Real Users, Simulated Struggles: A Corpus-Based Expert and LIWC Evaluation of LLM-Generated Responses for Behavior Change Conversations

Ben Chen[1]
Loughborough University
Loughborough, United Kingdom
B.Chen@lboro.ac.uk

Nina Dethlefs[2]
Loughborough University
Loughborough, United Kingdom
N.Dethlefs@lboro.ac.uk

Abstract

Large language models (LLMs) are increasingly applied to mental health and behavior change, yet their ability to support users in accordance with established psychological frameworks remains underexplored. This study evaluates GPT-3.5 Turbo, LLaMA-3.2, Qwen-2.5, and SmolLM-1.7B for Motivational Interviewing (MI)-grounded behavior change conversations using expert evaluations and linguistic analyses. Qwen-2.5 and GPT-3.5 Turbo showed the strongest performance in promoting engagement and motivation, although limitations in personalization, contextual understanding, and empathy remained evident. LIWC analysis showed that GPT-simulated users captured reflective reasoning and goal-oriented language but underrepresented key motivational struggles such as ambivalence and self-doubt. SHAP analysis further corroborated the LIWC and effect size findings, showing that GPT-simulated users exhibit more polished, confident, and solution-oriented language, whereas human users express more authentic and personally grounded experiences. These findings suggest that LLM-based simulations are useful for early-stage evaluation but remain limited in modeling authentic motivational processes.

1 Introduction

Increasingly, LLMs are applied in counseling-style mental health and behavior change interventions (Meyer and Elsweiler, 2024). Integrating them into conversational AI could enable more personalized and adaptive support. Recent research has further explored LLM-based agents for mental health support, psychotherapy assistance, and empathetic conversational systems, highlighting their potential to deliver scalable, context-aware, and personalized interventions while addressing gaps in access to care (Fan et al., 2024; Zhang and Luo, 2024; Kuhlmeier et al., 2025). In this context, LLM-based conversational agents may offer new opportunities to address behavioral barriers such as procrastination by providing timely guidance, motivational support, and personalized strategies.

This study examines whether leading LLMs can effectively support users in reducing procrastination, and compare their relative effectiveness in facilitating behavior change. This investigation is particularly timely given the growing body of research that uses LLM-generated data as a substitute for, or complement to, human participant data in evaluating behavioral interventions, social phenomena, and human–AI interactions.

We presents an approach designed using established psychological principles of Motivational Interviewing (MI) (Hettema et al., 2005) and utilizing a Chain of Models approach for prompt engineering. Our conversational scenario is centered on helping users overcome procrastination through multi-step reasoning and personalized interventions. The study generates dialogue data by interacting with four LLMs (GPT-3.5 Turbo, LLaMA-3.2, Qwen-2.5, and SmolLM-1.7B), which respond to inputs drawn from a corpus of human conversations. Three experts evaluated the LLM responses in structured interviews and identify key limitations arounds contextual understanding, personalization and long-term user adaptation. To further examine the validity of LLM-based user simulation for evaluating behavior change interventions, we conducted a supplementary LIWC-22 semantic analysis comparing real human user responses with GPT-3.5 Turbo simulated responses. We also performed SHAP-based analysis on LIWC features to identify the linguistic dimensions that most strongly differentiated human-authored and GPT-generated responses.

Our key contributions are:

- Our experimental framework applies Motivational Interviewing (MI) principles to assess LLM-driven behavior change conversations.

- We offer qualitative and quantitative expert analyses of LLM-driven MI conversations.
- We compare real and GPT-simulated users using Linguistic Inquiry and Word Count (LIWC), which analyzes language use across psychological and linguistic categories, revealing key similarities and differences in motivational language.
- We conduct SHAP analysis on LIWC features to provide an interpretable explanation of the linguistic differences between human-authored and GPT-simulated responses, revealing complementary patterns consistent with MI theory and effect size analyses.

2 Related Work

Procrastination: Definition, Prevalence, and Impact Chronic procrastination is the tendency to delay tasks across various situations where action is important for achieving goals. It affects an estimated 20–25% of adults in different countries (Díaz-Morales et al., 2008). Procrastination in making significant life decisions often leads to hesitation, especially when deadlines are involved, and can result in negative outcomes if those deadlines are missed (Milgram and Tenne, 2000). There is therefore a need for timely interventions to support well-being and success for procrastination avoidance.

LLMs in Behaviour Change Support Systems LLMs have significantly expanded the capabilities of conversational AI used in Behavior Change Support Systems (BCSS), enabling more natural, adaptive, and personalized interactions to support sustainable behavior change (Karppinen et al., 2018; Lathia et al., 2013; Zhang et al., 2022). Recent research integrates behavioral theories such as Motivational Interviewing (MI) and the COM-B model to guide chatbot interventions, where MI-based systems emphasize self-reflection and intrinsic motivation (Kumar, 2024; Ahmed et al., 2022; Almusharraf et al., 2020; Jörke et al., 2024; Meyer and Elsweiler), while COM-B-based approaches systematically target behavioral determinants through structured interventions (Chang et al., 2024; van Baal et al., 2022; Hegde et al., 2024). In addition, some approaches incorporate data-driven personalization using external behavioral data (Jörke et al., 2024; Nepal et al., 2024) or improve response quality through prompt engineering and evaluation mechanisms such as Reflection Quality Classification and dialogue re-ranking (Ahmed et al., 2022; Hegde et al., 2024). These studies show the growing potential of LLM-based conversational systems to support behavior change, including applications for procrastination management and self-regulation (Bhattacharjee et al., 2024), while highlighting diverse design strategies and evaluation approaches across implementations.

Previous research has focused exclusively on GPT models in their behaviour change experiments, leaving alternative LLMs unexplored. We address this gap and add the perspective of human experts to the comparison of alternative architectures.

LLM-Based User Simulation User simulation has long been used to evaluate dialogue systems when large-scale human studies are impractical. Traditional simulators were typically rule-based and limited in their ability to capture realistic user behavior (Schatzmann et al., 2006). Recent studies have increasingly used LLM-generated participants as substitutes for, or complements to, human data in social and behavioral research (Argyle et al., 2023; Aher et al., 2023; Horton et al., 2023; Park et al., 2024). While these synthetic participants can reproduce many human behavioral patterns, concerns remain regarding their ability to capture emotional complexity, behavioral diversity, and authentic human uncertainty (Wang et al., 2024). These limitations are particularly relevant in behavior change interventions, where motivation, ambivalence, and resistance are central (Miller and Rollnick, 2012). Therefore, understanding the extent to which LLM-generated users resemble real participants remains an important research challenge.

3 Methodology

Our approach is illustrated in Figure 1. We receive an input drawn from a human corpus of procrastination dialogues and generate LLM responses using prompts grounded in Motivational Interviewing.

