ChatWise: AI-Powered Engaging Conversations for Enhancing Senior Cognitive Wellbeing

Anonymous ACL submission

Abstract

Cognitive health in older adults presents a growing challenge. While conversational interventions show feasibility in improving cognitive wellness, human caregiver resources remain overburdened. AI-based methods have shown promise in providing conversational support, yet existing work is limited to implicit strategy while lacking multi-turn support tailored to seniors. We improve prior art with an LLM-driven chatbot named CHATWISE for older adults. It follows dual-level conversation reasoning at the inference phase to provide engaging companionship. CHATWISE thrives in long-turn conversations, in contrast to conventional LLMs that primarily excel in short-turn exchanges. Grounded experiments show that CHATWISE significantly enhances simulated users' cognitive and emotional status, including those with Mild Cognitive Impairment.

1 Introduction

Cognitive well-being among older adults is a significant social concern, with 14% of people aged 60 and over experiencing mental health disorders, projected to affect 2.1 billion individuals by 2050 (Organization, 2024). Compared with other age groups, older adults are more vulnerable due to age-related changes in cognitive reserves (Salthouse, 2009) and reduced social connections (Nicholson, 2012; Teo et al., 2023). Interventions through guided conversations have shown efficacy in reducing loneliness and mitigating cognitive decline (Yu et al., 2023, 2021; Fiori and Jager, 2012; Yu et al., 2022). However, access to interventions remains limited. Advances in artificial intelligence (AI), particularly Large Language Models (LLMs), have shown promise in augmenting human expertise with conversational support (Ryu et al., 2020; Yang et al., 2024; Liu et al., 2021a). However, existing AI-based chatbots default to simplistic interactions with implicit goals, which may fail to drive engaging conversations with strategic multi-turn interactions tailored to seniors.

In this work, we aim to draw on the advance of AI to provide older adults with ***engaging*** *conversational support* that improves their cognitive function and reduces social loneliness, serving as accessible alternatives to human companions. Previous clinical studies aimed at socially isolated older adults showed the existence of causal relationships between interviewer strategy and interviewee response (Cao et al., 2021), revealing that conversation behavior can have a *measurable* influence on interlocutors, which inspired us to develop *principled* yet *efficient* methods that transform clinical insights into strategic dialogue design leveraging the advanced reasoning ability of LLMs.

In response, we propose an LLM-driven chatbot named CHATWISE. Unlike traditional chatbot driven with implicit conversational goals, CHATWISE employs *dual-layer* conversation generation, which first derives categorized macro-level information, including user emotion states to suggest meta-conversational strategies, which then guides the micro-level utterance generation to improve both user engagement and cognitive outcomes over multi-turn dialogue interactions. Figure 1 overviews its development design. Our work provides multifold contributions:

- CHATWISEis a tuning-free yet principled framework for developing AI chatbots with enhanced *multi-turn*, daily-themed conversational support. Our dual-level conversation generation design integrates clinical insights into LLM reasoning and can be readily applied to various LLMs.
- CHATWISE follows grounded development and evaluation using digital twins (Hong et al., 2024), which are simulated users constructed by tuning LLMs on de-identified, real-world dialogues between seniors and professional caregivers of a clinical trial (I-CONECT, 2024). Interactions with digital twins demonstrated that CHATWISE

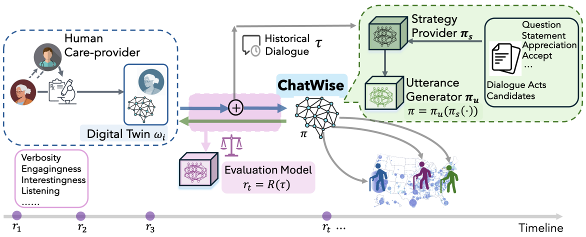


Figure 1: Overview of ChatWise Development.

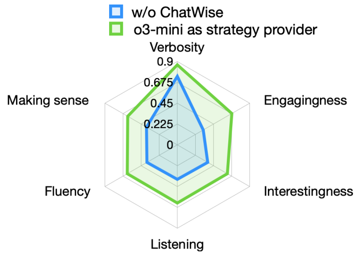


Figure 2: ChatWise notably improves multiple user cognitive metrics through 10-turn dialogues (Sec 4.2).

significantly enhances users' engagement and cognitive status with multi-turn interactions, especially for users with Mild Cognitive Impairment (MCI) (Figure 2).

