

Research Statement

Lizi Liao

School of Computing and Information Systems, Singapore Management University

Tel : (65) 6828-4882 ; Email : lzliao@smu.edu.sg

18 (Day) 12 (Month) 2025 (Year)

Background

My research revolves around developing proactive conversational systems that naturally and intelligently interact with humans to satisfy a range of goals, including information-seeking, recommendation, and emotionally supportive dialogue. My overarching vision is to create systems that can:

- Understand multifaceted user needs and contexts (textual, multimodal) through active interaction.
- Plan and anticipate possible user requirements, thus proactively offering help or insights.
- Respond effectively, blending knowledge from various sources (e.g., structured databases, commonsense, or user profiles).

In the early stage of my career, my research focused on multimodal conversational search and recommendation (MCSR) [15, 14], addressing the problem of information asymmetry between users and systems. Over time, I broadened my scope to proactive conversational scenarios, where the agent anticipates user needs and reasons on-the-fly [4, 3, 9], and to empathetic or emotional-support scenarios, where the agent processes not just factual context but also affective signals [6, 23].

This statement summarizes my central research concerns and strategy (§2), demonstrates how my work has evolved (§3), highlights my key accomplishments and recognitions (§4), and projects my future research trajectory (§5).

2 Central Concerns and Strategy

2.1 Central Research Questions

- **How to proactively assist and engage users?** Moving beyond purely reactive conversational models, I aim to design systems that can plan their interactions with users (e.g., question planning, policy learning), thereby preempting user requests and providing relevant prompts or suggestions [3, 7].
- **How to effectively integrate multimodal signals and ontology expansions?** Whether the user input is textual, visual, or spoken, the system must robustly interpret

and leverage these cues—and continuously expand or refine its domain knowledge [23, 13].

- **How to embed empathy in conversational systems?** Beyond factual correctness, the system should adapt its strategy and tone to users’ emotional states, offering emotional support when needed [6, 23].

- **How to evaluate complex, evolving dialogues?** As conversation becomes more dynamic, standard static metrics alone are insufficient. I develop user simulation [19] and reflection-based measures [7] to gauge real-world performance.

2.2 Strategy for Choosing Research Problems and Approaches

1. **Human-centric Design:** I prioritize tasks that address genuine user needs such as recommendation, information retrieval, or affective support. This ensures that solutions remain practical and beneficial.

2. **Proactive Modeling:** My core technical contribution lies in formulating proactive approaches, including simulation-free hierarchical latent policy [9], question planning with reflection [7], and multi-intent generative frameworks [22].

3. **Knowledge-enriched Neural Methods:** I integrate domain knowledge, ontology expansions, or large language models (LLMs) to tackle under-specified or evolving user demands [13, 12].

4. **Scalability and Adaptability:** My experiments typically span multiple domains or modalities, ensuring the techniques can generalize or transfer to new settings (for instance, from e-commerce to mental health support).

3 Evolution of My Research

Table 1 visualizes how my research has evolved since 2018, reflecting both technological advances and shifting real-world needs for conversational AI.

Table 1. High-Level Evolution of My Research and Representative Publications.

Phase	Focus and Key Publications
2018–2019	<ul style="list-style-type: none"> • Knowledge-aware Multimodal Dialogue: My early work examined integrating textual and visual signals with external knowledge to enhance user-agent interaction. • Best Paper Finalist at ACM Multimedia 2018: [15] on “Knowledge-aware Multimodal Dialogue Systems.”
2020–2021	<ul style="list-style-type: none"> • Task-oriented Dialogue and User Simulation: Explored robust data-driven approaches for end-to-end dialogue, e.g., unified user simulators [19], and incremental or recursive reasoning for state tracking. • Surveying Foundational Dialogue Tasks: Summarized key tasks and methods in end-to-end dialogues [18]

2022–2023	<ul style="list-style-type: none"> • Proactive Dialogue Systems: Studied the potential of question planning with reflection [7], proactive dialogue in the LLM era [3], and target-driven promotion strategies [2]. • New Intent Discovery and Ontology Maintenance: Proposed prompt-based cluster discovery [10] and slot discovery [21]. • SIGIR Tutorial: “Proactive Conversational Agent, in the Post-ChatGPT World” [16].
2024–2025	<ul style="list-style-type: none"> • Synergizing LLMs for Intent Discovery and Ontology Expansion: Emphasis on synergy between large language models and smaller pre-trained models [11] and advanced methods for category discovery [12]. • Hierarchical Policy Planning and Empathetic Dialogue: Simulation-free latent policy frameworks [9], dual-process conversation planning [8], and multimodal empathetic benchmarks [23]. • Human-centered and Emotional Support: Research on open-source avatar-based empathetic agents [6] and broadening demonstration-augmented prompting [1].