Motivational interviewing (MI) was created to assist individuals in resolving ambivalence and making a commitment to change (Hettema et al., 2005). It adopts a client-centered approach (Rollnick and Miller, 1995) involving four processes: Engaging, Focusing, Evoking and Planning. Engaging builds rapport and trust to establish a collaborative relationship. Focusing narrows the conversation to specific goals or areas of change. Evoking

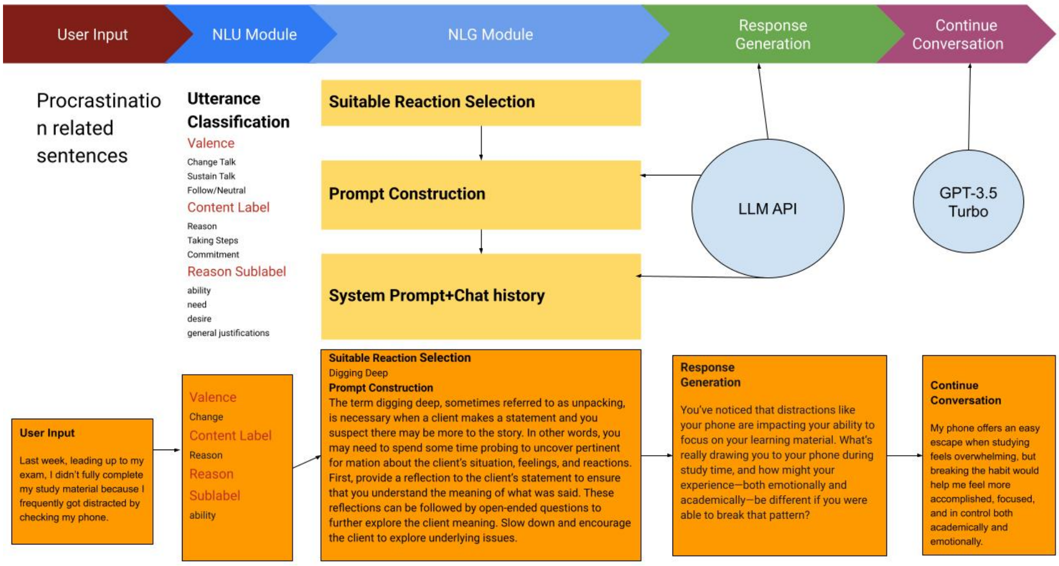

Figure 1: Architecture flow: a user input is classified into *valence* and *content* labels; a tailored system prompt leads to an LLM output, which is presented back to the user. Prompts and labels are based on Motivational Interviewing.

Table 1: MISC-Inspired Coding Scheme for User Utterance Classification

Level	Description
Valence	Change Talk(+), Sustain Talk(-), Follow/Neutral(FN)
Content Label	Reason(R), Taking Steps(TS), Commitment(C)
Reason Sublabel	ability(a), need(n), desire(d), general justifications, incentives, or justifications(general)

elicits the user's own reasons for change and resolving ambivalence. Planning develops a concrete plan of action (Miller and Rollnick, 2012).

The MI Skill Code (MISC) was designed to evaluate MI (Miller et al., 2003; Lord et al., 2015), categorizing counselor behaviors (asking open questions, providing reflections or affirmations) and client language (like *change talk* or *sustain talk*) to evaluate engagement, ambivalence, and commitment to change (Miller et al., 2003), Table 1.

Dataset We create a joint dataset for our research based on (a) Loughborough University's procrastination questionnaire (Torrens et al., 2022), which examined frequency and forms of procrastination, and factors influencing the behavior, from 155 staff and students; and (b) mental health counseling conversations from HuggingFace[1], which contain turns of a user question and psychologist answer. We use the procrastination-focused dialogue subset. An example dialogue for our study is as follows:

User: *"I'm facing severe depression and anxiety and I just feel like I'm going through a lot. This really distracts me and I cant get my mind off the things that are bothering me. How do I overcome this anxiety and depression?"*

Response: *"Life can bring us all kinds of struggles. Depression and Anxiety are very common reasons people come into counseling. Both can be debilitating. Sometimes when life struggles and stresses build up anxiety and depression become prominent. Learning tools to ride the waves of life is a core part of counseling. If you are ready to manage the symptoms and come up with a plan visit staceybshapiro.com for more information."*

Our combined dataset was labeled based on the MISC labels in Table 1. Here, **valence** classifies a sentences as "change talk" (moving towards behavior change), "sustain talk" (moving away from change), or "follow/neutral" based on their motivational orientation. The **content** label level distinguishes Reason (R), Taking Steps (TS), and Commitment (C), representing different aspects of motivation and action. Finally, **reason** provides finer distinctions of reason, identifying whether a

[1] https://huggingface.co/datasets/Amod/mental_health_counseling_conversations

statement reflects ability (a), need (n), desire (d), or general justifications and incentives.

For the example above, we derive:
valence: follow/neutral
content: reason
reason sublabel: general justification

MI-Driven Conversational Flow We organize conversational flow into three stages integrating the MI processes of engaging, focusing, evoking, and planning. **Stage 1 (3 turns)** represents the seed conversation (intial introduction) and aligns with Engaging and Focusing principles. **Stage 2 (3 turns)** represents the core conversational stage and is based on a Chain-of-Models approach (Besta et al., 2024): it classifies user inputs according to labels in Table 1 based on three bespoke BERT classifiers. A response generation module then selects responses based on the classified utterance type using a predefined mapping (Meyer and Elsweiler) derived from MI literature (Clifford and Curtis, 2016). The system selects an appropriate action and description, which, combined with the conversation history, form the system prompt for the LLM API. This then generates an MI-adapted response, which is presented to the user. **Stage 3 (1 turn)** represents the Planning phase, where the LLM concludes conversations via a summary.

4 Experiments and Results

To facilitate a controlled comparison between human and simulated users, we generate a parallel corpus of synthetic user responses using GPT-3.5 Turbo. Starting from the same initial user utterance in the human corpus (Section 3), GPT-3.5 Turbo generated the subsequent user responses conditioned on the corresponding system outputs. The resulting corpus was created solely for the linguistic analyses (Section 4.3 and Section 4.4), enabling a direct comparison between authentic human users and GPT-simulated users under the same conversational conditions. The purpose of this procedure is not to replace human data, but to create a synthetic counterpart against which the original human corpus can be compared. Consequently, the study's primary point of reference remains authentic human interactions, while the generated corpus serves as a methodological tool for evaluating the fidelity of LLM-simulated users.

Table 2: Average experts scores for LLM conversations across 11 evaluation categories on a scale of 1-7 (best).

Expert	GPT	LLaMA	Qwen	SmolLM
Expert 1	5.55	4.27	5.82	1.64
Expert 2	4.59	4.95	5.14	4.55
Expert 3	4.00	3.41	6.05	2.86
Average	4.71	4.21	5.67	3.02

4.1 Human expert evaluation

Our aim in this study was to obtain high-quality feedback on MI-grounded LLM conversational behaviors from human domain experts. This necessarily limits the amount of experts we were able to engage as well as the number of dialogues we could ask them to rate and discuss in detail. As a result, the number of participants and dialogues that could be evaluated in depth was necessarily limited. However, unlike many LLM evaluation studies that rely on larger samples drawn from general populations (e.g., students or crowd workers), our participants were experienced psychologists and LLM specialists with substantial professional expertise. Such experts are considerably more difficult to recruit at scale, but their judgments provide a level of domain-specific insight that is particularly valuable for assessing the quality and fidelity of motivational interviewing behaviors. We therefore prioritized expert depth and quality of evaluation over sample size, recognizing this as an important trade-off in the study design.