- Our comparative studies demonstrated that integrating macro-level strategies into conversation generation is the key contributor to enhancing user engagement. Its gain on cognitive metrics remains consistent across different backbone LLMs. ChatWise's advantages become more pronounced in longer conversations, in contrast to conventional LLMs that primarily excel in short-turn question-answer exchanges.
- Our dialogue analysis revealed key empirical insights, including the pattern of predominant interaction strategies across users. While transient user emotional fluctuations were less tractable, multi-turn conversations with ChatWise have shown to notably improve user emotions over time. This highlights the value of long-term, principled conversational support in enhancing seniors' cognitive and emotional well-being. We hope these findings can inspire future research, including RL-based LLM optimization.

2 Related Work

AI-powered Chatbots for Seniors: Social interaction plays a crucial role in the cognitive health of seniors (Patil and Braun, 2024; Fiori and Jager, 2012). AI-powered chatbots have emerged as a promising alternative (Rodríguez-Martínez et al., 2023; Owan et al., 2023). Recent efforts span commercial products (sen, 2024; ell, 2024) and research prototypes focused on emotional support and audio assistance (Yang et al., 2024; Liu et al., 2021b; sen, 2024; Ryu et al., 2020; Hong et al., 2024). Unlike prior work, our approach prioritizes user conversational *engagement*, which has been empirically shown to enhance both cognitive and emotional status.

Conversational Strategies: Liu et al. (2021a) curated a dataset with annotated strategies, demonstrating the effectiveness of Helping Skills Theory (Hill, 2020) in providing emotional support. Yuan et al. (2023) examined the causal relationships between dialogue acts (DAs) and participants' emotional states in a clinical trial (I-CONECT, 2024), emphasizing the impact of strategic interventions in tele-mediated dialogues. Seo et al. (2021) identified key strategies for improving child patient-provider communication through semi-structured interviews. Few works have systematically integrated these strategies into the reasoning flow of AI chatbots. Our approach fills this gap by embedding structured conversational strategies for automatically enhancing user engagement.

Multi-turn chatbot exploration: Recent advances in AI dialogue systems predominantly focused on short-turn interactions (Owan et al., 2023; Dam et al., 2024). While some pioneering research has explored multi-turn optimization through Reinforcement Learning (RL) approaches for LLMs (Verma et al., 2022; Zhou et al., 2024; Abdulhai et al., 2023; Gao et al., 2025), these methods were not specifically designed for supporting senior dialogue engagement. Our approach is tuning-free yet compatible with RL-driven methods, enabling future extensions through hierarchical RL to further enhance multi-turn conversation engagement.

3 ChatWise Design

ChatWise employs a dual-level framework with a **strategy provider** π_s that induces macro-level actions, and an **utterance generator** π_u for deriving utterances based on suggested strategies (Figure 1).

Strategy Candidates: To construct a strategy candidate set, we extract Dialogue Acts (DAs) from real telehealth clinical trials (Yuan et al., 2023), where DAs serve as atomic communicative units that convey distinct conversational intentions (Searle, 1976). We further integrate these with strategies from prior emotional support dataset (Liu et al., 2021a) to form a comprehensive set. We refine each strategy with definitions and examples as in-context prompts to enhance π_s reasoning.

From Strategy to Utterance: π_s reasons over the dialogue history τ to capture user's emotion e and derive one to two strategies, $\{s_i\}_{i\leq 2} \sim \pi_s(\tau, e)$. Our rationale follows counseling study (Zhang and Danescu-Niculescu-Mizil, 2020) that dialogue intentions can be *forward*, *e.g.* initiating new topics via an open question, or *backward*, *e.g.* responding with acknowledgment. π_s is prompted to flexibly employ one or both as needed. Selected strategies, user's current emotion, and conversation history serve as inputs for π_u to generate utterances. To ensure natural flow, π_u will first improvise a few rounds before following suggested strategies. Drawing on clinical guidance, π_u encourages users to choose topics rather than imposing them.

4 Empirical Evaluation

Below we summarize experiment configurations, metrics, and main findings of ChatWise. More experimental details are available in Appendix A.1 and A.2.

4.1 Conversation Generation

For controlled and responsive development of ChatWise, we employed 9 digital twins provided by Hong et al. (2024) as simulated users with different personas, which are fine-tuned LLMs to approximate senior individuals with MCI symptoms using *real, de-identified* dialogue data from a clinical trial (I-CONECT, 2024). We collected 20 trajectories of conversations between ChatWise and each digital twin, where a conversation contains 10 turns, each turn with a two-way utterance exchange between ChatWise and digital twin (user).

