Fuzzy Inference–Driven Knowledge Modeling for LLM-Assisted Human–Machine Co-Learning

Chang-Shing Lee, Mei-Hui Wang, Chao-Cyuan Yue, Sheng-Chi Yang
National University of Tainan
Tainan, Taiwan
leecs@mail.nutn.edu.tw

Yusuke Nojima, Naoki Masuyama, Takeru Konishi, Ryosuke Saga
Osaka Metropolitan University
Osaka, Japan
{nojima@, masuyama@, sc23197z@st., r.saga@}omu.ac.jp

***Abstract*—This paper introduces a fuzzy inference–driven knowledge modeling framework for evaluating LLM-assisted human–machine co-learning performance in multilingual speech-based learning environments. The proposed framework integrates Taiwanese-Japanese Open AI Whisper–based Automatic Speech Recognition (ASR), fuzzy Large Language Model (LLM)–based translation using the Trustworthy AI Dialogue Engine (TAIDE), and a personalized fine-tuned Meta AI Massively Multilingual Speech (MMS)-based Taiwanese Text-To-Speech (TTS) model. The framework has been successfully deployed in real-world human-machine co-learning scenarios in Taiwan and Japan. Based on the outputs of these components, three fuzzy evaluation variables are derived, including fuzzy ASR similarity (*ASRsim*), fuzzy LLM translation similarity (*LLMsim*), and Mean Opinion Score (MOS) of the fuzzy Taiwanese TTS model (*TTSmos*). These fuzzy variables are further utilized to construct a knowledge model and a quantum fuzzy inference model for assessing human–machine co-learning performance and generating a simulated quantum circuit. To evaluate the performance of the Taiwanese Whisper–based ASR model, the complete text of the *Heart Sutra*, together with selected passages from the *Diamond Sutra* and the *Tao Te Ching*, is used as benchmark data across different training checkpoints. In particular, the complete text of the *Heart Sutra* serves as the primary benchmark for evaluating the performance of the fine-tuned OpenAI Whisper Taiwanese-Japanese ASR model. Experimental results demonstrate that the proposed framework effectively supports students' learning processes and is capable of accurately recognizing Taiwanese-Japanese speech and mapping it to corresponding English and Chinese texts. These findings indicate that the proposed fuzzy inference–driven framework provides a practical and effective approach for multilingual human–machine co-learning evaluation.**

Keywords—Fuzzy Inference, Whisper-Taiwanese ASR, LLM, Taiwanese TTS, Human-Machine Co-Learning

I. INTRODUCTION

AI sovereignty is critical for ensuring national control over data, language resources, and AI technologies. This motivation has led the Taiwanese government to develop a large language model (LLM), the Trustworthy AI Dialogue Engine (TAIDE) [1]. Through collaboration among industry, government, and academia, TAIDE has been adopted to support applications in multiple domains, including education, healthcare, and law [2]. In Taiwan, however, the number of speakers of Taiwanese (Tâi-gí), also known as Taiwanese Hokkien, has decreased markedly—especially among younger generations—due to prolonged language repression, with more than 60% of the population viewing Taiwanese as endangered [12]. To preserve Taiwan's heritage languages, Lee et al. [3] integrated Quantum Computational Intelligence (QCI), LLMs, and English learning into educational environments to encourage active use of Taiwanese. By enabling students to speak Taiwanese while interacting with advanced AI technologies, the proposed framework not only enhances learning engagement but also facilitates the collection of learning data within a natural, speech-based human–machine co-learning environment.

Taiwanese Automatic Speech Recognition (ASR) faces significant challenges because Taiwanese is predominantly an oral language and lacks a fully standardized writing system, making conventional text-based translation approaches inadequate [4]. Meta's Universal Speech Translator (UST) [4] highlights the importance of robust Taiwanese Hokkien ASR for enabling multilingual learning and interaction in educational contexts. However, UST is not designed to fully address the diversity and domain specificity of instructional materials used in school-based environments. To address this limitation, the Whisper-Taiwanese model [5] was developed by fine-tuning the OpenAI Whisper foundation model [9] using speech data collected directly from real educational settings, thereby improving recognition performance for classroom-oriented Taiwanese speech. In addition to ASR, speech generation is a critical component of spoken human–machine interaction. Accordingly, the Taiwanese Text-To-Speech (TTS) model in this work is developed by fine-tuning the Meta AI Massively Multilingual Speech (MMS) foundation model [10]. Leveraging cross-lingual representations learned from large-scale multilingual speech data, the MMS model is used as the foundation to train a Taiwanese TTS model with high-quality Taiwanese speech recordings, supporting speech synthesis for a low-resource language and language preservation.