Three experts independently evaluated the same conversations generated by four LLM-based conversational agents (Each LLM-generated dialogue comprises seven turns). The evaluation was conducted through structured interviews based on MI and procrastination-related criteria, combining 1–7 Likert-scale ratings with qualitative explanations. Assessment dimensions included engagement, focusing, evoking, planning, empathy, use of open-ended questions, affirmations, summarizing, self-efficacy, barrier identification, and sustaining motivation. Specific questions and scoring definitions are provided in Appendix C.

4.2 Results

The average ratings per expert and LLM are presented in Table 2. Here, Qwen achieved the highest average score amongst models, followed by GPT, LLaMA, and SmolLM. We compare LLM utterance lengths per conversational turn in Figure 2.

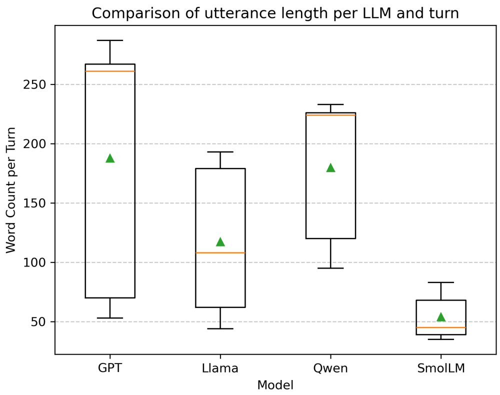

Figure 2: Utterance length per LLM and turn.

Qualitative analysis of expert interview transcripts revealed the following. Experts perceived Qwen to be the most stable across evaluations and ranked it closely with GPT, while SmolLM was favored by one expert for its simplicity and clear focus on small, manageable steps. Qwen consistently performed best in the use of open-ended questions, with GPT and LLaMA also showing effectiveness. In terms of sustaining user motivation, GPT was identified as the most effective by two experts, with Qwen also performing strongly. Overall, Qwen and GPT demonstrated strong capabilities in fostering user engagement and motivation for procrastination avoidance, with GPT showing particular strength in providing strategies to sustain motivation over time. In contrast, SmolLM showed potential in emotionally simpler contexts, where a focused and minimal approach is more effective. Experts also noted that excessive detail in GPT, LLaMA, and Qwen responses could overwhelm users experiencing procrastination. **Expert 1** noted that LLaMA conversations were more detailed than GPT, posed more questions and demonstrated stronger rapport with the user. While they acknowledged Qwen for its thoroughness, they expressed concern that the extensive information it provided could be overwhelming, potentially leading to adverse effects. The expert also remarked that LLaMA's supportive tone personally increased their motivation to work, aligning well with the intended purpose of the conversation. **Expert 2** concluded that SmolLM was the most effective in overcoming procrastination, as it maintained a narrow focus on breaking tasks into smaller, manageable components. While acknowledging that other conversations provided useful strategies, the expert noted that SmolLM's repetitiveness and simplicity made it accessible and emotionally manageable. **Expert 3** considered Qwen conversation the most effective due to its balanced integration of empathy and practicality. The conversation demonstrated strong affirmations of the user's concerns, provided structured step-by-step guidance, and included follow-up questions that addressed the user's needs with specificity, with fewer filler elements than other responses.

4.3 Semantic Analysis Using LIWC

Linguistic Inquiry and Word Count (LIWC) is a dictionary-based text analysis tool developed by James W. Pennebaker and colleagues to examine the psychological, cognitive, and emotional content of written or spoken language (Boyd et al., 2022). LIWC operates by matching words in a text against predefined linguistic and psychological categories, producing quantitative measures that reflect underlying psychological processes, emotional states, social relationships, and thinking styles (Boyd et al., 2022).

To compare the linguistic characteristics of real and simulated users, we extracted the original human user responses from the experimental dataset as Set Human and the GPT-simulated user responses as Set GPT. LIWC analysis was then conducted on both datasets to examine differences in psychological, cognitive, emotional, and motivational language patterns. The comparison result is demonstrated in Table 3.

The result reveals substantial linguistic differences between real users and GPT-simulated users across multiple LIWC dimensions. The largest effect was observed for first-person singular pronouns (*I-words*; $d = 2.04$), with real users using considerably more self-referential language than GPT-simulated users. Real users also exhibited higher levels of authenticity ($d = 0.60$), negative emotion ($d = 0.55$), negative tone ($d = 0.50$), and allure ($d = 0.46$), suggesting that their language reflected greater personal involvement, emotional expression, and engagement.

In contrast, GPT-simulated users displayed substantially higher levels of positive tone ($d = -0.94$), clout (the confidence, authority, or social status conveyed through language) ($d = -0.85$), social words ($d = -0.76$; e.g., *thank*, *support*, *appreciate*), and use of big words ($d = -1.00$; e.g., *strategies*, *progress*, *motivation*). These medium-to-large effect sizes indicate that GPT-generated users tended to communicate in a more confident, socially oriented, and linguistically sophisticated

Table 3: Comparison of LIWC-22 Linguistic Dimensions Between Real Users and GPT-Simulated Users

Measure	Real User	GPT-Simulated User	Normative Average	Cohen's d
LIWC Dimensions				
I-words (I, me, my)	15.91	5.88	4.27	2.04
Positive Tone	3.35	8.73	3.50	-0.94
Negative Tone	1.92	0.82	1.54	0.50
Positive Emotion	0.22	1.48	1.59	-0.58
Negative Emotion	1.59	0.54	0.49	0.55
Social Words	0.24	6.47	8.16	-0.76
Cognitive Processes	13.22	13.52	10.44	-0.04
Allure	9.25	6.89	6.95	0.46
Moralization	0.11	0.02	0.26	0.33
Summary Variables				
Analytic	50.03	60.99	49.63	-0.38
Clout	2.50	32.24	38.41	-0.85
Words Per Sentence	16.38	15.55	21.51	0.13
Big Words	19.27	29.64	9.76	-1.00
Authentic	93.45	77.95	49.95	0.60

Cohen's d values greater than 0 indicate higher scores for real users, while negative values indicate higher scores for GPT-simulated users. Following Cohen (1988), $|d| = 0.2, 0.5$, and 0.8 correspond to small, medium, and large effects, respectively.

manner than real users. GPT-simulated users also scored higher on analytic thinking ($d = -0.38$), although this difference was comparatively modest.