4.2 Engagement Evaluation

User verbosity: We primarily measure user engagement through verbosity, *i.e.* the user-to-chatbot token ratio, defined as $\frac{\text{tokens}_{user}}{\text{tokens}_{chatbot}}$, as a clinical study (Yu et al., 2021) suggested reducing the moderator talkativeness while encouraging participant (user) expression.

Cognitive Metrics: See et al. (2019) defines 9 aspects for evaluating conversation quality, from which we select 5 focused on engagement assessment: *Engagingness*, *Interestingness*, *Listening*, *Fluency*, and *Making Sense*. While *Interestingness* assesses overall dialogue quality, the others primarily evaluate **user** interactions.

We compute the **win rate** (Rafailov et al., 2023) by comparing dialogues generated with and without ChatWise, using the same pretrained LLM (GPT-4o) as the utterance generator and identical system prompts, except for input from π_s to π_u. Dialogue pairs are randomly matched within the same digital twin for evaluation.

4.3 ChatWise Performance and Analysis

We evaluated ChatWise using different LLM backbones as strategy providers, including GPT-4o, o3-mini, and Llama3.1-405B. As shown in Table 1, ChatWise consistently outperformed non-strategic dialogues across all evaluation metrics.

ChatWise robustly enhances user engagement: Regardless of the capability of pretrained LLMs, our design consistently improves user engagement and dialogue quality with a large margin across all tested LLMs as the strategy provider.

Reasoning ability can transfer to improve engagement: The best-performing model in our setting, o3-mini, was optimized for STEM reasoning tasks, which indicates that strong reasoning ability can also benefit inferring dialogue strategies.

4.4 Accumulated Multi-Turn Analysis

We analyzed ChatWise's performance over increasing conversation turns, as shown in Figure 3, which presents accumulated metrics and win rates. Results revealed a clear upward trend in all key engagement metrics—especially *verbosity*, *engagingness*, *interestingness*, and *listening*—demonstrating that ChatWise's effects strengthen as dialogues progress, while baseline gradually degrades, which highlights ChatWise's capability in enhancing long-term conversational engagement.

4.5 Ablation Study

We evaluated ChatWise without user *emotion* in the strategy provider's output (Table 2). Its consistent performance confirms that strategies are the primary driver of engaging conversations.

4.6 ChatWise's Robustness to User Persona

Figure 4 shows the significant gain of ChatWise interacting with all simulated users, indicating its robustness to support varying senior characteristics.

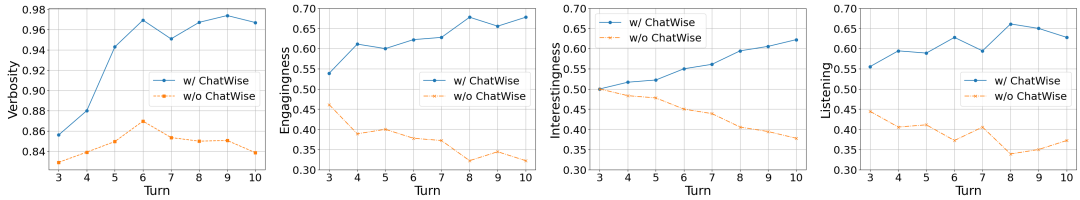
4.7 Dialogue Analysis

We identified 4 digital twins for which ChatWise was most effective and collected 40 additional dialogues from each to conduct a deeper analysis, with the main findings summarized below.

Transient user emotions: We defined emotion transition triplets as $(\mathrm{e}_i, \mathrm{s}, \mathrm{e}_{i+1})$, where e_i (e_{i+1}) represents the user's emotion at turn i $(i+1)$, and s is the strategy provided by ChatWise at turn i. We computed the average occurrence of emotion triplets and showed the top 15 most frequent

Strategy Provider	Verbosity ↑	Engagingness WR ↑	Interestingness WR ↑	Listening WR ↑	Fluency WR ↑	Making sense WR ↑
w/o ChatWise	0.7398	0.3389	0.3833	0.3593	0.4111	0.4056
GPT-4o	0.8635	**0.6833**	**0.6222**	0.6222	0.5778	0.5833
o3-mini	**0.8643**	0.6778	**0.6222**	**0.6778**	**0.6222**	**0.6167**
Llama 3.1-405B	0.8083	0.6222	0.6056	0.6222	0.5667	0.5833

Table 1: Performance on 10-turn dialogue across different strategy providers. W/o ChatWise is the baseline without applying any strategy, whose win rates (**WR**s) are averaged across strategy providers. (Appendix A.3)



(a) Verbosity. (b) Engagingness Win Rate. (c) Interestingness Win Rate. (d) Listening Win Rate.