4 Key Research Accomplishments

4.1 Foundational Work on Knowledge-aware, Multimodal Dialogue

My early work [15] (Best Paper Finalist at ACM Multimedia 2018) established the foundation for integrating structured knowledge into multimodal dialogue systems. The key challenge addressed was how to bridge semantic gaps between textual and visual inputs, a crucial problem in conversational AI. My research introduced knowledge-aware fusion models that enabled systems to incorporate external structured knowledge, improving dialogue comprehension, response generation, and information retrieval. This work has significantly influenced multimodal search, conversational recommendation, and interactive AI assistants by demonstrating how contextualized multimodal reasoning enhances conversational quality. Currently, I am continuing to explore multimodal foundation models via the AISG National Multimodal LLM Programme (NMLP) grant: “Trust-worthy Multimodal Foundation Models: A Scalable Multi-Agent Approach”. The output of this research grant will surely enhance multimodal dialogue systems from the application angle.

4.2 Advances in Proactive Dialogue and Policy Planning

Many of my recent contributions center on proactive, hierarchical, or simulation-free policy models, aiming to move beyond reactive dialogues and enable AI systems to anticipate user needs:

- **Simulation-Free Hierarchical Latent Policy Planning:** Traditional reinforcement learning methods in dialogue systems rely on user simulators, leading to bias

and inefficiencies. This work [9] introduced a hierarchical reinforcement learning framework that learns directly from real-world dialogues, enabling more scalable, flexible, and effective proactive policy planning. This innovation outperforms LLM-based baselines while maintaining computational efficiency.

- **Target-driven and Reflection-based Question Planning:** Proactively gathering user information is critical for goal-oriented conversations. My research [7,24] proposed reflection-based reasoning and demonstration-guided planning, allowing AI agents to ask more informative questions, optimize conversation flows, and increase user engagement. These approaches have led to improved task success rates and user satisfaction in real-world dialogue applications.
- **Multi-intent and Mixed Initiative Dialogue Management:** Handling evolving and ambiguous user intents is a major challenge in conversational AI. My research [22, 17] introduced multi-intent understanding and dynamic initiative control, allowing AI models to seamlessly adapt to changing user needs, generate flexible responses, and proactively guide conversations.
- **Reinforced Target-driven Conversational Promotion:** Conversational recommendation systems often struggle with passive user interactions. This work [2] established a reinforcement learning-based framework for conversational recommendation, demonstrating how goal-driven strategies can maximize engagement, optimize product exposure, and improve user retention in digital marketing and e-commerce applications.

4.3 Empathetic and Human-Centered Conversational AI

Understanding and responding to human emotions is essential for empathetic AI interactions. My research has tackled key challenges in multimodal emotion recognition, empathetic response generation, and human-centered conversational agents:

- **Multimodal Empathetic Benchmarks and Open-source Solutions:** Existing empathetic dialogue datasets lack multimodal signals. My research [23, 6] introduced new text-speech-vision benchmarks and open-source frameworks for emotion-aware AI. These contributions have paved the way for AI-driven mental health support and emotionally intelligent virtual assistants.
- **Self-Chats for Emotionally Intelligent AI:** Large-scale language models (LLMs) lack efficiency for deployment in low-resource settings. My work [25] explored knowledge distillation from LLM-generated self-dialogues to train lightweight emotion-aware chatbots, demonstrating improved scalability and response effectiveness.
- **Human-centered Adaptive AI Agents:** Building AI that adapts to user well-being and affective states is a critical challenge. My research [5] introduced frameworks for emotion-sensitive policy learning, allowing dialogue systems to

dynamically adjust conversational tone, engagement strategies, and response personalization.

4.4 Ontology Expansion and Discovery for Conversational AI

Conversational AI requires domain adaptability to remain relevant in evolving user scenarios. My research has contributed to scalable methods for automated ontology expansion and intent discovery:

- **Ontology Expansion for Conversational Understanding:** Traditional ontology development is manual and labor-intensive. My research [13] proposed semi-supervised expansion models that allow AI systems to dynamically learn and refine domain knowledge from structured and unstructured data.
- **Active Intent Discovery from Large Language Models:** Identifying new user intents without predefined labels is a major challenge. My work [12] developed prompt-driven active learning techniques, allowing AI models to autonomously detect emerging conversational needs and adapt their response strategies.

