Risk-Aware Bilingual Spoken Dialogue for Campus Mental Health Support

You-Teng Lin, Li-Yang Zhang, Yi-Tang Chen, Jen-Tzung Chien

Institute of Electrical and Computer Engineering
National Yang Ming Chiao Tung University, Hsinchu, Taiwan
{utungll04.ee12, tony.ee12, tom.ee13, jtchien}@nycu.edu.tw

Abstract

This work presented a web-based system which introduces an active-listening strategy in a spoken dialogue for self-disclosure to support mental health of a campus user. To enhance the system usability and safety, this demo is developed to conduct the bilingual (Mandarin/English) spoken dialogue where a high-risk dialogue detection during speech interaction is reliably augmented. In particular, a prompt-driven GPT classifier identifies the utterances indicating self-harm or suicide intent and triggers safety alerts with help center and counselor notification. We also integrate a TTS module for Taiwanese Mandarin and standard English, and redesign the user interface to automatically pop up alert messages when high-risk dialogue is detected. In addition, we collect speech data under diverse mental dialogue scenarios with bilingual speech to enable system analysis, evaluation and refinement. Overall, these extensions build a framework that promotes empathetic interactions, enables timely alert in critical cases, and improves the accessibility for diverse users.

Introduction

Mental health among university students, staffs and faculties has become increasingly challenging, while counseling resources remain limited, often leading to a long waiting time before professional support is available (Sinha et al. 2019; Roy et al. 2020). Spoken dialogue system has the potential to encourage self-disclosure and provide timely auxiliary support. A recent research in (Rohmatillah et al. 2024, 2023a) has introduced an active-listening spoken dialogue system for empathetic interactions in Mandarin.

Building on this foundation, this research addresses two key gaps for practical deployment: (1) early detection of high-risk dialogue content for real-time alert display, and (2) bilingual (Mandarin/English) speech interaction to broaden accessibility. We develop a prompt-based high-risk detection mechanism (Costello 2016), where GPT (OpenAI 2024) classifies user utterances into the predefined risk levels and triggers alerts with mental health resources and counselor notification when suicidal intent is detected. This work further integrates a bilingual TTS module for Taiwan-accented Mandarin and formal English, together with a redesigned user interface that provides real-time alert displays when high-risk dialogue is detected. In addition, we collect simulated counseling dialogues (ConvCounsel) (Chen et al. 2024) and tester logs (ConvCounsel 2.0), providing both strategy-rich annotations and real-world data.

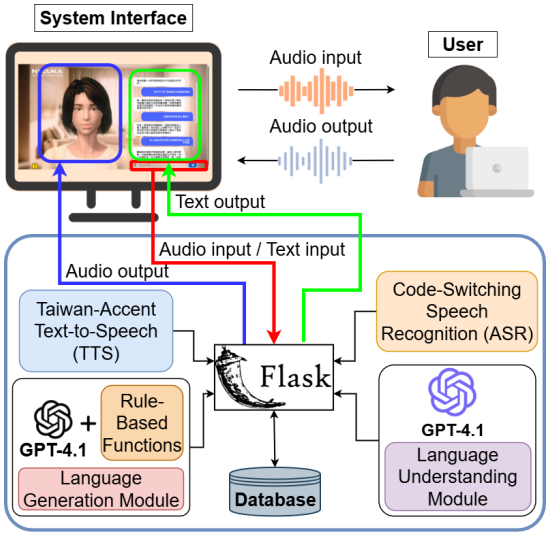

Figure 1: Nycuka dialogue system. User inputs are processed by ASR/NLU and middleware with database logging, while responses from LLM and rule-based modules are delivered via bilingual text or speech. Flask framework integrates modules and provides a unified web interface for real-time, safety-aware interaction with alert shown in Figure 2.

This work makes four contributions: (1) introducing multi-granularity high-risk detection with real-time alerts; (2) developing a bilingual Mandarin–English ASR/TTS (Zhang et al. 2023); (3) designing a safety-oriented interface; and (4) collecting new text and speech data to support evaluation of mental health for students, staffs and faculties.

System Overview

This study designs a real-time mental health dialogue system as an active listening agent as shown in Figure 1. It supports spoken dialogue in traditional Chinese and Taiwanese Mandarin, integrating automatic speech recognition (ASR), natural language understanding (NLU), dialogue management

Risk level	Prototypical user prompts
No Risk	“Everything’s fine.”; “Had a good lunch. Feel good.”
Low Risk	“I feel overwhelmed.”; “I’m down and don’t want to do anything.”
Moderate Risk	“I feel exhausted and nothing seems meaningful.”; “It feels like I’m only pretending to be fine.”
High Risk	“I’ve planned to end my life tomorrow.”; “I want to harm myself.”