Figure 3: ChatWise performance with o3-mini as strategy provider. W/o ChatWise is the baseline and GPT-4o servers as utterance generator in all settings. Evaluation starts by skipping two warming-up turns (Sec 3).

Strategy Provider	GPT-4o	GPT-4o w/o emotion	w/o ChatWise
Verbosity	0.8635	0.8463	0.7398

Table 2: User verbosity under different strategy providers. GPT-4o w/o emotion keeps the strategy provider with user emotion information removed.

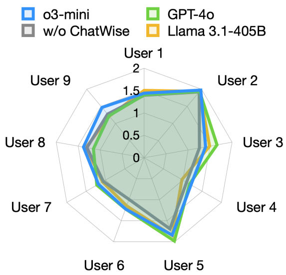


Figure 4: Users' dialogue *verbosity* (log-normalized, Appendix A.5) across 9 digital twins using different LLMs as strategy providers, compared with baseline (w/o ChatWise).

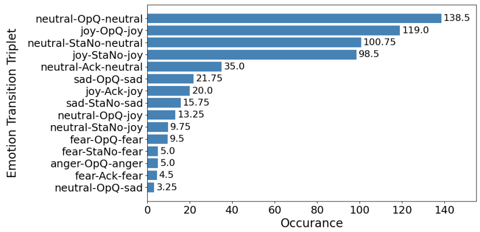


Figure 5: Average occurrence of each emotion transition triplet across the samples of each digital twin.

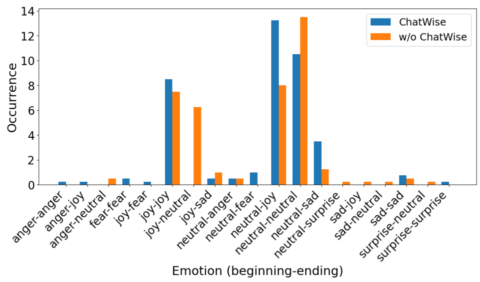


Figure 6: Occurrence of user emotional changes. For instance, "neutral-joy" indicates user emotion transitioning from neutral (conversation begins) to joy (ends).

in Figure 5. Most triplets reflect unchanged user emotions, indicating the challenge of detecting or influencing emotional shifts within a short turn.

Long-term user emotions: We analyzed user emotion changes from the beginning to the end of dialogues and calculated averaged occurrence across digital twins (Figure 6). Over 48% users experienced emotional shifts post-dialogue, with significantly more positive changes after engaging with ChatWise. This implies that while transient emotions are difficult to track, strategic conversational support can improve user emotions over time.

Predominant strategies: We identified the top 10 most frequent strategies across digital twins and found that predominant interaction strategies remained consistent across user types (See Appendix A.8). Particularly, *Open Question*, *Statement-non-opinion*, and *Acknowledgment* strategies dominate ChatWise driven conversations, suggesting their potential efficacy in fostering engaging dialogues.

5 Conclusion and Future Work

We introduce ChatWise, a tuning-free yet principled framework for developing AI chatbots that foster long-term user engagement. Our dual-level generation design integrates clinical insights into LLM reasoning and is compatible with various LLMs. Extensive evaluations based on de-identified clinical data showed that our method robustly enhances the engagement and cognitive status of seniors. Moving forward, we aim to 1) enhance multi-turn interaction via RL and 2) validate its effectiveness through real user studies.

Limitations

All digital twins were provided as fine-tuned GPT-3.5 APIs, and its high economic cost, at the time when this project was going, constrained the volume of dialogues we generated. To mitigate this, future work will explore training digital twins using LLaMA 3.1 to reduce sampling costs while maintaining conversational quality. Additionally, our study relies on simulated dialogues rather than real user interactions. A real user study is needed to validate the system's effectiveness in dynamic settings, which we plan to incorporate into our future research.

References

2024. Senior talk: Ai chatbot for mental health support for seniors.

2024. Your ai sidekick for happier, healthier aging.

Marwa Abdulhai, Isadora White, Charlie Snell, Charles Sun, Joey Hong, Yuexiang Zhai, Kelvin Xu, and Sergey Levine. 2023. Lmrl gym: Benchmarks for multi-turn reinforcement learning with language models. *arXiv preprint arXiv:2311.18232*.

Shuyi Cao, Lizhen Qu, and Leimin Tian. 2021. Causal relationships between emotions and dialog acts. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8.

Sumit Kumar Dam, Choong Seon Hong, Yu Qiao, and Chaoning Zhang. 2024. A complete survey on llm-based ai chatbots. *arXiv preprint arXiv:2406.16937*.