The *Heart Sutra* is widely regarded as a classic text embodying human wisdom. Its Chinese version consists of 260 characters, while the Japanese version contains 262 characters. Lee et al. [6] adopted the *Heart Sutra* as a foundational resource for modeling human–machine co-learning, interpreting its structure as a six-step learning process—(1) *observe and go*, (2) *study*, (3) *research*, (4) *utilize and follow*, (5) *understand and know*, and (6) *explain and speak*—that maps human learning behaviors to machine learning mechanisms. This *Heart Sutra*–inspired framework has been successfully applied in educational environments to support interactive co-learning between students and intelligent systems. This paper integrates Taiwanese-Japanese ASR, the TAIDE LLM, and Taiwanese TTS into a unified human–machine co-learning framework for

The authors would like to thank the financial support sponsored by the National Science and Technology Council (NSTC) of Taiwan under the grants NSTC 114-2622-E-024-002 and NSTC 114-2221-E-024-004.

school-based education. The proposed system enables machines to support listening, language understanding, and speech generation in interactive learning scenarios and has been successfully deployed in real-world classroom settings in Taiwan and Japan. Based on these components, three evaluation fuzzy variables: *ASR similarity* (*ASRsim*), *LLM translation similarity* (*LLMsim*), and *TTS mean opinion score* (*TTSmos*), are derived and incorporated into a knowledge model with a quantum fuzzy inference mechanism for co-learning ASR and TTS performance assessment. To evaluate the Taiwanese-Japanese ASR performance, benchmark data include the full *Heart Sutra*, along with selected passages from the *Diamond Sutra* and the *Tao Te Ching*, evaluated across different training checkpoints, with the *Heart Sutra* serving as the primary benchmark for Japanese speech recognition. Experimental results demonstrate that the proposed framework effectively supports student learning and accurately maps Japanese speech to corresponding English and Chinese texts, indicating its practicality for multilingual human–machine co-learning evaluation.

The remainder of this paper is organized as follows. Section II introduces the structure of the LLM-assisted human–machine co-learning framework. Section III presents the quantum fuzzy inference-driven knowledge modeling approach, including the knowledge base and fuzzy rule base, for performance evaluation. Section IV describes the human–machine co-learning evaluation framework for Taiwanese–Japanese ASR and MMS-based personalized Taiwanese TTS. Section V presents the experimental results, and Section VI concludes the paper.

II. LLM-Assisted Human-Machine Co-Learning

A. LLM-based Whisper Taiwanese ASR Model

Fig. 1 illustrates the overall structure of the LLM-based Whisper-Taiwanese ASR model for the Taiwanese-English–Japanese co-learning framework. The proposed AI learning companion system supports human–machine co-learning by generating multimodal learning data from both educational materials and classical texts of human wisdom, including the *Heart Sutra*, the *Diamond Sutra*, and the *Tao Te Ching*.

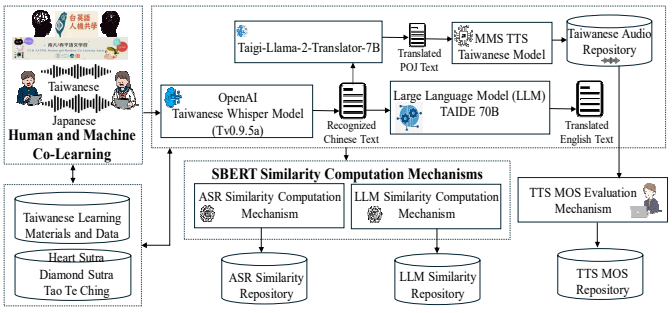


Fig. 1. LLM-based Taiwanese-English-Japanese co-learning structure.

During the learning process, Taiwanese or Japanese speech produced by learners is first transcribed into Chinese text by the *OpenAI Taiwanese Whisper ASR model* [5]. In this framework, the ASR model takes Taiwanese speech as input and outputs the corresponding recognized text in Chinese. This recognized Chinese text is then converted into POJ (Pe̍h-ōe-jī), a romanized writing system for Taiwanese pronunciation, using the Taigi-Llama2 translator model [8]. The resulting POJ text is subsequently used as the input to the Meta MMS Taiwanese TTS model to synthesize Taiwanese speech audio. In parallel, the recognized Chinese text is also translated into English by the TAIDE [1] large language model. The synthesized Taiwanese speech is stored in the Taiwanese audio repository and returned to the learner as auditory feedback.

To evaluate co-learning performance, multiple fuzzy-valued assessment variables are defined. Specifically, fuzzy ASR similarity and fuzzy LLM translation similarity are computed using SBERT-based semantic similarity mechanisms, while synthesized speech quality is evaluated using a MOS–based fuzzy TTS evaluation mechanism. These evaluation results are stored in corresponding repositories and further integrated into a knowledge model with a quantum fuzzy inference mechanism, enabling comprehensive assessment of human–machine co-learning effectiveness across fuzzy speech recognition, fuzzy language understanding, and fuzzy speech generation tasks.