Table 1: Summary of four risk levels and the corresponding prototypical user prompts.

(DM) (Traum and Larsson 2003), natural language generation (NLG), text-to-speech (TTS), and high-risk detection. All conversations are securely stored, and key information can be extracted to assist counselor for decision-making. System website is provided in https://www.nycuka.com.tw/.

Speech Component

The speech component integrates ASR (Radford et al. 2023) and TTS (Zhang et al. 2023). ASR module is based on Whisper (Radford et al. 2023) and fine-tuned with Taiwan-accented data (Rohmatillah et al. 2023b) to capture phonetic and prosodic variations, while further supporting the Mandarin-English code-switching (Aditya et al. 2024). TTS module adopts a two-stage strategy: first, pre-training on large China-accented speech corpora to learn general acoustic patterns; then, fine-tuning with Taiwan-accented speech data to model pronunciation and prosody accurately. In addition, with the punctuation-aware prosody control, the system is feasible to generate fluent and natural speech responses.

Text Component

The text component integrates NLU, DM, NLG, and high-risk detection. NLU extracts the intent, emotions, and semantic features. DM manages dialogue flow with counselor-informed strategies through prompts (Rohmatillah and Chien 2021, 2024a,b; Chien and Huang 2025). NLG produces fluent, empathetic responses (Chien and Wu 2024; Chien and Liu 2025). A large language model (LLM)-based classifier assigns each utterance a predefined risk level, triggering an alert to notify counselors when a high-risk intent (e.g. self-harm or suicide) is detected. This unified pipeline supports a natural interaction with a campus user while ensuring timely intervention in critical cases.

Risk-Aware Prompt Design

The high-risk detection module uses a prompt-based LLM classifier to evaluate and assign each user utterance into four predefined risk levels: no, low, moderate, and high as summarized with examples in Table 1. When a high-risk case is detected, the back-end generates a trigger token monitored by the front-end, which activates a prevention alert containing counseling resources and notifying counselors. This decoupled design keeps the risk classification separate from the user interface (UI) rendering, ensuring easy updates and immediate alert displays with minimal latency.

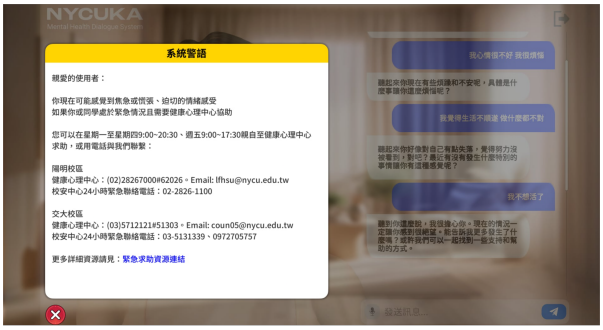

Figure 2: When high-risk utterances are detected, the system immediately displays an alert with counselor and emergency contact information to guide users towards supports.

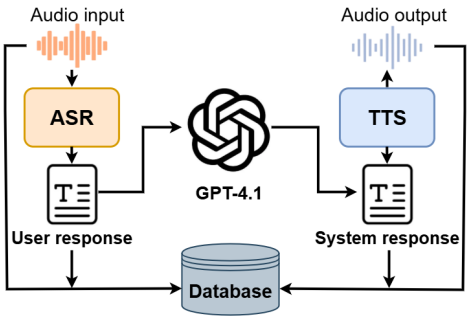

Figure 3: Data collection pipeline. User utterance is transcribed by ASR, processed by large language model for dialogue flow and risk detection, and synthesized by TTS for dialogue response. All interactions are stored in the database.

Data Collection

For speech data, this study collected and preprocessed 400 hours of audio for ASR/TTS model training. For text data, two sources are utilized: ConvCounsel (Chen et al. 2024), comprising 40 simulated counseling sessions with annotations of roles, emotions, and strategies; and ConvCounsel 2.0, derived from interaction logs of 15 participants, which include user inputs, system responses, and associated risk levels. Each log entry further records the session date, user identifier, dialogue turn index, and utterance content. The data collection and processing pipeline is shown in Figure 3.