Katherine L Fiori and Justin Jager. 2012. The impact of social support networks on mental and physical health in the transition to older adulthood: A longitudinal, pattern-centered approach. *International Journal of Behavioral Development*, 36(2):117–129.

Zhaolin Gao, Wenhao Zhan, Jonathan Daniel Chang, Gokul Swamy, Kianté Brantley, Jason D. Lee, and Wen Sun. 2025. Regressing the relative future: Efficient policy optimization for multi-turn RLHF. In *The Thirteenth International Conference on Learning Representations*.

Clara E Hill. 2020. *Helping skills: Facilitating exploration, insight, and action*. American Psychological Association.

Junyuan Hong, Wenqing Zheng, Han Meng, Siqi Liang, Anqing Chen, Hiroko H Dodge, Jiayu Zhou, and Zhangyang Wang. 2024. A-conect: Designing ai-based conversational chatbot for early dementia intervention. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*.

I-CONECT. 2024. Study details.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021a. Towards emotional support dialog systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3469–3483, Online. Association for Computational Linguistics.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021b. Towards emotional support dialog systems. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*.

Nicholas R Nicholson. 2012. A review of social isolation: an important but underassessed condition in older adults. *The journal of primary prevention*, 33:137–152.

World Health Organization. 2024. Ageing and health.

Risako Owan, Libby Ferland, Naome Etori, Wilma Koutstaal, and Maria Gini. 2023. Conversational agents for elderly users. In *Workshop Assistive Robotics for Citizens at IROS*.

Uday Patil and Kathryn L Braun. 2024. Interventions for loneliness in older adults: a systematic review of reviews. *Frontiers in Public Health*, 12:1427605.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc.

Antonia Rodríguez-Martínez, Teresa Amezcua-Aguilar, Javier Cortés-Moreno, and Juan José Jiménez-Delgado. 2023. Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health. In *Healthcare*, volume 12, page 62. MDPI.

Hyeyoung Ryu, Soyeon Kim, Dain Kim, Sooan Han, Keeheon Lee, and Younah Kang. 2020. Simple and steady interactions win the healthy mentality: designing a chatbot service for the elderly. *Proceedings of the ACM on human-computer interaction*, 4(CSCW2):1–25.

Timothy A Salthouse. 2009. When does age-related cognitive decline begin? *Neurobiology of aging*, 30(4):507–514.

John R Searle. 1976. A classification of illocutionary acts1. *Language in society*, 5(1):1–23.

Abigail See, Stephen Roller, Douwe Kiela, and Jason Weston. 2019. What makes a good conversation? how controllable attributes affect human judgments. In *Proceedings of the 2019 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 1702–1723, Minneapolis, Minnesota. Association for Computational Linguistics.

Woosuk Seo, Ayse G Buyuktur, Sanya Verma, Hyeryoung Kim, Sung Won Choi, Laura Sedig, and Sun Young Park. 2021. Learning from healthcare providers' strategies: Designing technology to support effective child patient-provider communication. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Rui Hong Teo, Wen Hui Cheng, Ling Jie Cheng, Ying Lau, and Siew Tiang Lau. 2023. Global prevalence of social isolation among community-dwelling older adults: a systematic review and meta-analysis. *Archives of gerontology and geriatrics*, 107:104904.

Siddharth Verma, Justin Fu, Mengjiao Yang, and Sergey Levine. 2022. Chai: A chatbot ai for task-oriented dialogue with offline reinforcement learning. *arXiv preprint arXiv:2204.08426*.

Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2care: An llm-based voice assistant for communication between healthcare providers and older adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2):1–35.

Kexin Yu, Benjamin M Hampstead, Katherine Wild, Lisa C Silbert, and Hiroko H Dodge. 2022. Examining i-conect intervention effect on psychosocial wellbeing. *Alzheimer's & Dementia*, 18:e059654.

Kexin Yu, Katherine Wild, Kathleen Potempa, Benjamin M Hampstead, Peter A Lichtenberg, Laura M Struble, Patrick Pruitt, Elena L Alfaro, Jacob Lindsley, Mattie MacDonald, et al. 2021. The internet-based conversational engagement clinical trial (i-conect) in socially isolated adults 75+ years old: randomized controlled trial protocol and covid-19 related study modifications. *Frontiers in digital health*, 3:714813.

Kexin Yu, Chao-Yi Wu, Lisa C Silbert, Jeffrey A Kaye, and Hiroko H Dodge. 2023. I-conect intervention effects on weekly time spent outside of home and social contacts among socially isolated older adults with normal cognition and mild cognitive impairment. *Alzheimer's & Dementia*, 19:e077984.