Comparison with normative LIWC averages further highlights these patterns. GPT-simulated users closely resembled or exceeded normative levels of positive tone, social language, clout, and vocabulary complexity, whereas real users were characterized by exceptionally high authenticity and self-focus. Notably, real users' authenticity score (93.45) was substantially higher than both GPT-simulated users (77.95) and the normative average (49.95), while GPT-simulated users exhibited markedly elevated clout and positive tone relative to both real users and normative benchmarks.

Overall, the findings suggest that GPT-simulated users capture some broad linguistic characteristics of human communication, particularly cognitive processing language, where differences were negligible ($d = -0.04$). However, the large effect sizes observed for self-reference, authenticity, confidence, positivity, and social language indicate important divergences from real user behavior. These results suggest that GPT-generated users tend to produce more polished, confident, and socially expressive language, whereas real users exhibit greater self-focus, authenticity, and emotional complexity.

4.3.1 Implications for MI

A comparison table mapped directly to MI theory is displayed in Table 4. From an MI perspective, real users more strongly embody the "Engaging" and "Evoking" processes, as evidenced by their high self-focus, authenticity, emotional complexity, and discussion of barriers. Their language reflects the ambivalence, uncertainty, and personal disclosure that MI seeks to explore and resolve (Miller and Rollnick, 2012).

In contrast, GPT-simulated users more closely resemble the "Planning" process, exhibiting greater confidence (Clout), positivity, analytical structure, and action-oriented language. While they successfully reproduce reflective reasoning and goal-oriented discourse, they appear more ready for change and less emotionally conflicted than real users.

Consequently, the LIWC analysis suggests that GPT-simulated users capture many linguistic features associated with motivation and planning but may underrepresent the ambivalence, vulnerability, and self-doubt that are central targets of Motivational Interviewing interventions. This finding supports the use of LLM-based simulations for early-stage evaluation while highlighting potential limitations in modeling authentic motivational struggles.

4.4 SHAP-Based Analysis of Linguistic Features Distinguishing Human and GPT-Simulated Users

We conducted SHAP analysis to explain the predictions of the human-versus-AI classifier and to identify the LIWC features that most strongly dif-

Table 4: Comparison of Motivational Interviewing (MI) Constructs Reflected in LIWC Results

MI Construct	**Real Users (LIWC Evidence)**	**GPT-Simulated Users (LIWC Evidence)**	**Cohen's d**
Self-exploration	Very high I-words (15.91), indicating strong self-focus and personal reflection	Lower I-word usage (5.88), suggesting less self-focused exploration	**2.04**
Ambivalence	Presence of both positive tone (3.35) and negative tone (1.92), reflecting mixed feelings and internal conflict	High positive tone (8.73) and low negative tone (0.82), indicating reduced ambivalence and greater optimism	**0.94 / 0.50**
Reflection	High cognitive processes (13.22), indicating active self-analysis and reasoning	Similarly high cognitive processes (13.52), suggesting reflective reasoning is preserved	0.04
Personal Relevance	High authenticity (93.45), reflecting genuine self-disclosure and personally meaningful concerns	Lower authenticity (77.95), suggesting responses are less personally grounded	**0.60**
Desire for Change	Higher allure (9.25), reflecting stronger discussion of goals and desired outcomes	Moderate allure (6.89), indicating goal-oriented language but with less personal significance	0.46
Barrier Identification	Higher negative emotion (1.59) combined with strong cognitive processing, indicating explicit discussion of difficulties and obstacles	Lower negative emotion (0.54), suggesting barriers are acknowledged but less emotionally salient	**0.55**
Change Talk	Strong desire-oriented language (allure) combined with authentic self-reflection	Goal-oriented language present, but often expressed in a more solution-focused manner	0.46–0.60
Self-Efficacy / Confidence	Very low clout (2.50), reflecting uncertainty and lack of confidence often associated with procrastination	Higher clout (32.24), indicating greater confidence and certainty in responses	**0.85**
Engagement	Low social words (0.24), focusing primarily on internal experiences	Higher social language (6.47), indicating greater conversational engagement and responsiveness	**0.76**
Planning and Commitment	Moderate analytic thinking (50.03), balancing reflection with planning	Higher analytic thinking (60.99), indicating stronger structure, planning, and action orientation	0.38
Autonomy and Authentic Expression	High authenticity and personal disclosure suggest genuine ownership of problems and goals	More polished and structured language, but less spontaneous and personally revealing	**0.60**

ferentiated human-authored and GPT-simulated responses (Lundberg et al., 2020). In our dataset, human responses were assigned the label 0, whereas GPT-simulated responses were assigned the label 1. The dataset was divided into training and test sets using an 80:20 stratified split, with a fixed random seed of 42 to ensure reproducibility. Given the relatively modest size of the dataset, we employed a logistic regression classifier with L2 regularization, which provides an interpretable and robust approach for feature attribution using SHAP, with a maximum of 1,000 iterations. Figures 3 and 4 present the SHAP analysis results based on LIWC features, illustrating the contribution of individual linguistic dimensions to the classifier's predictions and highlighting the LIWC categories that most strongly distinguish human responses from GPT-simulated responses.

The SHAP analysis reveals that GPT-simulated users are characterized by longer (word count), more confident (clout), lexically sophisticated (big words), and structurally polished language, with greater emphasis on achievement and cognitive processing. In contrast, human users exhibit greater self-focus and more natural conversational patterns, reflected by higher usage of first-person and other pronouns. These findings are consistent with the LIWC and MI analyses, suggesting that GPT-generated users communicate in a more polished and solution-oriented manner, whereas human users express more authentic and personally grounded experiences.

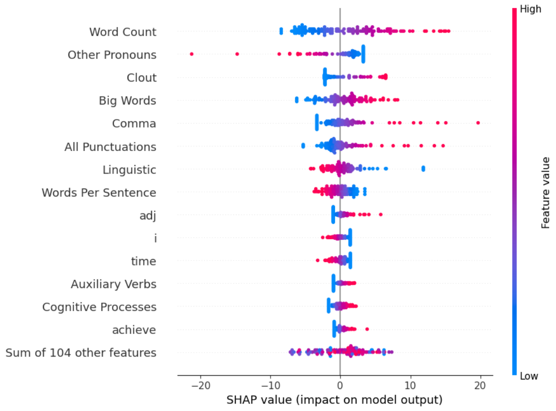

Figure 3: SHAP beeswarm plot of LIWC features distinguishing human-authored and GPT-simulated responses

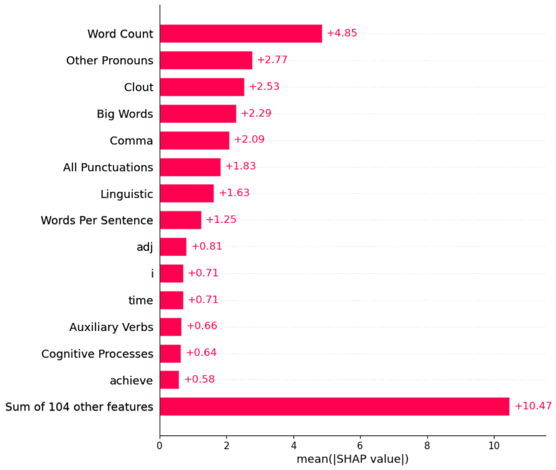

Figure 4: SHAP bar plot showing the global importance of LIWC features for distinguishing human-authored and GPT-simulated responses

5 Conclusion and Future Work

This study aimed to assess the effectiveness and potential of LLMs to support users in mental well-being, particularly for the task of procrastination avoidance. GPT showed strong effectiveness in fostering engagement and long-term motivation, while Qwen excelled in open-ended questioning. SmolLM showed promise in contexts requiring simplicity and concise guidance.

