the specific needs of each user. To enhance user engagement and provide comprehensive support, the system could also offer supplemental resources that are tailored to each specific task [95].

6.1.2 Integration with User Routines and Deadlines. Participants in our study highlighted the necessity for deadline-oriented interactions facilitated by LLMs as a means of seamlessly integrating into their daily academic activities. However, it is important to acknowledge the individual variations in perceived effort related to both the target task and other concurrent responsibilities. This aspect of our findings enhances and broadens the scope of existing literature,

which underscores that effective deadline management should encompass more than just scheduling; it needs to take into account the comprehensive context of a user's life [5, 31, 46, 76, 121, 123]. Building on this, our research advocates for a dynamic approach to task prioritization, one that is attuned to the urgency of deadlines while being sensitive to the individual's overall workload and personal commitments.

Design Recommendations: Future procrastination management tools could benefit from synchronization with commonly used platforms such as Google Calendar, Microsoft Teams, and Slack. This integration would facilitate the automatic import of scheduled tasks and deadlines, thereby aiding in task organization [26, 36, 43]. Additionally, integrating mobile phone sensor data could provide further information about users' activity levels and sleep patterns, potentially enhancing the relevance of task management suggestions [27, 93]. Based on this imported information and any user-inputted priorities, LLM-based tools could then dynamically generate a ranked task list. These tools could also include a *effort estimator* to allow users to adjust task rankings based on their estimated completion time and task difficulty [2].

6.1.3 Collaborative Support in Using LLM-based Tools. Participants of our study expressed interest in learning from others' examples to both navigate our LLM-based tool and to learn about effective strategies for combating procrastination. This reflects the broader psychological need for collaborative support in combating procrastination [18, 45, 108, 110]. Implementing such collaborative support features could aid users in navigating the complexities of interacting with the LLM-based tool and offer them guidelines or best practices through the experiences of others [11].

Design Recommendations: Future procrastination management tools could provide users with a communal repository where they could access a variety of strategies and solutions that have been effective in managing tasks and deadlines, bearing in mind that user contributions would need to be thoroughly de-identified to respect people's privacy [104]. Specifically for LLM-based tools, this concept could be further developed into a shared prompt library where users could view and contribute diverse examples of prompts and the corresponding responses generated by the LLM. This feature would be particularly useful as it would allow users to explore the range of the tool's capabilities, providing insights into how to interact more effectively with it. Furthermore, allowing users to rate the utility, relevance, and clarity of the LLM's responses in the communal repository could provide an effective means for users to identify the examples that best suit their needs.

6.2 RQ2: Potential Limitations and Tensions

6.2.1 Boundaries in Addressing Emotions. Our study sheds light on the limitations of LLMs in the realm of procrastination management, particularly concerning emotional and therapeutic support. These findings align with recent skepticism among clinicians regarding the suitability of LLMs for emotional counseling and therapeutic support [52, 54, 79]. While LLMs inherently lack the ability to discern human emotions, they might encounter user-expressed

emotions in written form. The experts consulted in our study underscored the need for LLM-based tools to acknowledge and validate these emotions without crossing into the territory of clinical or therapeutic advice [52]. Their comments reinforce the importance of carefully defining the scope and limitations of LLM-based tools in managing tasks and emotions, ensuring their responsible and effective use.

Design Recommendations: Future procrastination management tools should prominently feature a detailed disclaimer on the landing page to ensure clarity and responsible usage [11, 12]. This disclaimer should concisely state the purpose and limitations of the tool, particularly in relation to emotional and psychological support. In the context of LLM-based tools, disclaimers should highlight that while they can assist with procrastination management strategies like scheduling or time management, they are not equipped to provide therapeutic support [32, 63]. For users who might require specialized emotional or psychological assistance, the system should provide a place where users can access a vetted list of well-established resources for mental health support: informational websites, helpline numbers, or links to accredited mental health service providers [11, 12].

6.2.2 Balancing Guidance and Independent Exploration. The expert interviewees expressed concerns about the potential for students to become overreliant on LLM-based tools [50, 54, 79]. They warned that easy access to a tool that breaks down complex tasks might inhibit the development of independent problem-solving skills. This aligns with recent literature that emphasizes the need for a balance between providing guided assistance and fostering independent problem-solving skills [1, 104].