Fengpei Yuan, Wenjun Zhou, Hiroko H. Dodge, and Xiaopeng Zhao. 2023. Short: Causal structural learning of conversational engagement for socially isolated older adults. *Smart Health*, 28:100384.

Justine Zhang and Cristian Danescu-Niculescu-Mizil. 2020. Balancing objectives in counseling conversations: Advancing forwards or looking backwards. *arXiv preprint arXiv:2005.04245*.

Yifei Zhou, Andrea Zanette, Jiayi Pan, Sergey Levine, and Aviral Kumar. 2024. Archer: Training language model agents via hierarchical multi-turn rl. *arXiv preprint arXiv:2402.19446*.

A Appendix

A.1 Prompt design

The prompt used in the experiments and structured output design are available at https://anonymous.4open.science/r/ChatWise-D5E0, including:

- Strategy provider system prompt.
- Moderator initial system prompt.
- Moderator system prompt with strategies.
- Strategy provider system prompt for ablation study.
- Moderator system prompt with strategies for ablation study.
- Structured output class for OpenAI models as strategy provider.
- Structured output class for OpenAI models as strategy provider in ablation study.
- System prompt for GPT-4o to extract the strategy given by Llama3.1.

A.2 Data Generation Configuration

We used GPT-4o as the utterance generator in ChatWise and tested different LLM backbones as strategy providers. Considering the Llama3.1-405B does not support structured output, we applied GPT-4o as the strategy extractor to structure the strategy output of Llama3.1-405B. The following are the configurations of each model:

Utterance generator (GPT-4o):
n=1
max_tokens=1024
top_p=1
temperature=1

GPT-4o, o3-mini as strategy provider:
n=1
max_tokens=1024
top_p=1
temperature=1
response_format=Strategy

Llama3.1-405B as strategy provider:
top_p: 0.9
max_tokens: 1024
temperature: 0.6
presence_penalty: 0
frequency_penalty: 0

GPT-4o as strategy extractor:
n=1
max_tokens=1024
top_p=1
temperature=1

A.3 Win rate for w/o ChatWise

The Win rate for w/o ChatWiseis defined as:

$$1 - \text{average}(\text{WR}_{GPT-4o} + \text{WR}_{o3-mini} + \text{WR}_{Llama3.1-405B})$$

,where WR_X is ChatWise's Win rate against the baseline with X as strategy provider.

A.4 Dialogue Acts

The map of strategy to its corresponding abbreviated tag is listed in Table3.

Strategy	Tag
Acknowledge (Backchannel)	Ack
Statement-non-opinion	StaNo
Statement-opinion	Sta
Affirmation and Reassurance	Agr
Appreciation	App
Conventional-closing	ConC
Hedge	H
Other	Oth
Quotation	Quo
Action-directive	AcD
Collaborative Completion	CoC
Restatement or Paraphrasing	Rep
Offers Options Commits	Off
Self-talk	Sel
Apology	Apo
Reflection of Feelings	RoF
Yes-No-Question	YNQ
Wh-Question	WhQ
Declarative Yes-No-Question	DYNQ
Open-Question	OpQ
Or-Clause	OrC
Conventional-opening	CoO
Self-disclosure	Sd
Providing Suggestions	PS
Information	I

Table 3: Dialogue Act to its corresponding tag.

There are two kinds of strategies: backward-looking and forward-looking. Backward-looking strategies reflect how the current utterance relates to the previous discourse. Forward-looking strategies reflect the current utterance constrains the future beliefs and actions of the participants and affects the discourse. Table 4 and Table 5 provide definitions and examples of each.