4.5 Real-World Applications and Broader Impact

My research has had significant academic and practical impact, including:

- **Recognition in Top AI Conferences:** My work has been presented in ACL, EMNLP, SIGIR, AAI, and TOIS, influencing both academia and industry.
- **Citation and Adoption:** Many of my research contributions have been widely cited and adopted in follow-up studies, demonstrating their influence in proactive dialogue planning, conversational recommendation, and empathetic AI.
- **Industry and Open-source Contributions:** Several techniques I developed have been incorporated into AI-driven systems. For example, we adopted these to an industry project from DSO National Laboratories for Text Style Transfer with Pre-trained Language Models. We also published our codes and data for most of our works via platforms such as github. Through these efforts, my research continues to push the boundaries of human-centered, proactive, and emotionally intelligent conversational AI.

5 Future Directions

5.1 Synergizing LLMs with Enhanced Planning Architectures

Although Large Language Models excel at producing fluent and coherent text, they often lack strong grounding and robust planning ability. To address these limitations, I propose a dual-stage or dual-process framework [8], wherein an LLM handles open-domain generation and linguistic flexibility, while a smaller, specialized module is responsible for structured planning, domain constraints, and user modeling [11]. This design leverages the strengths of each component combining the generative prowess

of LLMs with a more interpretable, domain-aware backbone that ensures accuracy and contextual relevance.

5.2 Towards Proactive and Psychosocially Aware Systems

Moving beyond conventional Q&A or transactional tasks, I envision empathetic systems designed for mental health, emotional well-being, and broader social good. Building on insights from [23], I plan to incorporate more nuanced models of user affect and social context, necessitating interdisciplinary collaborations with psychologists, HCI researchers, and healthcare professionals. In parallel, I aim to unify proactive planning with empathy modeling, allowing a system to consider both task success and emotional resonance. This entails dynamically adapting the agent's style, content, and initiative level to user context [20], ultimately enabling more supportive and human-centric dialogue experiences.

5.3 Adaptive Personalization and Continual Learning

Increasing personalization demands that systems accumulate user-specific knowledge and adapt swiftly to changing preferences or emotional states [6]. In this vein, I plan to develop robust continual learning strategies that effectively manage user drift while preserving privacy. Beyond short-term dialogues, I will also investigate long-term personalization, wherein the agent iteratively refines its understanding of each user over repeated sessions (e.g., emotional patterns, domain preferences). This approach promises more adaptive recommendations and deeper emotional rapport, ultimately benefiting applications such as mental-health counseling or lifelong learning.

5.4 Next-generation Evaluation and User Simulation

Robust evaluation remains a critical bottleneck for proactive and empathetic dialogue systems. While many studies rely on static corpora or crowd-sourced approximations, I aim to develop more realistic user simulators that account for multiple intents, emotional fluctuations, and ongoing knowledge expansion [19]. This will allow us to conduct large-scale tests of proactive strategies without incurring the prohibitive costs of human-in-the-loop experimentation. In tandem, I will design novel in-the-wild user studies to ensure the resulting systems are safe, trustworthy, and beneficial in real-world settings.

6 Conclusion

My research continues to push the boundaries of human-centered, proactive conversational AI, weaving together key ideas from knowledge representation, multimodal understanding, policy learning, and empathy modeling. By iterating through foundational knowledge-aware systems, real-time user simulation, and proactive policy frameworks, I aim to create next-generation conversational agents that not only answer user queries but also anticipate, support, and care about their

users' needs and well-being. I look forward to future collaborations and interdisciplinary opportunities to realize this vision.

References

[1] Huy Dao, Yang Deng, Dung Le, and Lizi Liao. Broadening the view: Demonstration-augmented prompt learning for conversational recommendation. In Proceedings of the 2024 ACM SIGIR Conference on Research and Development in Information Retrieval, 2024.

[2] Huy Dao, Lizi Liao, Dung D. Le, and Yuxiang Nie. Reinforced target-driven conversational promotion. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[3] Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. Proactive dialogue

system, in the era of large language models: Evaluating from a prompting perspective. In Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[4] Yang Deng, Lizi Liao, Wenqiang Lei, Grace Hui Yang, Wai Lam, and Tat-Seng Chua. Proactive conversational ai: A comprehensive survey of advancements and opportunities. TOIS, 2025.