Design Recommendations: Future procrastination management tools may occasionally restrict users' access to functionalities that produce direct step-by-step solutions and instead prompt them to independently seek out potential strategies [22, 59]. These restrictions could be strategically timed, for instance, after a user completes a certain number of sub-tasks or reaches key milestones in a project. LLM-based tools could go a step further by generating *independent exploration prompts* (e.g., “research further”, “think critically”) that encourage problem-solving strategies [51, 73]. Each option could come with a brief description of what the activity entails and why it might be beneficial, providing a nudge without forcing a specific course of action [9]. Users could also have the option to schedule independent exploration prompts at intervals of their choosing, offering customization to better suit individual work habits and needs [9].

6.2.3 Allowing Flexible Engagement. We recognize a design tension regarding the extent of guidance users may wish to see when describing their situation. In some instances, users may prefer a set of structured questions to shape their responses, which may be more feasible during idle periods and moments of downtime [12, 90]. Conversely, a single open-ended question could be more suitable when users are initially acquainting themselves with the system or are in the midst of many other activities [62]. The same contextual variability applies to the level of customization available in the tool, whether it be related to the tone, directedness, or length

of the advice given [6, 34, 75]. While these features may be beneficial, there are moments when the user may not have the bandwidth to engage with such granularity and would prefer the option to bypass these decisions to avoid overwhelm [12].

Design Recommendations: Our findings highlight the need for procrastination management tools that permit flexible engagement, allowing users to maintain a minimal level of interaction even when they are not fully engaged. Studies in educational psychology and social media behaviors indicate that different online activities demand various levels of cognitive involvement [12, 19, 39, 113]. While active engagement is typically more beneficial for enhancing learning and wellbeing, our findings suggest that there are times when users may be reluctant to invest significant cognitive effort, particularly during periods of low energy or mood [12]. In such instances, LLM-based tools should allow users to access key functionalities without demanding intensive engagement. While our findings are situated within the use of an LLM-based probe, this need for flexible engagement could apply to general procrastination management tools as well.

6.3 Contributions to Human-LLM Collaboration

Our findings, although mainly situated within the context of managing academic procrastination, contribute to the evolving body of literature on human-LLM collaboration. This study highlights the nuanced ways in which users engage with LLM-based tools, underscoring the potential for these models to offer more than just information processing; they can act as collaborative partners in personal and academic development [21, 29, 38, 57, 65]. The desire for structured action steps and deadline-driven interactions aligns with the broader understanding of how individuals seek both assistance and partnership in managing their tasks. This perspective is bolstered by recent studies showing the promise of LLMs in exhibiting social intelligence and human-like collaborative behaviors [38, 86, 98, 125]. With this paradigm shift in AI that is moving towards systems that are capable of complex and empathetic interactions, there is a need to understand the unique contexts of individual users in order to enhance their downstream utility.

The varied opinions on structured questioning and the concerns raised about emotional support challenges resonate with ongoing discussions about the ethical and practical implications of human-LLM interactions [52, 100, 104]. Our findings suggest that while users appreciate the personalized assistance offered by LLMs, they also value retaining a degree of autonomy and control over the interaction. This balance between guidance and independence could be a critical aspect of designing effective LLM-based tools, as it may directly impact user engagement and satisfaction [10]. Moreover, the caution advised by experts about the role of LLMs in emotional support and the need for transparency in their capabilities reflects a growing awareness of the ethical responsibilities inherent in deploying AI in sensitive areas like mental health [52, 54, 79]. The ethical implications of AI tools offering mental health support are complex, particularly given the risk of users relying on these tools for serious mental health issues without professional guidance [54]. Our study contributes to the broader discourse on the ethical deployment of AI in sensitive domains, advocating for a cautious,

well-informed approach that prioritizes user safety and wellbeing while exploring the beneficial aspects of human-LLM collaboration.

6.4 Limitations and Future Works

In acknowledging the limitations of our research, we first note that our participant group was composed of individuals residing in North America at the time of the study. Although this cohort included people from various racial groups, the geographical and cultural context in which they live may have influenced their perspectives on the use of LLM-based tools for managing procrastination. Future studies should aim to include a more geographically and culturally diverse participant group. Engaging with populations from different regions and cultural backgrounds would provide a broader perspective on the use of LLM-based tools for managing procrastination, potentially uncovering unique challenges, preferences, and expectations.

Additionally, we designed a technology probe to explore user expectations regarding LLM-based tools for managing academic procrastination. To mitigate academic and emotional risks that could arise from deploying an unvalidated tool, we chose not to conduct a longitudinal study without researcher oversight. Nevertheless, we recognize the potential benefits of long-term studies for a more comprehensive understanding of the real-world effectiveness and long-term implications of LLM-based tools in this context. With appropriate researcher oversight and ethical considerations, future studies could entail longitudinal deployments that examine the impact of various features on acceptability and perceived effectiveness.