B. Fuzzy MMS Taiwanese TTS Model

As illustrated in Fig. 2, this study adopts POJ as the phonetic representation for fuzzy Taiwanese speech modeling. As a phoneme-oriented romanization system that explicitly encodes initials, finals, and tonal information, POJ provides a personalized speech–text mapping and avoids pronunciation ambiguity, making it suitable for personalized ASR and TTS training. POJ is also widely used in historical Taiwanese resources and serves as a standardized linguistic representation within the fuzzy MMS framework.

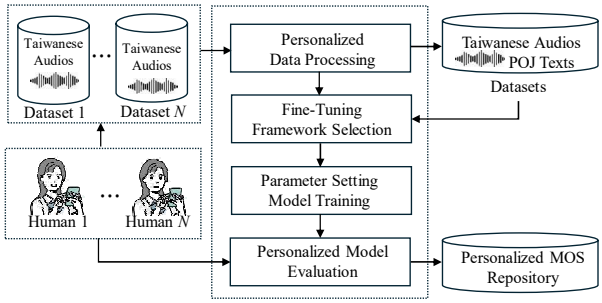


Fig. 2. Fuzzy MMS Taiwanese TTS structure.

The proposed Taiwanese speech synthesis system is built on the MMS framework developed by Meta AI and follows a pipeline comprising personalized data processing, fine-tuning framework selection, model training, and evaluation. Taiwanese speech recordings and their corresponding POJ texts are first prepared through personal data processing, where the Taiwanese text is converted into phoneme representations using the *Taiwanese Hokkien Transliterator* and *Tokeniser toolkit*. A suitable fine-tuning framework is then selected, followed by parameter configuration and model training using paired POJ phoneme sequences and a personalized Taiwanese audio data set. Finally, model performance is evaluated through a human-in-the-loop assessment process. Synthesized personal Taiwanese speech is reviewed by human listeners, and the evaluation results are recorded in a MOS repository using a five-point MOS

scale, providing a subjective but practical measure of fuzzy speech synthesis quality.

III. Quantum Fuzzy Inference-Driven Knowledge Modeling for Performance Evaluation

A. Quantum Fuzzy Inference-Driven Knowledge Model

This section introduces a fuzzy inference–driven knowledge model that integrates three input fuzzy variables derived from system evaluation metrics: *ASRsim*, *LLMsim*, and *TTSmos*. *ASRsim* and *LLMsim* were partitioned into three linguistic labels to achieve a balance between model simplicity and interpretability. *ASRsim* quantifies the semantic similarity between the reference Chinese text and the Chinese text recognized by the *Whisper-Taiwanese ASR model*. Sentence-BERT (SBERT) [11] is employed to encode both texts into dense semantic embeddings, and cosine similarity is used to compute a normalized similarity score in the range [0, 1]. The SBERT-based similarity is computed by Eq. (1). *LLMsim* is computed in the same manner and measures the semantic similarity between the English text generated by the TAIDE large language model and the corresponding reference English text, with values also normalized to the range [0,1].

$$sim(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \tag{1}$$

where **u** and **v** denote the SBERT embedding vectors of the reference text and the generated (or recognized) text, respectively.

TABLE I. Parameters of the Fuzzy Linguistic Terms.

Fuzzy Variable	Linguistic Term	Trapezoidal Parameters [a, b, c, d]
ASRsim	*Low*	[0, 0, 0.3, 0.4]
	Medium	[0.3, 0.4, 0.6, 0.7]
	High	[0.6, 0.7, 1, 1]
LLMsim	*Low*	[0, 0, 0.3, 0.4]
	Medium	[0.3, 0.4, 0.6, 0.7]
	High	[0.6, 0.7, 1, 1]
TTSmos	*Bad*	[1, 1, 1, 2]
	Poor	[1, 2, 2, 3]
	Fair	[2, 3, 3, 4]
	Good	[3, 4, 4, 5]
	Excellent	[4, 5, 5, 5]
Performance	*Low*	[0, 0, 5, 6]
	Medium	[5, 6, 7, 8]
	High	[7, 8, 10, 10]