[5] Yang Deng, Lizi Liao, Zhonghua Zheng, Grace Hui Yang, and Tat-Seng Chua. Towards human-centered proactive conversational agents. In Proceedings of the 2024 ACM SIGIR Conference on Research and Development in Information Retrieval, 2024.

[6] Hao Fei, Han Zhang, Bin Wang, Lizi Liao, Qian Liu, and Erik Cambria. EmpathyEar: An open-source avatar multimodal empathetic chatbot. In Proceedings of the 2024 ACL (Demonstrations), 2024.

[7] Shasha Guo, Lizi Liao, Jing Zhang, Cuiping Li, and Hong Chen. PCQPR: Proactive conversational question planning with reflection. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[8] Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Ming Liu, Zerui Chen, and Bing Qin. Planning like human: A dual-processframeworkfordialogueplanning. In Proceedings of the 2024 Annual Meeting of the Association for Computational Linguistics (ACL), 2024.

[9] Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Yiheng Sun, Zerui Chen, Ming Liu, and Bing Qin. Simulation-free hierarchical latent policy planning for proactive dialogues. In Proceedings of the AAAI Conference on Artificial Intelligence, 2025.

[10] Jinggui Liang and Lizi Liao. ClusterPrompt: Cluster semantic enhanced prompt learning for new intent discovery. In Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[11] Jinggui Liang, Lizi Liao, Hao Fei, and Jing Jiang. Synergizing large language models and pre-trained smaller models for conversational intent discovery. In Findings of the 2024 Annual Meeting of the Association for Computational Linguistics (ACL Findings), 2024.

[12] Jinggui Liang, Lizi Liao, Hao Fei, Bobo Li, and Jing Jiang. Actively learn from llms with uncertainty propagation for generalized category discovery. In Proceedings of the 2024 Conference of the North American

Chapter of the Association for Computational Linguistics (NAACL), 2024.

[13] Jinggui Liang, Yuxia Wu, Yuan Fang, Hao Fei, and Lizi Liao. A survey of ontology expansion for conversational understanding. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[14] Lizi Liao, Le Hong Long, Zheng Zhang, Minlie Huang, and Tat-Seng Chua. MMConv: An environment for multimodal conversational search across multiple domains. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021.

[15] Lizi Liao, Yunshan Ma, Xiangnan He, Richang Hong, and Tat-Seng Chua. Knowledge-aware multimodal dialogue systems. In Proceedings of the 26th ACM International Conference on Multimedia (ACM MM), 2018. Best Paper Final List.

[16] Lizi Liao, Grace Hui Yang, and Chirag Shah. Proactive conversational agent, in the post-ChatGPT world. In Proceedings of the 2023 ACM SIGIR Conference on Research and Development in Information Retrieval (Tutorial), 2023.

[17] Yuxiang Nie, Heyan Huang, Xian-Ling Mao, and Lizi Liao. Mix-initiative response generation with dynamic prefix tuning. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2024.

[18] Libo Qin, Wenbo Pan, Qiguang Chen, Lizi Liao, Zhou Yu, Yue Zhang, Wanxiang Che, and Min Li. End-to-end task-oriented dialogue: A survey of tasks, methods, and future directions. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[19] Dazhen Wan, Zheng Zhang, Qi Zhu, Lizi Liao, and Minlie Huang. Unified dialogue user simulator for few-shot data augmentation. In Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.

[20] Zhaohui Wei, Lizi Liao, Xiaoyu Du, and Xinguang Xiang. Balancing visual context understanding, in dialogue for image retrieval. In Findings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[21] Yuxia Wu, Tianhao Dai, Zedong Zheng, and Lizi Liao. Actively discovering new slots for task-oriented conversation. IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP), 2024.

[22] Bowen Xing, Lizi Liao, Minlie Huang, and Ivor Tsang. DC-Instruct: An effective framework for generative multi-intent spoken language understanding. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[23] Han Zhang, Zixiang Meng, Meng Luo, Hong Han, Lizi Liao, Erik Cambria, and Hao Fei. Towards multimodal empathetic response generation: A rich text-speech-vision avatar-based benchmark. In Proceedings of The Web Conference, 2025.

[24] Zhonghua Zheng, Lizi Liao, Yang Deng, Ee-Peng Lim, Minlie Huang, and Liqiang Nie. Thoughts to target: Enhance planning for target-driven conversation. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[25] Zhonghua Zheng, Lizi Liao, Yang Deng, Libo Qin, and Liqiang Nie. Self-chats from large language models Make small ChatPal better. In Proceedings of the 2024 Annual Meeting of the Association for Computational Linguistics (ACL), 2024.