Finally, our findings are situated in the domain of academic procrastination management and, therefore, might not directly translate to other behavioral or psychological areas. Future research could broaden its scope to assess the efficacy of LLM-based tools in tackling a diverse array of challenges. Such exploration could range from addressing mental health concerns to enhancing personal productivity or promoting healthy behaviors. Investigating the application of these tools in varied scenarios will provide a deeper understanding of their adaptability and constraints, enriching the knowledge base about the full potential of LLM-based tools in different contexts.

7 CONCLUSION

Traditional interventions often fail to capture the unique, person-specific factors that contribute to academic procrastination. LLMs offer a promising avenue to bridge this gap, as they allow for open-ended, individualized responses through sophisticated data analysis and natural language processing capabilities. Yet, the expectations and constraints of employing LLMs for this purpose have not been thoroughly investigated. To fill this knowledge void, we conducted in-depth interviews and focus group discussions with 15 university students and consulted six experts, using a technology probe for context. Our findings underscore the need for LLM-based tools to offer structured guidance with a focus on deadlines and to include enhanced mechanisms for user support. The results also emphasize the importance of adaptability in the query process, taking into

account factors such as the individual's level of busyness, and advocate for caution when considering the use of LLMs for therapeutic advice. These insights enabled us to make several design recommendations surrounding structured guidance with adaptive complexity, boundaries in addressing user emotions, and collaborative support in using LLMs. We believe our findings and recommendations will serve as important guidelines for developers, UX/UI designers, and researchers working on LLM-based intervention tools.

ACKNOWLEDGMENTS

We thank Pranav Rao, Cynthia Sa, and Carolyn Lee for their help in collecting data, drawing figures, and proofreading the manuscript. This work was supported by grants from the National Institute of Mental Health (K01MH125172, R34MH124960), the Office of Naval Research (N00014-18-1-2755, N00014-21-1-2576), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-06968), and the National Science Foundation (2209819). In addition, we acknowledge a gift from the Microsoft AI for Accessibility program to the Center for Behavioral Intervention Technologies that, in part, supported this work (<http://aka.ms/ai4a>), and an award from the Wolfond Scholarship Program in Wireless Information Technology.

REFERENCES

- [1] Alaa Abd-Alrazak, Rawan AlSaad, Dari Alhuwail, Arfan Ahmed, Pdraig Mark Healy, Syed Latifi, Sarah Aziz, Rafat Damsch, Sadam Alabed Alrazak, Javaid Sheikh, et al. 2023. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Medical Education* 9, 1 (2023), e48291.
- [2] Elena Agapie, Patricia A Areán, Gary Hsieh, and Sean A Munson. 2022. A Longitudinal Goal Setting Model for Addressing Complex Personal Problems in Mental Health. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.
- [3] Riku Arakawa, Hiromu Yakura, and Masataka Goto. 2023. CatAlyst: Domain-Extensible Intervention for Preventing Task Procrastination Using Large Generative Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [4] Simon Arvidsson and Johan Axell. 2023. Prompt engineering guidelines for LLMs in Requirements Engineering. (2023).
- [5] Rachel Baker, Brent Evans, and Thomas Dee. 2016. A randomized experiment testing the efficacy of a scheduling nudge in a Massive Open Online Course (MOOC). *Aera Open* 2, 4 (2016), 2332858416674007.
- [6] Albert Bandura and Sebastian Wessels. 1994. *Self-efficacy*. Vol. 4. na.
- [7] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.
- [8] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [9] Ananya Bhattacharjee, Jiyau Pang, Angelina Liu, Alex Mariakakis, and Joseph Jay Williams. 2023. Design implications for one-way text messaging services that support psychological wellbeing. *ACM Transactions on Computer-Human Interaction* 30, 3 (2023), 1–29.
- [10] Ananya Bhattacharjee, Haochen Song, Xuening Wu, Justice Tomlinson, Mohi Reza, Akmar Ehsan Chowdhury, Nina Deliu, Thomas W Price, and Joseph Jay Williams. 2023. Informing Users about Data Imputation: Exploring the Design Space for Dealing with Non-Responses. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 11. 14–26.
- [11] Ananya Bhattacharjee, Joseph Jay Williams, Karrie Chou, Justice Tomlinson, Jonah Meyerhoff, Alex Mariakakis, and Rachel Kornfield. 2022. "I Kind of Bounce off It": Translating Mental Health Principles into Real Life Through Story-Based Text Messages. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW2 (2022), 1–31.
- [12] Ananya Bhattacharjee, Joseph Jay Williams, Jonah Meyerhoff, Harsh Kumar, Alex Mariakakis, and Rachel Kornfield. 2023. Investigating the Role of Context in the Delivery of Text Messages for Supporting Psychological Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