Conclusions

This work extends an active-listening mental health dialogue system with high-risk detection, bilingual speech interaction, and a safety-oriented interface, supported by new collected text and speech data. A prompting mechanism based on large language models enables real-time identification of high-risk intent (e.g. self-harm) and automatic alerts with counselor notification. A bilingual TTS module and redesigned UI improve accessibility and usability. Together, these components create a framework that balances empathetic interaction with real-time alerting in critical cases, demonstrating the value of integrating counseling expertise into dialogue systems.

Acknowledgments

This work is supported by the National Science and Technology Council, Taiwan, under NSTC 113-2634-F-A49-006.

References

Aditya, B.; Rohmatillah, M.; Tai, L.-H.; and Chien, J.-T. 2024. Attention-guided adaptation for code-switching speech recognition. In *Proc. of IEEE International Conference On Acoustics, Speech And Signal Processing*, 10256–10260.

Chen, P.-C.; Rohmatillah, M.; Lin, Y.-T.; and Chien, J.-T. 2024. ConvCounsel: a conversational dataset for student counseling. In *Proc. of Conference of the Oriental COCOSDA*, 1–6.

Chien, J.; and Liu, H. 2025. Contrastive Disentanglement Learning for Empathetic Generation. In *Proc. of International Workshop on Machine Learning for Signal Processing*, 1–6.

Chien, J.-T.; and Huang, P.-C. 2025. CAPR: Confidence-Aware Prompt Refinement in Large Language Models. In *Proc. of Annual Conference of the International Speech Communication Association*, 3264–3268.

Chien, J.-T.; and Wu, Y.-C. 2024. Empathetic Response Generation via Regularized Q-Learning. In *Proc. of Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 1–6.

Costello, E. J. 2016. Early detection and prevention of mental health problems: developmental epidemiology and systems of support. *Journal of Clinical Child & Adolescent Psychology*.

OpenAI. 2024. GPT-4 Technical Report.

Radford, A.; Kim, J. W.; Xu, T.; Brockman, G.; McLeavey, C.; and Sutskever, I. 2023. Robust speech recognition via large-scale weak supervision. In *Proc. of International Conference on Machine Learning*.

Rohmatillah, M.; Aditya, B.; Ngo, B. G.; Chen, P. C.; Sulaiman, W.; and Chien, J.-T. 2023a. NYCUKA: a self-disclosure mental health spoken dialogue system. In *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop*.

Rohmatillah, M.; Aditya, B.; Yang, L.-J.; Ngo, B. G.; Sulaiman, W.; and Chien, J.-T. 2023b. Promoting Mental Self-Disclosure in a Spoken Dialogue System. In *Proc. of Annual Conference of the International Speech Communication Association*, 670–671.

Rohmatillah, M.; and Chien, J.-T. 2021. Corrective guidance and learning for dialogue management. In *Proc. of ACM International Conference on Information & Knowledge Management*, 1548–1557.

Rohmatillah, M.; and Chien, J.-T. 2024a. Revise the NLU: A prompting strategy for robust dialogue system. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 10956–10960.

Rohmatillah, M.; and Chien, J.-T. 2024b. Taming NLU Noise: Student-Teacher Learning for Robust Dialogue Policy. In *Proc. of IEEE Spoken Language Technology Workshop*, 849–856.

Rohmatillah, M.; Ngo, B. G.; Sulaiman, W.; Chen, P.-C.; and Chien, J.-T. 2024. Reliable dialogue system for facilitating student-counselor communication. In *Proc. of Annual Conference of the International Speech Communication Association*, 1003–1004.

Roy, A.; Nikolitch, K.; McGinn, R.; Jinah, S.; Klement, W.; and Kaminsky, Z. A. 2020. A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digital Medicine*.

Sinha, P. P.; Mishra, R.; Sawhney, R.; Mahata, D.; Shah, R. R.; and Liu, H. 2019. #suicidal-A multipronged approach to identify and explore suicidal ideation in twitter. In *Proc. of the ACM International Conference on Information and Knowledge Management*, 941–950.

Traum, D. R.; and Larsson, S. 2003. The information state approach to dialogue management. In *Current and New Directions in Discourse and Dialogue*. Springer.

Zhang, Z.; Zhou, L.; Wang, C.; Chen, S.; Wu, Y.; Liu, S.; Chen, Z.; Liu, Y.; Wang, H.; Li, J.; et al. 2023. Speak foreign languages with your own voice: cross-lingual neural codec language modeling. *arXiv preprint arXiv:2303.03926*.