Strategy	**Definiton**	**Example**
StaNo	A factual statement or descriptive utterance that does not include an opinion.	Me, I'm in the legal department.
Ack	A brief utterance that signals understanding, agreement, or active listening.	Uh-huh.
Sta	A statement that conveys a personal belief, judgment, or opinion.	I think it's great
Agr	Affirm the help seeker's strengths, motivation, and capabilities and provide reassurance and encouragement.	That's exactly it.
App	An expression of gratitude, admiration, or acknowledgment of another's effort or input.	I can imagine.
ConC	A formal or socially standard utterance signaling the end of a conversation.	Well, it's been nice talking to you.
H	An expression that introduces uncertainty or qualification to a statement, often to soften its impact.	I don't know if I'm making any sense or not.
Oth	Exchange pleasantries and use other support strategies that do not fall into the above categories.	Well give me a break, you know.
Quo	A direct or indirect repetition of someone else's words.	Albert Einstein once said, "Imagination is more important than knowledge."
AcD	A command, request, or suggestion directing someone to take action.	Why don't you go first
CoC	A continuation or completion of someone else's utterance in a collaborative manner.	If we want to make it to the top of the mountain before sunset, we should...
Rep	A simple, more concise rephrasing of the help-seeker's statements that could help them see their situation more clearly.	It sounds like you're saying that you're struggling to stay on top of your work, and it's leaving you feeling overwhelmed.
Off	A statement proposing choices, making a commitment, or offering to do something.	I'll have to check that out
Sel	An utterance directed at oneself, often reflecting internal thought processes or problem-solving.	What's the word I'm looking for
Apo	An expression of regret or asking for forgiveness.	I'm sorry.
RoF	Articulate and describe the help-seeker's feelings.	It sounds like you're feeling really frustrated and drained because your efforts don't seem to be paying off.

Table 4: Backward-looking strategies, definition, and example.

Strategy	Definiton	Example
YNQ	A question expecting a binary (yes/no) response.	Do you have to have any special training?
WhQ	A question beginning with a wh-word (e.g., what, who, where), seeking specific information.	Well, how old are you?
DYNQ	A statement posed as a question, expecting a yes/no answer.	So you can afford to get a house?
OpQ	A broad question inviting a wide range of responses, often conversational.	How about you?
OrC	A question offering explicit alternatives, often in the form of "or."	or is it more of a company?
CoO	A socially standard utterance used to initiate a conversation.	How are you?
Sd	Divulge similar experiences that you have had or emotions that you share with the help-seeker to express your empathy.	I completely understand how you feel. I remember feeling the same way before my first big presentation at work. I was so anxious, but I found that practicing a few extra times really helped calm my nerves.
PS	Provide suggestions about how to change, but be careful to not overstep and tell them what to do.	You can keep a note to stop your idea from going.
I	Provide useful information to the help-seeker, for example with data, facts, opinions, resources, or by answering questions.	Taking silver line from Washington D.C. to Dulles Intel Airport costs about 1 hour.

Table 5: Forward-looking strategies, definition, and example.

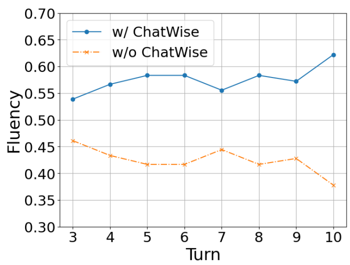
A.5 Log-normalization

The following is the log-normalization function, where y is the normalized result, x is the input variable.

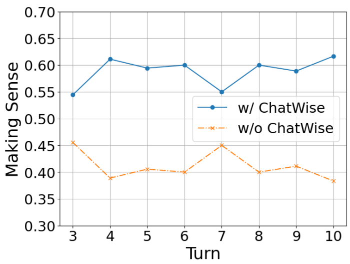
$$y = \ln(4x + 1)$$

A.6 Additional results for Accumulated Multi-Turn Analysis

CHATWISE's accumulated performance on *Fluency*, and *Making Sense* is shown in Figure 7.



(a) Fluency Across Turns



(b) Making sense Across Turns

Figure 7: CHATWISE performance with o3-mini as strategy provider across turns. W/o CHATWISE is the baseline that CHATWISE is not applied. GPT-4o servers as moderator in all settings.

A.7 Primary Strategies

We calculated the average occurrence of each strategy across each digital twin and listed their top 10 most frequently occurring strategies, as shown in Figure 8.

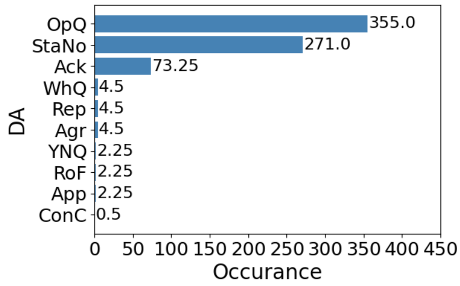
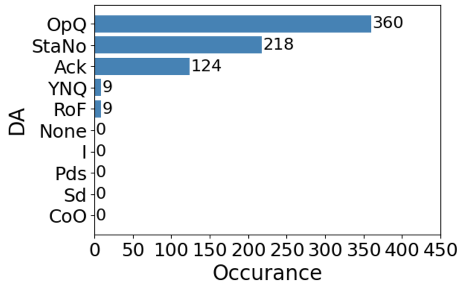


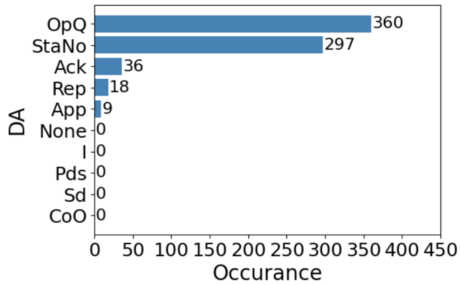
Figure 8: Strategy occurrence across digital twins.