- [13] Aditi Bhutoria. 2022. Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence* 3 (2022), 100068.
- [14] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1077–1086.
- [15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [16] Jan Brechmann, Marc Hassenzahl, Matthias Laschke, and Marion Digel. 2013. ReMind: a transformational object for procrastinators. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2811–2812.
- [17] D Brent, C Lucas, M Gould, B Stanley, G Brown, P Fisher, J Zelazny, A Burke, et al. 2008. Columbia-suicide severity rating scale (C-SSRS). (2008).
- [18] Eleanor R Burgess, Kathryn E Ringland, Jennifer Nicholas, Ashley A Knapp, Jordan Eschler, David C Mohr, and Madhu C Reddy. 2019. "I think people are powerful" The Sociality of Individuals Managing Depression. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29.
- [19] Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social network activity and social well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1909–1912.
- [20] Rafael A Calvo, Dorian Peters, Daniel Johnson, and Yvonne Rogers. 2014. Autonomy in technology design. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 37–40.
- [21] Sze-yi Chan, Jiachen Li, Bingsheng Yao, Amama Mahmood, Chien-Ming Huang, Holly Jimison, Elizabeth D Mynatt, and Dakuo Wang. 2023. "Mango Mango, How to Let The Lettuce Dry Without A Spinner?": Exploring User Perceptions of Using An LLM-Based Conversational Assistant Toward Cooking Partner. *arXiv preprint arXiv:2310.05853* (2023).
- [22] Francesco Cirillo. 2018. *The Pomodoro technique: The acclaimed time-management system that has transformed how we work*. Currency.
- [23] Harris Ed Cooper, Paul M Camic, Debra L Long, AT Panter, David Ed Rindskopf, and Kenneth J Sher. 2012. *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association.
- [24] David Dale, Elena Voita, Loïc Barrault, and Marta R Costa-jussà. 2022. Detecting and Mitigating Hallucinations in Machine Translation: Model Internal Workings Alone Do Well, Sentence Similarity Even Better. *arXiv preprint arXiv:2212.08597* (2022).
- [25] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [26] Vedant Das Swain, Javier Hernandez, Brian Houck, Koustuv Saha, Jina Suh, Ahad Chaudhry, Tenny Cho, Wendy Guo, Shamsi Iqbal, and Mary P Czerwinski. 2023. Focused Time Saves Nine: Evaluating Computer-Assisted Protected Time for Hybrid Information Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [27] Nediya Daskalova, Jina Yoon, Yibing Wang, Cintia Araujo, Guillermo Beltran Jr, Nicole Nugent, John McGeary, Joseph Jay Williams, and Jeff Huang. 2020. SleepBandits: Guided flexible self-experiments for sleep. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [28] Victor Day, David Mensink, and Michael O'Sullivan. 2000. Patterns of academic procrastination. *Journal of College Reading and Learning* 30, 2 (2000), 120–134.
- [29] Paul Denny, Viraj Kumar, and Nasser Giacaman. 2023. Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 1136–1142.
- [30] Paul Denny, Juho Leinonen, James Prather, Andrew Luxton-Reilly, Thezyrie Amarouche, Brett A Becker, and Brent N Reeves. 2023. Promptly: Using Prompt Problems to Teach Learners How to Effectively Utilize AI Code Generators. *arXiv preprint arXiv:2307.16364* (2023).
- [31] Stephen H Edwards, Joshua Martin, and Clifford A Shaffer. 2015. Examining classroom interventions to reduce procrastination. In *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*. 254–259.
- [32] Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L Beam. 2021. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 3, 11 (2021), e745–e750.
- [33] Frederic Gmeiner and Nur Yildirim. 2023. Dimensions for Designing LLM-based Writing Support. In *In2Writing Workshop at CHI*.
- [34] Timo Gnamb, Markus Appel, and Bernad Batinic. 2010. Color red in web-based knowledge testing. *Computers in Human Behavior* 26, 6 (2010), 1625–1631.
- [35] Peter M Gollwitzer and Paschal Sheeran. 2006. Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in experimental social psychology* 38 (2006), 69–119.
- [36] Ted Grover, Kael Rowan, Jina Suh, Daniel McDuff, and Mary Czerwinski. 2020. Design and evaluation of intelligent agent prototypes for assistance with focus and productivity at work. In *Proceedings of the 25th international conference on intelligent user interfaces*. 390–400.
- [37] Carola Grunschel and Lena Schopenhauer. 2015. Why are students (not) motivated to change academic procrastination?: An investigation based on the transtheoretical model of change. *Journal of College Student Development* 56, 2 (2015), 187–200.
- [38] Rose Guingrich and Michael SA Graziano. 2023. Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines. *arXiv preprint arXiv:2311.10599* (2023).
- [39] Paul Haidet, Robert O Morgan, Kimberly O'malley, Betty Jeanne Moran, and Boyd F Richards. 2004. A controlled trial of active versus passive learning strategies in a large group setting. *Advances in Health Sciences Education* 9, 1 (2004), 15–27.
- [40] Mohanad Halaweh. 2023. ChatGPT in education: Strategies for responsible implementation. (2023).
- [41] Scarlett R Herring, Chia-Chen Chang, Jesse Krantzler, and Brian P Bailey. 2009. Getting inspired! Understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 87–96.
- [42] Clara E Hill, Barbara J Thompson, and Elizabeth Nutt Williams. 1997. A guide to conducting consensual qualitative research. *The counseling psychologist* 25, 4 (1997), 517–572.
- [43] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [44] Gary Hsieh, Jennifer Lai, Scott E Hudson, and Robert Kraut. 2008. Using tags to assist near-synchronous communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 223–226.
- [45] Chia En Hsueh, Hsinju Lee, Yu-Han Lu, Nien-Hsin Wu, Yu-Wei Wu, Shao En Lin, and Pei-Yi Kuo. 2020. Exploring the Effect of Group Size on Goal Setting & Sharing to Reduce Procrastination: Lessons Learned in a Field Study. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [46] William Humphrey Jr, Debbie Laverie, and Alison Shields. 2021. Exploring the effects of encouraging student performance with text assignment reminders. *Journal of Marketing Education* 43, 1 (2021), 91–102.
- [47] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 17–24.
- [48] Shima Imani, Liang Du, and Harsh Shrivastava. 2023. [Industry] MathPrompter: Mathematical Reasoning using Large Language Models. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.
- [49] Nanna Inie and Mircea F Lungu. 2021. Aiki-turning online procrastination into microlearning. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–13.
- [50] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [51] Hyungwook Jin, Minsuk Chang, and Juho Kim. 2019. SolveDeep: A system for supporting subgoal learning in online math problem solving. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.
- [52] Eunkyoung Jo, Daniel A Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [53] Kazi Sinthia Kabir, Stacey A Kenfield, Erin L Van Blarigan, June M Chan, and Jason Wiese. 2022. Ask the Users: A Case Study of Leveraging User-Centered Design for Designing Just-in-Time Adaptive Interventions (JITAs). *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–21.
- [54] Enkelejd 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.
- [55] Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J Ericson, David Weintrop, and Tovi Grossman. 2023. Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–23.
- [56] Shahedul Huq Khandkar. 2009. Open coding. *University of Calgary* 23, 2009 (2009).
- [57] Jinhee Kim, Hyunkyung Lee, and Young Hoan Cho. 2022. Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies* 27, 5 (2022), 6069–6104.