TTSmos was defined with five linguistic labels to follow the conventional 5-point *Mean Opinion Score* (*MOS*) scale for speech quality evaluation, where higher scores indicate better perceptual quality and can help teachers identify areas in which a student may need improvement, such as recognition accuracy, semantic adequacy, and speech quality. *Performance* denotes the integrated evaluation result of human–machine co-learning and helps teachers identify areas in which a student may need improvement, such as recognition accuracy, semantic adequacy, and speech quality. Table I summarizes the fuzzy linguistic terms and corresponding trapezoidal membership function parameters of the fuzzy variables. A trapezoidal membership function is specified by four parameters [a, b, c, d] as Eq. (2). Fig. 3 shows the fuzzy sets with membership functions of *ASRsim*, *LLMsim*, *TTSmos*, and *Performance*.

$$\text{trapezoid}(x{:}a, b, c, d) = \begin{cases} 0, & x<a \\ (x-a)/(b-a), & a \le x<b \\ 1, & b \le x<c \\ (d-x)/(d-c), & c \le x<d \\ 0, & x \ge d \end{cases} \tag{2}$$

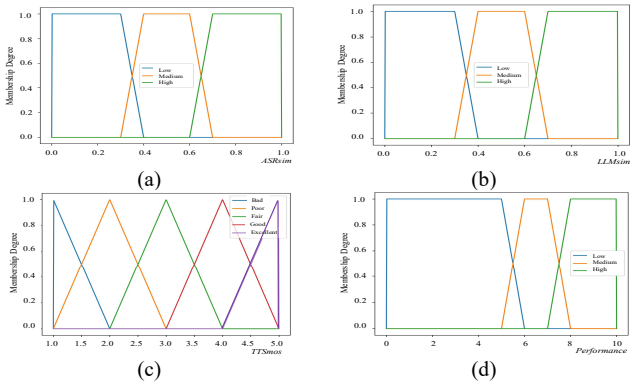


Fig. 3. Fuzzy sets with membership functions of (a) *ASRsim*, (b) *LLMsim*, (c) *TTSmos*, and (d) *Performance*.

B. Human Linguistic Variables and Rules

This subsection demonstrates how fuzzy variables and rules can be encoded and evaluated within a quantum circuit framework for human–machine co-learning. Based on the proposed fuzzy inference–driven knowledge model, a total of 45 fuzzy inference rules are established. *ASRsim* and *TTSmos,* respectively, reflect speech recognition accuracy and speech synthesis quality, which interact most directly with learners in human–machine co-learning scenarios and therefore exert a greater impact on the learning experience than *LLMsim*. Accordingly, the relative importance ratio of *ASRsim*, *LLMsim*, and *TTSmos* is set to 2:1:2. After normalization, the resulting normalized rule score is used to determine the consequent part of each fuzzy rule. Table II presents partial fuzzy rules.

TABLE II. Partial Fuzzy Rules.

Rule No.	Fuzzy Variables			
	ASRsim	*LLMsim*	*TTSmos*	*Performance*
1	Low	Low	Bad	Low
2	Low	Low	Poor	Low
3	Low	Low	Fair	Low
4	Low	Low	Good	Low
5	Low	Low	Excellent	Medium
⋮				
42	High	High	Poor	Medium
43	High	High	Fair	Medium
44	High	High	Good	High
45	High	High	Excellent	High

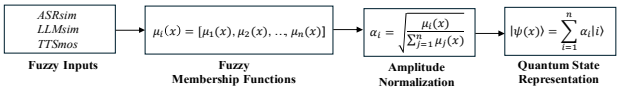


Fig. 4. Input fuzzy variable-to-quantum state representation encoding.

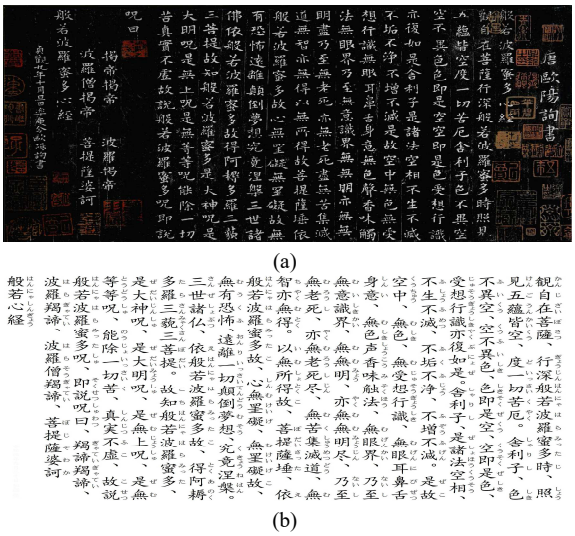
To integrate the fuzzy membership functions into the Quantum Fuzzy Inference Engine (QFIE), Fig. 4 illustrates the *input fuzzy variable-to-quantum state representation* encoding process. Three fuzzy input variables are first mapped to membership degrees through fuzzy membership functions to obtain the membership degrees of each linguistic term, denoted as $\mu_i(x)$. These membership degrees are then transformed into quantum amplitudes via amplitude normalization and finally encoded as a quantum state representation. For each linguistic term i, the corresponding quantum amplitude is defined as α_i. Using the normalized amplitudes, the fuzzy information is encoded into a quantum state representation expressed as $|\varphi\rangle$.