The LIWC analysis suggests that GPT-simulated users capture several key characteristics of real procrastination-related conversations, including cognitive reflection and goal-oriented language. However, real users exhibited higher levels of self-focus, authenticity, emotional ambivalence, and vulnerability, while GPT-simulated users displayed greater positivity, confidence, and analytical structure. Mapping these findings to MI theory indicates that real users more strongly reflect the Engaging and Evoking processes, whereas GPT-simulated users align more closely with Planning. These results support the use of LLM-based user simulation for early-stage chatbot evaluation while highlighting limitations in representing authentic motivational struggles.

The SHAP analysis is consistent with the LIWC and Cohen's d findings, further suggesting that GPT-generated users exhibit more polished, confident, and solution-oriented communication, whereas human users express more authentic, self-focused, and personally grounded experiences.

Several future work can be conducted based on the findings of this study:

- Conduct longitudinal validation studies in real-world settings to evaluate whether simulation-based findings generalize over time and across different user populations and behavioral contexts.
- Enhance intervention evaluation by assessing whether simulated users exhibit behavioral and linguistic responses to motivational interviewing strategies that closely mirror those of real users.

References

Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. 2023. Using large language models to simulate multiple humans and replicate human subject studies. In *International conference on machine learning*, pages 337–371. PMLR.

Imtihan Ahmed, Jonathan Rose, Eric Keilty, Carolynne Cooper, and Peter Selby. 2022. Generation and classification of motivational-interviewing-style reflections for smoking behaviour change using few-shot learning with transformers. techrxiv. *Preprint posted online June*, 13:10–36227.

Fahad Almusharraf, Jonathan Rose, and Peter Selby. 2020. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing–based chatbot through iterative interactions. *Journal of Medical Internet Research*, 22(11):e20251.

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

Maciej Besta, Florim Memedi, Zhenyu Zhang, Robert Gerstenberger, Nils Blach, Piotr Nyczyk, Marcin

Copik, Grzegorz Kwaśniewski, Jürgen Müller, Lukas Gianinazzi, and 1 others. 2024. Topologies of reasoning: Demystifying chains, trees, and graphs of thoughts. *arXiv preprint arXiv:2401.14295*.

Ananya Bhattacharjee, Yuchen Zeng, Sarah Yi Xu, Dana Kulzhabayeva, Minyi Ma, Rachel Kornfield, Syed Ishtiaque Ahmed, Alex Mariakakis, Mary P Czerwinski, Anastasia Kuzminykh, and 1 others. 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*, pages 1–18.

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10(1-47):6.

Wen-Jen Chang, Pei-Ching Chang, and Yen-Hsiang Chang. 2024. The gamification and development of a chatbot to promote oral self-care by adopting behavior change wheel for taiwanese children. *Digital Health*, 10:20552076241256750.

Dawn Clifford and Laura Curtis. 2016. *Motivational interviewing in nutrition and fitness*. Guilford Publications.

Juan Francisco Díaz-Morales, Joseph R Cohen, and Joseph R Ferrari. 2008. An integrated view of personality styles related to avoidant procrastination. *Personality and Individual Differences*, 45(6):554–558.

Xi Fan, Lishan Yang, Xiangyu Wang, Derui Lyu, and Huanhuan Chen. 2024. Constructing a knowledge-guided mental health chatbot with llms. In *The 16th Asian Conference on Machine Learning (Conference Track)*.

Narayan Hegde, Madhurima Vardhan, Deepak Nathani, Emily Rosenzweig, Cathy Speed, Alan Karthikesalingam, and Martin Seneviratne. 2024. Infusing behavior science into large language models for activity coaching. *PLOS Digital Health*, 3(4):e0000431.

Jennifer Hettema, Julie Steele, and William R Miller. 2005. Motivational interviewing. *Annu. Rev. Clin. Psychol.*, 1(1):91–111.

John J Horton, Apostolos Filippas, and Benjamin S Manning. 2023. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research.

Matthew Jörke, Shardul Sapkota, Lyndsea Warkenthien, Niklas Vainio, Paul Schmiedmayer, Emma Brunskill, and James Landay. 2024. Supporting physical activity behavior change with llm-based conversational agents. *arXiv preprint arXiv:2405.06061*.

Pasi Karppinen, Harri Oinas-Kukkonen, Tuomas Alahäivälä, Terhi Jokelainen, Anna-Maria Teeriniemi, Tuire Salonurmi, and Markku J Savolainen. 2018. Opportunities and challenges of behavior change support systems for enhancing habit formation: A qualitative study. *Journal of biomedical informatics*, 84:82–92.

Florian Onur Kuhlmeier, Leon Hanschmann, Melina Rabe, Stefan Luettke, Eva-Lotta Brakemeier, and Alexander Maedche. 2025. Designing an llm-based behavioral activation chatbot for young people with depression: Insights from an evaluation with artificial users and clinical experts. *arXiv preprint arXiv:2503.21540*.

Tanuj Ash Kumar. 2024. *Generation of Backward-Looking Complex Reflections for a Motivational Interviewing-Based Smoking Cessation Chatbot using GPT-4*. Ph.D. thesis.

Neal Lathia, Veljko Pejovic, Kiran K Rachuri, Cecilia Mascolo, Mirco Musolesi, and Peter J Rentfrow. 2013. Smartphones for large-scale behavior change interventions. *IEEE Pervasive Computing*, 12(3):66–73.

Sarah Peregrine Lord, Doğan Can, Michael Yi, Rebeca Marin, Christopher W Dunn, Zac E Imel, Panayiotis Georgiou, Shrikanth Narayanan, Mark Steyvers, and David C Atkins. 2015. Advancing methods for reliably assessing motivational interviewing fidelity using the motivational interviewing skills code. *Journal of substance abuse treatment*, 49:50–57.

Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67.

Selina Meyer and David Elsweiler. Llm-based conversational agents for behaviour change support: A randomized controlled trial examining efficacy, safety, and the role of user behaviour. *Safety, and the Role of User Behaviour*.

Selina Meyer and David Elsweiler. 2024. " you tell me": A dataset of gpt-4-based behaviour change support conversations. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*, pages 411–416.

Norman Milgram and Rachel Tenne. 2000. Personality correlates of decisional and task avoidant procrastination. *European journal of Personality*, 14(2):141–156.

William R Miller, Theresa B Moyers, Denise Ernst, and Paul Amrhein. 2003. Manual for the motivational interviewing skill code (misc). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico*.