- [58] Yewon Kim, Mina Lee, Donghwi Kim, and Sung-Ju Lee. 2023. Towards Explainable AI Writing Assistants for Non-native English Speakers. *arXiv preprint arXiv:2304.02625* (2023).
- [59] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 272–283.
- [60] Felipe C Kitamura. 2023. ChatGPT is shaping the future of medical writing but still requires human judgment. , e230171 pages.
- [61] ByeongJo Kong, Erik Hemberg, Ana Bell, and Una-May O'Reilly. 2023. Investigating Student's Problem-solving Approaches in MOOCs using Natural Language Processing. In *LAK23: 13th International Learning Analytics and Knowledge Conference*. 262–272.
- [62] Rachel Kornfield, Renwen Zhang, Jennifer Nicholas, Stephen M Schueller, Scott A Cambo, David C Mohr, and Madhu Reddy. 2020. "Energy is a Finite Resource": Designing Technology to Support Individuals across Fluctuating Symptoms of Depression. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [63] Diane M Korngiebel and Sean D Mooney. 2021. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *NPJ Digital Medicine* 4, 1 (2021), 93.
- [64] Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S Bernstein. 2019. Conservation of procrastination: Do productivity interventions save time or just redistribute it?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [65] Harsh Kumar, David M Rothschild, Daniel G Goldstein, and Jake M Hofman. 2023. Math Education with Large Language Models: Peril or Promise? Available at SSRN 4641653 (2023).
- [66] Harsh Kumar, Yiyi Wang, Jiakai Shi, Ilya Musabirov, Norman AS Farb, and Joseph Jay Williams. 2023. Exploring the Use of Large Language Models for Improving the Awareness of Mindfulness. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [67] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.
- [68] Juho Leinonen, Arto Hellas, Sami Sarsa, Brent Reeves, Paul Denny, James Prather, and Brett A Becker. 2023. Using large language models to enhance programming error messages. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 563–569.
- [69] Q Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. 2022. Connecting algorithmic research and usage contexts: a perspective of contextualized evaluation for explainable AI. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 10. 147–159.
- [70] Mark Liffiton, Brad Sheese, Jaromir Savelka, and Paul Denny. 2023. CodeHelp: Using Large Language Models with Guardrails for Scalable Support in Programming Classes. *arXiv preprint arXiv:2308.06921* (2023).
- [71] Jiayin Lin, Geng Sun, Ghassan Beydoun, and Li Li. 2022. Applying machine translation and language modelling strategies for the recommendation task of micro learning service. *Educational Technology & Society* 25, 1 (2022), 205–212.
- [72] Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–23.
- [73] Dastyni Loksa, Amy J Ko, Will Jernigan, Alannah Oleson, Christopher J Mendez, and Margaret M Burnett. 2016. Programming, problem solving, and self-awareness: Effects of explicit guidance. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1449–1461.
- [74] Brady D Lund, Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang. 2023. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology* 74, 5 (2023), 570–581.
- [75] Bertram F Malle and Gale E Pearce. 2001. Attention to behavioral events during interaction: Two actor–observer gaps and three attempts to close them. *Journal of Personality and Social Psychology* 81, 2 (2001), 278.
- [76] Joshua Martin, Stephen H Edwards, and Clifford A Shaffer. 2015. The effects of procrastination interventions on programming project success. In *Proceedings of the eleventh annual International Conference on International Computing Education Research*. 3–11.
- [77] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.
- [78] Silvia Milano, Joshua A McGrane, and Sabina Leonelli. 2023. Large language models challenge the future of higher education. *Nature Machine Intelligence* 5, 4 (2023), 333–334.
- [79] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, et al. 2023. Exploring user perspectives on chatgpt: Applications, perceptions, and implications for ai-integrated education. *arXiv preprint arXiv:2305.13114* (2023).