Take *ASRsim* = 0.927, *LLMsim* = 1.0, and *TTSmos* = 2.6 as an example. The corresponding fuzzy membership vectors are $\mu_{ASRsim}(0.927) = [0, 0, 1]$, $\mu_{LLMsim}(1.0) = [0, 0, 1]$, $\mu_{TTSmos}(2.6) = [0, 0.4, 0.6, 0, 0]$. After amplitude normalization, the nonzero amplitudes for *TTSmos* are α_{Poor} = sqrt (0.4 / (0.4 + 0.6)) ≈ 0.6325 and α_{Fair} = sqrt (0.6 / (0.4 + 0.6)) ≈ 0.7746. Thus, the corresponding quantum state is $|\psi_{TTSmos}\rangle = 0.6325|Poor\rangle + 0.7746|Fair\rangle$. Since both *ASRsim* and *LLMsim* are fully mapped to *High*, their quantum states are $|\psi_{ASRsim}\rangle = |High\rangle$ and $|\psi_{LLMsim}\rangle = |High\rangle$. This example shows that *ASRsim* and *LLMsim* are encoded as crisp *High* states, while *TTSmos* is encoded as a superposition of *Poor* and *Fair*, thereby preserving the fuzzy uncertainty in the quantum state representation.

IV. HUMAN–MACHINE CO-LEARNING EVALUATION FOR ASR AND TTS

A. Dataset Construction and ASR Evaluation

The *Heart Sutra* is not only a Buddhist scripture but also a significant text in the history of human wisdom. In Japanese cultural traditions, it is one of the most frequently recited scriptures, despite being written in Chinese characters. Fig. 5(a) and (b) show the *Heart Sutra* [6] in Chinese and Japanese, respectively, illustrating how classical wisdom texts adapt across cultures while preserving their core ideas.



(a)

(b)

Fig. 5. Heart Sutra versions in (a) Chinese and (b) Japanese.

Fig. 6 illustrates the overall framework for LLM dataset construction, model fine-tuning, and performance evaluation. Classical texts representing human intelligence, including the *Heart Sutra*, the *Diamond Sutra*, and the *Tao Te Ching*, are incorporated into learner-generated speech data through human–machine interaction. The collected raw data are processed by a data preprocessing mechanism and subsequently organized into training, validation, and testing datasets. These datasets are used to fine-tune the Whisper-Taiwanese models based on the OpenAI Whisper model, producing multiple training checkpoints, such as 1,000, 2,000, …, and 10,000 steps. To evaluate the effectiveness of each checkpoint model, selected subjects are invited to validate recognition performance across different checkpoints by reciting the classical texts in Taiwanese and Japanese using a dedicated model performance evaluation mechanism.

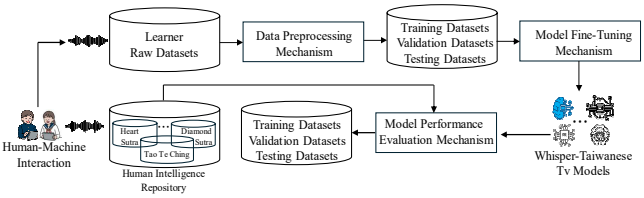


Fig. 6. Framework for LLM dataset construction and model evaluation using classical texts.

B. Taiwanese TTS Model with Human MOS Evaluation

In this paper, two lessons selected from the Jen Pin elementary-school Taiwanese textbook—a traditional fable narrative (*The Crow Drinks Water*) and a lesson on modern digital life (*The Online Shopping*)—were used as test texts for evaluating the performance of the Taiwanese TTS model through human MOS assessment. These two lessons contain a total of 31 sentences, including 17 sentences from *The Crow Drinks Water* and 14 sentences from *The Online Shopping*. Table III shows the texts of the *crow drinking water*.

TABLE III. TEXTS OF THE CROW DRINK WATER.