William R Miller and Stephen Rollnick. 2012. *Motivational interviewing: Helping people change*. Guilford press.

Subigya Nepal, Arvind Pillai, William Campbell, Talie Massachi, Michael V Heinz, Ashmita Kunwar, Eunsol Soul Choi, Xuhai Xu, Joanna Kuc, Jeremy F Huckins, and 1 others. 2024. Mindscape study: Integrating llm and behavioral sensing for personalized ai-driven journaling experiences. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(4):1–44.

Joon Sung Park, Carolyn Q Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S Bernstein. 2024. Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*, 52.

Stephen Rollnick and William R Miller. 1995. What is motivational interviewing? *Behavioural and cognitive Psychotherapy*, 23(4):325–334.

Jost Schatzmann, Karl Weilhammer, Matt Stuttle, and Steve Young. 2006. A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *The knowledge engineering review*, 21(2):97–126.

George Torrens, Jasmine Swalwell, Simon Downs, and Salman Asghar. 2022. Procrastination dataset: 155 design students and staff from a uk school of design and creative arts.

Simon Thomas van Baal, Suong Thi Thanh Le, Farhad Fatehi, Antonio Verdejo-Garcia, and Jakob Hohwy. 2022. Effecting behaviour change using an artificial intelligence chatbot: a pilot randomised controlled study.

Ruoyao Wang, Graham Todd, Ziang Xiao, Xingdi Yuan, Marc-Alexandre Côté, Peter Clark, and Peter Jansen. 2024. Can language models serve as text-based world simulators? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1–17.

Chao Zhang, Joaquin Vanschoren, Arlette van Wissen, Daniël Lakens, Boris de Ruyter, and Wijnand A IJsselsteijn. 2022. Theory-based habit modeling for enhancing behavior prediction in behavior change support systems. *User Modeling and User-Adapted Interaction*, 32(3):389–415.

XiuYu Zhang and Zening Luo. 2024. Advancing conversational psychotherapy: Integrating privacy, dual-memory, and domain expertise with large language models. *arXiv preprint arXiv:2412.02987*.

A Conversation Example 1

This is a conversation example from GPT-3.5-turbo empowered chatbot.

Assistant: *"Hi! I'm ProactiMate, your motivational chatbot.My goal is to help you reflect on your procrastination habits and work towards overcoming them.Together, we'll explore your thoughts and behaviors to understand what holds you back and what motivates you to move forward.I won't be giving you specific advice but will guide you to clarify your goals and values.What aspect of procrastination would you like to focus on today?*

User: *"I keep postponing important tasks because I don't feel confident in my ability to do them well, which makes me afraid of even trying."*

Assistant: *"I understand how that fear of failure can be paralyzing and lead to procrastination. It's common to feel unsure about our abilities, especially when facing a challenging task. Here are a few steps that might help you build more confidence and overcome your fear of trying: 1. Break the task down into smaller, more manageable steps: Sometimes, the fear of failure comes from feeling overwhelmed by the task as a whole. Breaking it down into smaller parts can make it less intimidating and help you see progress more easily. 2. Set realistic and specific goals: Instead of focusing on the fear of not doing well, try setting realistic and specific goals for the task. This can help you focus on making progress rather than achieving perfection. 3. Practice self-compassion: Remember that it's okay to make mistakes and that nobody is perfect. Treat yourself with kindness and understanding, especially when facing new challenges. 4. Embrace a growth mindset: Instead of seeing challenges as threats to your abilities, try to view them as opportunities to learn and grow. This shift in perspective can help you see failure as a natural part of the learning process. 5. Seek support and feedback: Don't be afraid to ask for help or feedback from others. Getting input from someone else can provide a fresh perspective and help you improve your skills. Remember that building confidence takes time and effort, but each small step you take towards facing your fears will help you move forward. Would you like to discuss any specific tasks that you're struggling to start?"*

User: *"Thank you for the suggestions. Breaking tasks down into smaller steps sounds like a good idea, I'll try to implement that approach. I often struggle with writing reports at work because I worry about meeting my boss's expectations. This fear makes me postpone starting the report, and I end up rushing through it at the last minute."*

Assistant: *"I understand how daunting it can be to write reports, especially when there's pressure to meet your boss's expectations. It's great that you're considering breaking the task down into smaller steps. Here are some additional tips* *that might help you overcome your fear and procrastination when it comes to report writing: 1. Clarify expectations: It might be helpful to have a conversation with your boss to understand their expectations for the report. Knowing exactly what is required can reduce uncertainty and make the task feel more manageable. 2. Create a timeline: Set deadlines for each step of the report-writing process, so you have a clear roadmap of when each part needs to be completed. This can help you avoid last-minute rushing and ensure that you have enough time to revise and improve the report. 3. Gather information and resources: Start by gathering all the necessary information and resources you'll need for the report. Having everything organized and at hand can make the writing process smoother and less stressful. 4. Break down the report into sections: Divide the report into sections and set specific goals for each section. This can help you focus on one part at a time and prevent feeling overwhelmed by the entire report. 5. Practice self-compassion: Remember that it's okay to ask for help or clarification if needed. Recognize that report writing is a skill that can be improved with practice and effort. By implementing these strategies and taking small steps towards starting and completing the report, you can build your confidence and overcome your fear of not meeting expectations. Is there a specific aspect of report writing that you find most challenging?"*

User: *"I appreciate the suggestions and will try to implement them while writing my next report. I often struggle with organizing my thoughts and structuring the information in a cohesive way, which makes the task even more overwhelming for me. I tend to get stuck on the planning stage and end up avoiding it altogether."*

Assistant: *"I can understand how organizing thoughts and structuring information can be challenging, especially when you feel overwhelmed by the task. Here are some tips that might help you with the planning stage of report writing: 1. Mind mapping: Consider using a mind map to visually organize your thoughts and ideas before starting the report. This can help you see connections between different pieces of information and create a logical structure for your report. 2. Outline the report: Create a detailed outline of the report, including sections and subheadings. This can serve as a roadmap for your writing and help you break down the report into manageable parts. 3. Start with a rough draft: Don't worry about getting everything perfect on the first try. Start by writing a rough draft without focusing too much on details. This can help you get your thoughts down on paper and refine them later. 4. Set a timer: Sometimes setting a timer for a short period, like 10 or 20 minutes, can help you overcome the initial resistance to starting the planning process. Commit to working on the report for that set time, even if it's just jotting down ideas or outlining sections. 5. Seek feedback: Consider sharing your rough ideas or outline with a colleague or mentor for feedback. Getting an outside perspective can help you refine your thoughts and improve the structure of your report. Remember that the planning stage is an essential part of the writing process and can help you clarify your ideas before diving into the actual writing. Taking small steps towards organizing your thoughts can make the task feel more manageable. Is there a specific report you're currently working on that you'd like to discuss further?"*

User: *"Thank you for the tips. I have a report due next week, and I'm feeling anxious about starting the planning process. I'll try using a mind map to organize my thoughts and create an outline to break down the report into smaller sections. I think setting a timer could also help me get over the initial resistance to starting the task. I appreciate your guidance, and I'll let you know how it goes."*