- [80] Steven Moore, Richard Tong, Anjali Singh, Zitao Liu, Xiangen Hu, Yu Lu, Joleen Liang, Chen Cao, Hassan Khosravi, Paul Denny, et al. 2023. Empowering Education with LLMs-The Next-Gen Interface and Content Generation. In *International Conference on Artificial Intelligence in Education*. Springer, 32–37.
- [81] Stefan Morreel, Danny Mathysen, and Veronique Verhoeven. 2023. Aye, AI! ChatGPT passes multiple-choice family medicine exam. *Medical Teacher* 45, 6 (2023), 665–666.
- [82] Gabriele Oettingen. 2015. *Rethinking positive thinking: Inside the new science of motivation*. Current.
- [83] Kathleen O'Leary, Stephen M Schueller, Jacob O Wobbrock, and Wanda Pratt. 2018. "Suddenly, we got to become therapists for each other" Designing Peer Support Chats for Mental Health. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.
- [84] OpenAI. 2023. GPT-4 technical report. <https://arxiv.org/abs/2303.08774>
- [85] Artidoro Pagnoni, Vidhisha Balachandran, and Yulia Tsvetkov. 2021. Understanding factuality in abstractive summarization with FRANK: A benchmark for factuality metrics. *arXiv preprint arXiv:2104.13346* (2021).
- [86] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulators of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
- [87] John V Pavlik. 2023. Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator* 78, 1 (2023), 84–93.
- [88] Michael Pearce, Ashok K Goel, IL Kolodner, Craig Zimring, Lucas Sentosa, and Richard Billington. 1992. Case-based design support: A case study in architectural design. *Ieee Expert* 7, 5 (1992), 14–20.
- [89] Juanan Pereira and Óscar Diaz. 2021. Struggling to keep tabs on capstone projects: a chatbot to tackle student procrastination. *ACM Transactions on Computing Education (TOCE)* 22, 1 (2021), 1–22.
- [90] Erika S Poole. 2013. HCI and mobile health interventions: how human–computer interaction can contribute to successful mobile health interventions. *Translational behavioral medicine* 3, 4 (2013), 402–405.
- [91] Michael I Posner and Mary K Rothbart. 2000. Developing mechanisms of self-regulation. *Development and psychopathology* 12, 3 (2000), 427–441.
- [92] Timothy A Pynchyl. 2013. *Solving the procrastination puzzle: A concise guide to strategies for change*. TarcherPerigee.
- [93] Ruth Ravichandran, Sang-Wha Sien, Shwetak N Patel, Julie A Kientz, and Laura R Pina. 2017. Making sense of sleep sensors: How sleep sensing technologies support and undermine sleep health. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6864–6875.
- [94] Alexander Rozental, Erik Forsell, Andreas Svensson, Gerhard Andersson, and Per Carlborg. 2015. Internet-based cognitive–behavior therapy for procrastination: A randomized controlled trial. *Journal of Consulting and Clinical Psychology* 83, 4 (2015), 808.
- [95] James F Sallis, Neville Owen, E Fisher, et al. 2015. Ecological models of health behavior. *Health behavior: Theory, research, and practice* 5, 43–64 (2015).
- [96] Camille L Scent and Susan R Boes. 2014. Acceptance and commitment training: A brief intervention to reduce procrastination among college students. *Journal of College Student Psychotherapy* 28, 2 (2014), 144–156.
- [97] Henri C Schouwenburg. 2004. Perspectives on Counseling the Procrastinator. (2004).
- [98] Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* (2023), 1–6.
- [99] Paschal Sheeran and Thomas L Webb. 2016. The intention–behavior gap. *Social and personality psychology compass* 10, 9 (2016), 503–518.
- [100] Sofia Eleni Spatharioti, David M Rothschild, Daniel G Goldstein, and Jake M Hofman. 2023. Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment. *arXiv preprint arXiv:2307.03744* (2023).
- [101] Piers Steel. 2007. The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological bulletin* 133, 1 (2007), 65.
- [102] Frode Svartdal. 2017. Measuring procrastination: Psychometric properties of the Norwegian versions of the Irrational Procrastination Scale (IPS) and the Pure Procrastination Scale (PPS). *Scandinavian Journal of Educational Research* 61, 1 (2017), 18–30.
- [103] Liyan Tang, Zhaoyi Sun, Betina Idnay, Jordan G Nestor, Ali Soroush, Pierre A Elias, Ziyang Xu, Ying Ding, Greg Durrett, Justin F Rousseau, et al. 2023. Evaluating large language models on medical evidence summarization. *npj Digital Medicine* 6, 1 (2023), 158.