No.	Taiwanese Text	Chinese Text	TAIDE-based English Text
1	烏鴉食水	烏鴉喝水	The crow drinks water.
2	一隻烏鴉飛出去□ 𣍐	有一隻烏鴉飛出去玩	A crow flew out to play.
3	飛𣍐久嘛飛誠遠	飛了很久也飛了很遠	It flew for a long time and traveled very far.
4	感覺喙焦	牠覺得口很渴	It felt very thirsty.
5	毋過搵無半滴水通啉	不過卻找不到半滴水可以喝	But it couldn't find a single drop of water to drink.
6	伊看著樹跤一支矸仔	牠看到樹下有一只瓶子	It saw a bottle under a tree.
7	內底有寡水	裡面有些水	There was some water inside.
8	趕緊共喙襬入去	急忙把嘴塞進去	It quickly put its beak into the bottle.
8	無疑惜矸仔喙傷細	不料瓶口太小	But the bottle's opening was too small.
10	猶原食袂著水	一樣喝不到水	It still couldn't drink the water.
11	伊想欲共矸仔僅予倒	牠想把瓶子推倒	It tried to knock the bottle over.
12	閣驚水流了去	又怕水流光了	But it was afraid that the water would spill out.
13	行來想去的時陡踢著石頭	思來想去的時候卻踢到石頭	While thinking about it, it accidentally kicked a stone.
14	伊喝一聲	牠大喊一聲	It shouted:
15	啊有矣！來共石頭咬入去矸仔內	啊有了！把石頭咬進去瓶子裡	Ah! I've got it! Let's put the stones into the bottle.'
16	水就會𥍉𥍉仔衝懇	水會逐漸升高	The water will gradually rise.
17	按呢我就會當止喙焦	這樣就可以解渴	Then I can quench my thirst!

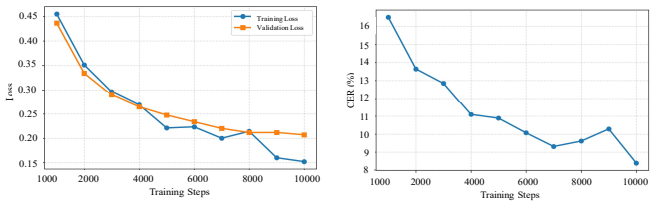
When a learner speaks in Taiwanese, the utterance is first transcribed into Chinese text by the Whisper ASR model. The recognized Chinese text is then translated into POJ representation, which serves as the input to the trained MMS-based Taiwanese TTS model for speech synthesis. To conduct a human MOS evaluation, five Taiwanese domain experts are invited to assess the quality of the synthesized speech. Since the materials were selected from a Grades 5–6 elementary-school Taiwanese textbook, highly specialized expertise was not required. However, the evaluators were expected to have sufficient proficiency in Taiwanese to understand the content and assess the synthesized speech. The evaluation scores range from 1 to 5, where 5 indicates excellent quality, and 1 denotes poor quality. The subjective scores provided by the experts are

aggregated to evaluate the overall performance of the synthesized Taiwanese speech.

V. Experimental Results

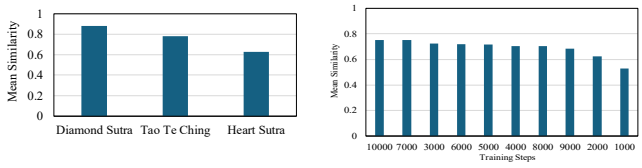
A. CER Performance Evaluation for Taiwanese ASR

This section describes the experimental evaluation conducted to assess the Taiwanese-Japanese ASR model performance. Fig. 7 shows the training and validation loss trends and the corresponding Character Error Rate (CER) across different training checkpoints, with decreasing values indicating improved recognition performance as training progresses. Fig. 8 presents the mean similarity results under different evaluation settings. The evaluation dataset consists of 612 samples from the *Heart Sutra*, 144 samples from the *Tao Te Ching*, and 121 samples from the *Diamond Sutra*, with audio recordings in both Chinese and Taiwanese. Fig. 8(a) reports the mean similarity of the three classical texts evaluated using the final Whisper-Taiwanese model (checkpoint = 10,000), while Fig. 8(b) shows the mean similarity averaged over all texts for different checkpoint models. Consistent with the decreasing loss and CER trends in Fig. 7, recognition performance improves as training progresses across checkpoints.



(a) (b)

Fig. 7. (a) Training and validation loss curves and (b) CER performance.



(a) (b)

Fig. 8. Mean similarity comparison for (a) using the final model (checkpoint = 10,000) and (b) different checkpoint models averaged over all texts.

B. Taiwanese-Japanese ASR Performance on Heart Sutra

To evaluate the performance of the fine-tuned Whisper-Taiwanese-Japanese ASR model in recognizing Japanese speech, the *Heart Sutra* is used as the test material.