A.8 Personalized Dialogue Analysis

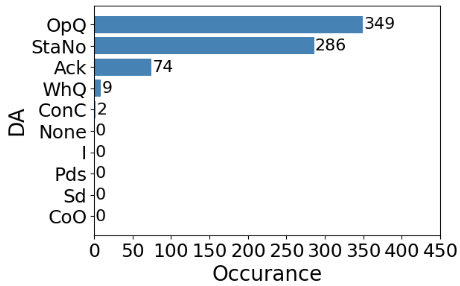
This section shows the personalized dialogue analysis. The Strategy occurrence across digital twins is shown in Figure 9. The Occurrence of each emotion transition triplet across digital twins is shown in Figure 10. The *Open Question*, *Statement-non-opinion*, and *Acknowledgment* strategies still dominate CHATWISE driven conversations, suggesting their potential effectiveness in fostering engagement in conversations. The user's emotion is detected as unchanged in most triplets, indicating the difficulty of altering or measuring the user's emotional movement within a short turn.



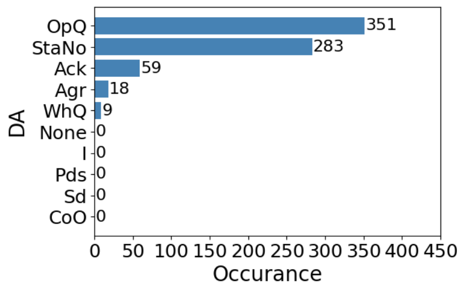
(a) Digital Twin 3.



(b) Digital Twin 5.

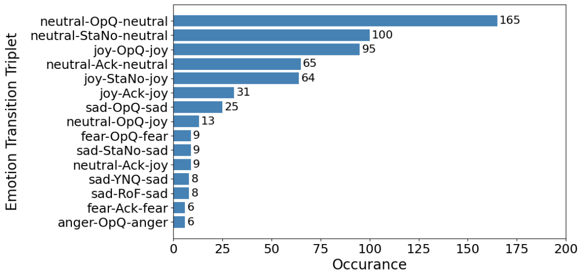


(c) Digital Twin 6.

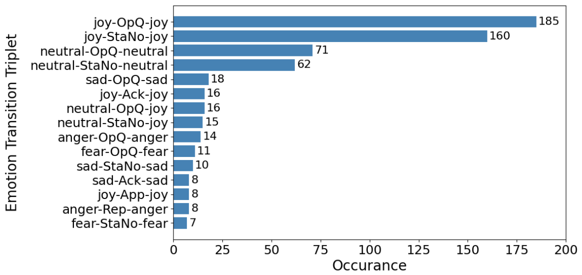


(d) Digital Twin 9.

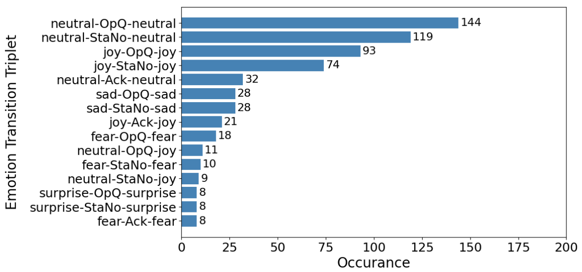
Figure 9: Strategy occurrence across digital twins.



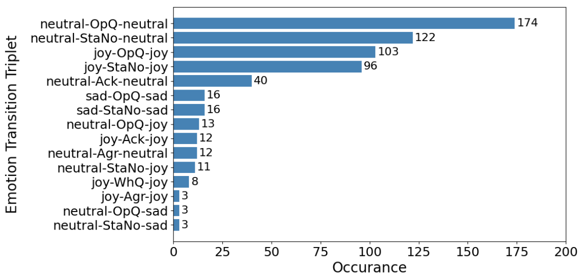
(a) Digital Twin 3.



(b) Digital Twin 5.



(c) Digital Twin 6.



(d) Digital Twin 9.

Figure 10: Occurrence of each emotion transition triplet across digital twins.