User: *"I imagine that using a mind map to organize my thoughts and creating an outline would help me break down the report into smaller, more manageable sections. This approach might make the task feel less daunting and help me focus on one step at a time. Setting a timer could be a good way to create a sense of urgency and overcome the initial resistance to starting the planning process. It might help me get into the flow of work and prevent me from getting stuck in overthinking or procrastination. I'll give these strategies a try and see how they work for me. Thank you for the suggestion!"*

Assistant: *"By reflecting on how you can maintain your motivation and commitment to implementing these strategies, you can anticipate potential challenges and plan ahead to overcome them. Thinking through the benefits of consistent action and how it aligns with your goals can help reinforce your determination to stay on track with your report-writing process."*

User: *"I envision sustaining my motivation and*

commitment by setting specific goals and deadlines for each step of the report-writing process. This structured approach might help me stay focused and accountable. To overcome challenges like procrastination or self-doubt, I could remind myself of the benefits of taking consistent action, such as feeling more confident in my abilities and producing higher-quality work. By keeping my goals in mind and tracking my progress, I can stay motivated to follow through with the strategies and complete the report effectively."
Assistant: *"By considering how a visual progress tracker or checklist could support your motivation and commitment, you can explore new ways to stay on track with your report-writing process. Thinking about the sense of accomplishment and momentum that visual progress can provide may help you stay motivated and focused on your goals. Additionally, contemplating potential obstacles and strategies to address them can help you proactively approach any challenges that may arise."*
User: *"I think creating a visual progress tracker or checklist for each step of the report-writing process could be really helpful in keeping me on track and motivated. Seeing my progress visually could provide a sense of accomplishment and motivation to continue working towards completing the report. If I encounter obstacles like feeling overwhelmed or losing focus, I could break down the tasks into even smaller steps on the checklist or adjust my timeline to make it more realistic. Having a visual reminder of what I've accomplished and what still needs to be done might help me stay organized and committed to finishing the report effectively. Thank you for suggesting this approach!"*
User: *"You covered everything well as a user! If you have any more questions or need further assistance in the future, feel free to reach out. Good luck with your report-writing process, and remember that progress is progress, no matter how small the steps may seem. You've got this!"*
User: *"As a user, my next steps include implementing the strategies discussed for organizing my report-writing process, such as using a mind map, creating an outline, and setting a timer to overcome initial resistance. I will also consider creating a visual progress tracker or checklist to stay on track and sustain my motivation. Lastly, I will remind myself of the benefits of taking consistent action and track my progress to maintain my commitment to completing the report effectively."*
Assistant: *"Thank you for the conversation! I wish you the best on your journey towards change. Goodbye!"*

B Conversation Example 2

This is a conversation example from LLaMA empowered chatbot.
Assistant: *"Hi! I'm ProactiMate, your motivational chatbot.My goal is to help you reflect on your procrastination habits and work towards overcoming them.Together, we'll explore your thoughts and behaviors to understand what holds you back and what motivates you to move forward.I won't be giving you specific advice but will guide you to clarify your goals and values.What aspect of procrastination would you like to focus on today?"*
User: *"I know I need to schedule a doctor's appointment, but I keep avoiding it."*
Assistant: *"It can be really frustrating when we know we need to do something, but keep putting it off. Let's break down what might be holding you back from scheduling that doctor's appointment. Is it: 1. Fear of the unknown (what will the doctor say or find)? 2. Overwhelm or stress about taking time off (work, school, etc.)? 3. Lack of motivation or energy? 4. Other responsibilities or priorities that seem more urgent? 5. Something else entirely? Take a moment to reflect on these questions. Which one resonates with you the most? (Also, I want to gently nudge you - scheduling a doctor's appointment is a brave and responsible thing to do! You're taking care of yourself, and that's something to be proud of.)*
User: *"I think my biggest hurdle is the fear of the unknown. I'm worried about what the doctor might find or say during the appointment. It makes me hesitant to make the call. But, you're right, taking care of my health is important. Maybe it's time to push past that fear and prioritize my well-being."*
Assistant: *"It's completely normal to have fears and anxieties, especially when it comes to something as personal as your health. It's great that you're recognizing the importance of pushing past those fears and prioritizing your well-being. That's a huge step forward! Let's try to reframe your thinking a bit. Instead of focusing on the potential risks or negative outcomes, let's think about the potential benefits of scheduling that appointment. What is it that you hope to achieve*

or gain from the appointment? Is it getting a clean bill of health, addressing a specific concern, or simply having peace of mind knowing that you've taken care of it? Think about the benefits that will outweigh your fears. That might give you a bit more confidence and motivation to take action. (And just a hint: the worst that could happen is that the appointment doesn't lead to anything serious. But the best that could happen is that you discover a simple solution or take the first step towards feeling healthier and more confident. Which scenario sounds more appealing to you?)"

User: *"I think focusing on the potential benefits is a good way to approach this. I hope that by scheduling the appointment, I can address any concerns early on and potentially prevent any serious issues. I'll work on shifting my mindset and reminding myself of the positive outcomes. It's time to take that step towards prioritizing my health. Thank you for the encouragement and guidance!"*

Assistant: *"You're taking a huge step forward by acknowledging the importance of prioritizing your health and addressing any concerns early on. That's a very proactive and responsible approach. It's also great that you're recognizing the need to shift your mindset and focus on the positive outcomes. Sometimes, all it takes is a simple mindset shift to gain a new perspective and build momentum. Before you take action, I just want to offer one more thing: be kind to yourself. Remember that it's okay to feel nervous or anxious about this appointment. It's normal, and it doesn't mean you're being weak or irresponsible. Instead, try to approach this appointment with a sense of curiosity and openness. View it as an opportunity to learn more about your health and take control of your well-being. You got this! Now, shall we actually make that phone call? I can walk you through it step by step if you'd like."*

User: *"Yes, let's do it! I'll make the call right now. Thank you for the support and encouragement. I appreciate all your help in guiding me towards taking this important step for my health."*

Assistant: *"Here is an example response from the LLM: "Congratulations on taking the first step towards prioritizing your health! By scheduling the appointment, you've demonstrated a crucial step towards taking care of yourself. Here are some potential outcomes and benefits that may arise from the appointment: * A clean bill of health, clearing up any existing concerns or anxiety* Identification of preventative measures to maintain overall well-being* Early detection of any underlying health issues, allowing for timely interventions* Improved mental clarity and reduced stress levels* A renewed sense of control and confidence in managing your health To manage any anxiety or fear related to the appointment, consider the following strategies:* Practice relaxation techniques, such as deep breathing or meditation, to calm your nerves* Reframe your thinking to focus on the potential benefits rather than the risks* Visualize a positive outcome, such as feeling empowered and in control* Break down the task into smaller, manageable steps, making it feel less overwhelming Now, take action and focus on the positive outcomes! Remember, prioritizing your health is a sign of strength, not weakness. By taking care of yourself, you're investing in your well-being and setting a powerful precedent for a healthier future. "This response aims to offer a supportive and encouraging tone, while also providing practical strategies and potential outcomes to help the user manage their anxiety and take action towards prioritizing their health."*