- [104] Anja Thieme, Aditya Nori, Marzyeh Ghassemi, Rishi Bommasani, Tariq Osman Andersen, and Ewa Luger. 2023. Foundation Models in Healthcare: Opportunities, Risks & Strategies Forward. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–4.
- [105] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature Medicine* (2023), 1–11.
- [106] Yao Tian, Chengwei Tong, Lik-Hang Lee, Reza Hadi Mogavi, Yong Liao, and Pengyuan Zhou. 2023. Last Week with ChatGPT: A Weibo Study on Social Perspective regarding ChatGPT for Education and Beyond. *arXiv preprint arXiv:2306.04325* (2023).
- [107] Betül Tokur and Raşit Avcı. 2015. Effect of cognitive-behavioral-theory-based skill training on academic procrastination behaviors of university students. (2015).
- [108] Rahmat Fitrah Tuasikal and Bhina Patria. 2019. Role of social support and self-concept clarity as predictors on thesis writing procrastination. *Journal of Psychology and Instruction* 3, 3 (2019), 76–82.
- [109] Shintaro Uchiyama, Kyoji Umemura, and Yusuke Morita. 2023. Large Language Model-based System to Provide Immediate Feedback to Students in Flipped Classroom Preparation Learning. *arXiv preprint arXiv:2307.11388* (2023).
- [110] Bilge Uzun Ozer, Ayhan Demir, and Joseph R Ferrari. 2013. Reducing academic procrastination through a group treatment program: A pilot study. *Journal of Rational-Emotive & Cognitive-Behavior Therapy* 31 (2013), 127–135.
- [111] Wendelien van Eerde and Katrin B Klingsieck. 2018. Overcoming procrastination? A meta-analysis of intervention studies. *Educational Research Review* 25 (2018), 73–85.
- [112] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [113] Philippe Verduyn, Nino Gugushvili, and Ethan Kross. 2021. The impact of social network sites on mental health: distinguishing active from passive use. *World Psychiatry* 20, 1 (2021), 133.
- [114] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling conversational interaction with mobile ui using large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [115] Jiaqi Wang, Enze Shi, Sigang Yu, Zihao Wu, Chong Ma, Haixing Dai, Qiushi Yang, Yanqing Kang, Jinru Wu, Huawen Hu, et al. 2023. Prompt engineering for healthcare: Methodologies and applications. *arXiv preprint arXiv:2304.14670* (2023).
- [116] Leonard Kuan-Pei Wang, Praneet Sai Paidisetty, and Alicia Magdalena Cano. 2023. The next paradigm shift? ChatGPT, artificial intelligence, and medical education. *Medical Teacher* (2023), 1–1.
- [117] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.
- [118] Michael Williams and Tami Moser. 2019. The art of coding and thematic exploration in qualitative research. *International Management Review* 15, 1 (2019), 45–55.
- [119] Ben Williamson, Felicitas Macgillchrist, and John Potter. 2023. Re-examining AI, automation and datafication in education. , 5 pages.
- [120] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864* (2023).
- [121] Runlong Ye, Pan Chen, Yini Mao, Angela Wang-Lin, Hammad Shaikh, Angela Zavaleta Bernuy, and Joseph Jay Williams. 2022. Behavioral Consequences of Reminder Emails on Students' Academic Performance: a Real-world Deployment. In *Proceedings of the 23rd Annual Conference on Information Technology Education*. 16–22.
- [122] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *27th International Conference on Intelligent User Interfaces*. 841–852.
- [123] Angela Zavaleta Bernuy, Ziwen Han, Hammad Shaikh, Qi Yin Zheng, Lisa-Angelique Lim, Anna Rafferty, Andrew Petersen, and Joseph Jay Williams. 2022. How can Email Interventions Increase Students' Completion of Online Homework? A Case Study Using A/B Comparisons. In *LAK22: 12th International Learning Analytics and Knowledge Conference*. 107–118.
- [124] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B Hashimoto. 2023. Benchmarking large language models for news summarization. *arXiv preprint arXiv:2301.13848* (2023).
- [125] Zoe Zhao, Sophie Song, Bridget Duah, Jamie Macbeth, Scott Carter, Monica P Van, Nayeli Suseth Bravo, Matthew Klenk, Kate Sick, and Alexandre LS Filipowicz. 2023. More human than human: LLM-generated narratives outperform human-LLM interleaved narratives. In *Proceedings of the 15th Conference on Creativity and Cognition*. 368–370.
- [126] Qingxiaoyang Zhu, Yi-Chieh Lee, and Hao-Chuan Wang. 2022. Action-a-Bot: Exploring Human-Chatbot Conversations for Actionable Instruction Giving