TABLE IV. Mean similarity of Japanese ASR on the *Heart Sutra*

Japanese No.	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
Human Evaluation						
J1_HI	1.000	1.000	1.000	0.960	1.000	1.000
J2_HI	0.975	1.000	0.972	1.000	1.000	1.000
J3_HI	1.000	0.950	0.995	0.948	1.000	1.000
Machine Evaluation						
J1_MI	1.000	1.000	0.978	0.948	0.991	0.974
J2_MI	0.99	1.00	0.98	1.00	1.00	0.97
J3_MI	1.00	1.00	1.00	0.96	0.98	1.00

Three Japanese speakers are invited to recite the *Heart Sutra* in Japanese, and the text is divided into six segments to align with the human–machine co-learning framework and to facilitate both human and machine evaluations. For each sentence, the best result from up to three utterances is selected to compute SBERT-based similarity. Table IV shows the mean similarity scores of Japanese ASR on the *Heart Sutra* evaluated by humans and machines. It demonstrates that the fine-tuned Whisper-Taiwanese-Japanese ASR model can effectively recognize Japanese speech and accurately map it to the corresponding Chinese text.

C. MOS Fuzzy Evaluation for MMS Taiwanese TTS

To evaluate synthesized Taiwanese speech quality, texts from *The Crow Drinks Water* and *Online Shopping* are used as inputs to the MMS-based Taiwanese TTS model to generate speech outputs. A subjective listening test was conducted with five expert evaluators proficient in Taiwanese. The human-based fuzzy evaluation focused on speech naturalness, fluency, and intelligibility, using the MOS as the primary metric. In Section V.C, variables are categorial, and their corresponding linguistic labels are given in Section V.D.

Fig. 9 shows the human-evaluated MOS results of the model for *The Crow Drinks Water*. In Fig. 9, the bars show individual MOS scores from five evaluators, while the red curve denotes the average MOS. The results indicate that the proposed Meta AI MMS-based Taiwanese TTS system achieves consistently high MOS scores for both contents, demonstrating robust generalization across different contexts. Overall, the fuzzy evaluation confirms that the MMS-based Taiwanese TTS model generates natural Taiwanese speech from POJ-translated Chinese texts, with stable performance validated by human listening tests.

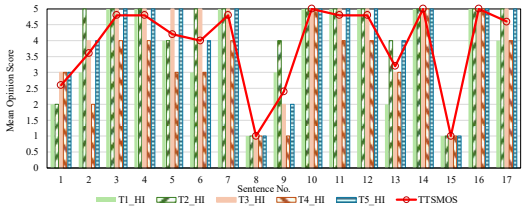


Fig. 9. Human-evaluated MOS results of MMS-based Taiwanese TTS on *The Crow Drinks Water*.

D. Quantum Fuzzy Inference Engine with Quantum Circuit

In this subsection, two datasets are employed for model training and evaluation. The first dataset (DS1) comprises 31 real human–machine co-learning samples collected from The Crow Drinks Water and Online Shopping. The second dataset (DS2) consists of 300 simulated samples generated for model training. The knowledge model is trained on DS2 using Particle Swarm Optimization (PSO) with 10 particles and 150 generations. Fig. 10(a)–(d) presents the learned membership functions corresponding to *ASRsim*, *LLMsim*, *TTSmos*, and *Performance*, respectively. Before training, the semantic accuracy is 0.774 on DS1 and 0.513 on DS2. After PSO-based learning, the semantic accuracy increases to 0.806 on DS1 and 0.907 on DS2, demonstrating the effectiveness of the proposed training strategy.

Fig. 11 illustrates the partial quantum circuit for quantum fuzzy inference on the first sample in DS1. Given *ASRsim* = 0.927, *LLMsim* = 1.0, and *TTSmos* = 2.6, both quantum and traditional fuzzy inference produce an output of 6.5 under the pre-learning model, compared with the domain expert score of 7.308. After learning, the quantum fuzzy inference output increases from 6.5 to 6.759, whereas the traditional fuzzy inference output decreases from 6.5 to 6.253. This result indicates that the learned quantum inference model achieves closer agreement with the expert assessment. However, there is still a notable difference with the expert assessment, so there is some room to improve the approach.

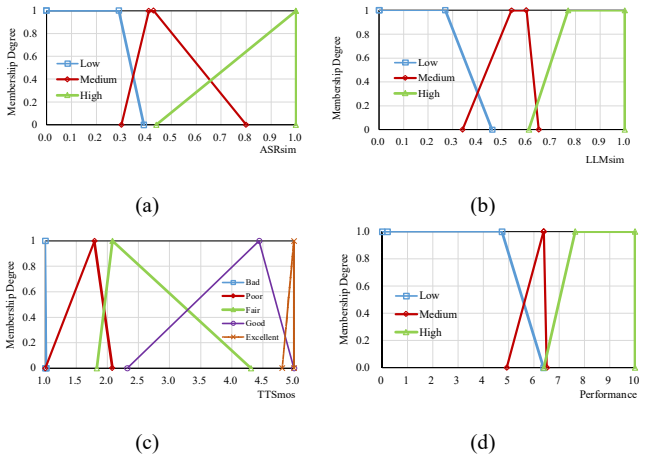


Fig. 10. Learned knowledge model based on PSO (a) *ASRsim*, (b) *LLMsim*, (c) *TTSmos*, and (d) *Performance*.