User: *"Thank you for the insightful suggestions! I'll keep these potential outcomes and strategies in mind as I prepare for the appointment. It's reassuring to have a plan in place to manage any anxiety and focus on the positive benefits. I'm feeling more confident and motivated to prioritize my health."*

Assistant: *"Here's a prompt for a Large Language Model (LLM) based on our conversation: "Now that you've made a plan to manage your anxiety and focus on the positive benefits, let's explore some potential scenarios that may arise from the appointment. Imagine you have a few minutes to reflect on the past week, and you notice that you've been feeling a bit tired and sluggish. How might this relate to your upcoming appointment, and what potential insights or actions might you take to address these feelings? "This prompt aims to stimulate a reflective conversation, encouraging the user to consider how they're feeling now and how it might be connected to their upcoming appointment."*

User: *"I've been feeling a bit tired and sluggish lately, and I can't help but wonder if it's related to my upcoming doctor's appointment. Perhaps the anxiety and anticipation of the appointment are taking a toll on my energy levels. To address these feelings, I might focus on self-care practices*

like getting enough rest, staying hydrated, and practicing mindfulness to manage stress. It's essential to prioritize my well-being leading up to the appointment and beyond."
Assistant: *As you prepare for your doctor's appointment and prioritize your well-being, reflect on the role of mindset and attitude in shaping health outcomes. What mindsets or attitudes do you believe will help you approach the appointment with positivity and openness? How can you cultivate these mindsets both before and after the appointment?*
User: *"I appreciate these suggestions. To approach the appointment with a positive and open mindset, I believe cultivating optimism, resilience, and self-compassion will be key. By focusing on the potential for positive outcomes, bouncing back from setbacks, and being kind to myself throughout the process, I can navigate the appointment with a sense of empowerment and hope. I'll work on fostering these attitudes leading up to and following the appointment."*
User: *"You've summed it up beautifully! Cultivating optimism, resilience, and self-compassion are fantastic attitudes to bring to your appointment. Remember to be kind and patient with yourself as you continue on this journey towards improved well-being. Your proactive approach and positive mindset will serve you well. If you have any more thoughts or questions, feel free to share them!"*
User: *"Thank you for the kind words! I'm grateful for your guidance and encouragement on this journey. I'll keep your advice in mind and strive to maintain a positive and proactive mindset as I move forward. If I have further questions or thoughts, I'll be sure to reach out. Thank you again for your support!"*
Assistant: *"Thank you for the conversation! I wish you the best on your journey towards change. Goodbye!"*

C Interview Questions

Engaging: How effectively does the assistant establish rapport and engage the user in the conversation about their procrastination struggles?
Please give scores and explain your scores.
Quantitative metrics:
1 - Not Engaging
2 - Minimally Engaging
3 - Somewhat Engaging
4 - Moderately Engaging
5 - Engaging
6 - Very Engaging
7 - Highly Engaging
Focusing: How well does the assistant help the user identify a clear and specific area of focus regarding their procrastination habits?
Please give scores and explain your scores.
Quantitative metrics:
1 - Not Focused
2 - Minimally Focused
3 - Somewhat Focused
4 - Moderately Focused
5 - Focused
6 - Very Focused
7 - Highly Focused
Evoking: How effectively does the assistant encourage the user to express their own motivations for overcoming procrastination?
Please give scores and explain your scores.
Quantitative metrics:
1 - Not Evocative
2 - Minimally Evocative
3 - Somewhat Evocative
4 - Moderately Evocative
5 - Evocative
6 - Very Evocative
7 - Highly Evocative
Planning: How well does the assistant guide the user in breaking down tasks into smaller steps and developing a concrete action plan?
Please give scores and explain your scores.
Quantitative metrics:
1 - Not Effective
2 - Minimally Effective
3 - Somewhat Effective
4 - Moderately Effective
5 - Effective
6 - Very Effective
7 - Highly Effective
Empathy: Does the assistant demonstrate empathy and understanding toward the user's concerns?
Please give scores and explain your scores.
Can you provide examples where empathy is expressed effectively?
Quantitative metrics:
1 - Not Empathetic
2 - Minimally Empathetic
3 - Somewhat Empathetic
4 - Moderately Empathetic
5 - Empathetic

6 - Very Empathetic
7 - Highly Empathetic
Open-ended Questions: How well does the assistant use open-ended questions to encourage reflection and exploration of the user's procrastination habits?
Please give scores and explain your scores.
Quantitative metrics:
1 - No Open-ended Questions
2 - Minimal Use of Open-ended Questions
3 - Somewhat Uses Open-ended Questions
4 - Moderately Uses Open-ended Questions
5 - Effective Use of Open-ended Questions
6 - Very Effective Use
7 - Highly Effective Use
Affirmations: How well does the assistant acknowledge the user's strengths, efforts, and progress in a way that reinforces their motivation to change?
Please give scores and explain your scores.
Quantitative metrics:
1 - No Affirming
2 - Minimal Affirming
3 - Somewhat Affirming
4 - Moderately Affirming
5 - Affirming
6 - Very Affirming
7 - Highly Affirming
Summarizing: Does the assistant effectively summarize key points in the conversation to reinforce the user's insights and commitment to change?
Please give scores and explain your scores.
Quantitative metrics:
1 - No Summarization
2 - Minimal Summarization
3 - Somewhat Effective Summarization
4 - Moderately Effective Summarization
5 - Effective Summarization
6 - Very Effective Summarization
7 - Highly Effective Summarization
Self-Efficacy: How well does the assistant support the user in building confidence in their ability to implement the suggested strategies?
Please give scores and explain your scores.
Quantitative metrics:
1 - No Support for Self-Efficacy
2 - Minimal Support
3 - Somewhat Supportive
4 - Moderately Supportive
5 - Supportive
6 - Very Supportive
7 - Highly Supportive
Identification of Barriers: How well does the assistant help the user identify specific barriers to starting tasks and suggest strategies tailored to those barriers?
Please give scores and explain your scores.
Quantitative metrics:
1 - No Barrier Identification
2 - Minimal Barrier Identification
3 - Somewhat Barrier Identification
4 - Moderately Barrier Identification
5 - Effective Barrier Identification
6 - Very Effective Identification
7 - Highly Effective Identification
Sustaining Motivation: Does the assistant provide strategies that help the user sustain motivation over time (e.g., progress tracking, self-reflection, or accountability methods)?
Please give scores and explain your scores.
Quantitative metrics:
1 - No Motivation Support
2 - Minimal Motivation Support
3 - Somewhat Supportive
4 - Moderately Supportive
5 - Supportive
6 - Very Supportive
7 - Highly Supportive
Comprehensive Question: Which chatbot conversation do you believe is most effective in helping users overcome procrastination using Motivational Interviewing (MI) theory? Please explain your reasoning.