and Following. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 145–149.

A CODE DEFINITIONS

Our final codebook had 19 codes in total. We provide their definitions below:

- **Need for direct solutions:** Preferences for straightforward, practical advice or solutions that directly address various aspects of academic procrastination.
- **Step-by-step guidelines:** The desire for detailed instructions that are laid out in a sequential order.
- **Structuring recommendations:** Well-organized, systematic advice that helps in planning and organizing tasks. Structuring recommendations could involve breaking down large tasks into smaller, manageable parts.
- **Overreliance:** Concerns about the possibility of becoming too dependent on Large Language Models for solutions, potentially diminishing one's own problem-solving skills.
- **Adaptive breakdown:** The desire for advice that can be customized or broken down according to the individual's specific needs and circumstances.
- **Reliance on deadlines:** Making recommendations around important deadlines.
- **Use of calendars:** Incorporating calendars as tools for organization, planning, and keeping track of deadlines and tasks.
- **Documenting goals:** The practice of writing down goals as a means to keep track of progress.
- **Need for reminders:** The desire for prompts or alerts to help individuals remember and adhere to their planned tasks and deadlines.
- **Diversity in motivation:** Presenting a variety of motivational prompts to users to help them maintain motivation over time.
- **Varied levels of guidance:** The need for different levels of guidance and support, based on the preferences of individuals and their circumstances.
- **Providing multiple layers of features:** Offering a range of features that allow for varied forms of interaction and support, catering to different user preferences and needs.
- **Flexible engagement:** Allowing users to choose their level of interaction and engagement with the tool, accommodating different styles of learning and task management.
- **Need for emotional support:** Expressing a desire for assistance that goes beyond task management to include emotional support.
- **Limitations of LLMs in addressing emotions:** Recognizing that while LLMs can offer task-related support, their ability to provide emotional support is limited.
- **Clear disclaimers:** The importance of having explicit disclaimers about the capabilities and limitations of LLMs, especially regarding emotional support and therapeutic advice.
- **Examples and instructions:** The request for practical examples and clear instructions on how to effectively use LLMs for managing academic procrastination.

- **Collaborative elements:** The interest in features that enable collaboration, either with the LLM or with other users.
- **Feedback to LLMs:** Ways to integrate user feedback with LLM-based tools for improving their effectiveness and relevance.