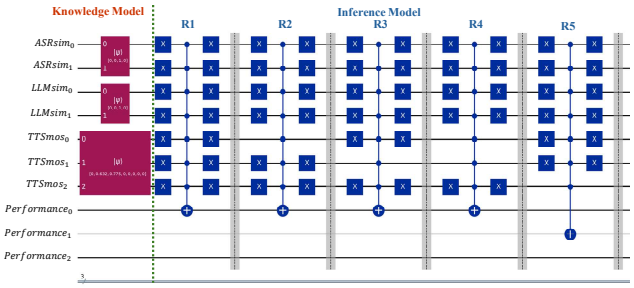


Fig. 11. Partial quantum circuit for quantum fuzzy inference.

VI. Conclusion

This paper proposes a fuzzy inference–driven knowledge modeling framework for evaluating LLM-assisted human–machine co-learning in multilingual speech-based learning environments. The framework integrates Whisper-based Taiwanese ASR, the TAIDE large language model, and an MMS-based personalized Taiwanese TTS system, with evaluation variables derived from ASR similarity, LLM translation similarity, and TTS Mean Opinion Score. Furthermore, a quantum fuzzy inference–based knowledge model is introduced to assess co-learning performance. Experiments using classical texts such as the Heart Sutra demonstrate accurate Taiwanese–Japanese speech recognition and reliable mapping to corresponding Chinese and English texts, as well as effective evaluation of MMS-based Taiwanese TTS performance. These results confirm the effectiveness and practicality of the proposed framework for human–machine co-learning evaluation. However, the ASR and TTS evaluations were conducted on a relatively small scale, so the current findings mainly demonstrate feasibility rather than full generalization to real-world speech conditions. Future work will include more diverse speakers, more natural conversational data, larger-scale human evaluation, stricter data partitioning to reduce possible data leakage, and broader applications in tourism, healthcare, and elderly companionship.

Acknowledgment

The authors would like to thank the students from NUTN, Taiwan, and OMU, Japan, who were involved in this project, and express their gratitude to ChatGPT 5.2 for its grammar checking and English proofreading of the paper.

References

[1] National Institutes of Applied Research, Trustworthy AI Dialogue Engine (TAIDE), [Online]. Available: https://huggingface.co/taide.

[2] E. Tseng, “Putting AI to work: TAIDE speaks in tongues,” Taiwan Panorama, Jan. 2026. [Online]. Available: https://is.gd/2LUPXi.

[3] C. S. Lee, M. H. Wang, C. Y. Chen, S. C. Yang, M. Reformat, N. Kubota, and A. Pourabdollah, “Integrating quantum CI and generative AI for Taiwanese/English co-learning,” *Quantum Machine Intelligence*, vol. 6, 64, pp. 1-19, 2024.

[4] P. J. Chen, K. Tran, Y. Yang, J. Du, J. Kao, Y. A. Chung, P. Tomasello, P. A. Duquenne, H. Schwenk, H. Gong, H. Inaguma, S. Popuri, C. Wang, J. Pino, W. N. Hsu, and A. Lee, “Speech-to-speech translation for a realworld unwritten language,” in A. Rogers, J. Boyd-Graber, N. Okazaki (editors), In Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 4969–4983.

[5] C. S. Lee, M. H. Wang, Y. H. Lee, C. C. Yue, S. C. Yang, Y. J. Lin, Y. Nojima, and M. Reformat, “LLM-based fuzzy agent for human-machine interaction in STEM education and Taiwanese-English co-learning application,” 2025. [Online]. Available: https://www.researchsquare.com/article/rs-6301579/v1.

[6] C. S. Lee, M. H. Wang, M. Reformat, S. H. Huang, “Human intelligence-based Metaverse for co-learning of students and smart machines,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 7695-7718, 2023.

[7] G. Acampora, R. Schiattarella, and A. Vitiello, “On the implementation of fuzzy inference engines on quantum Computers,” *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 5, pp. 1419–1433, 2023.

[8] B. H. Lu, Y. H. Lin, E. S. Lee, and T. H. Tsai, “Enhancing Taiwanese Hokkien dual translation by exploring and standardizing of four writing systems,” The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italy, May 20-25, 2024.

[9] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. Mcleavey, and I. Sutskever, “Robust speech recognition via large-scale weak supervision,” in Proceedings of the 40th International Conference on Machine Learning (ICML 2023), Hawaii, USA, 2023, pp. 28492-28518.

[10] V. Pratap, A. Tjandra, B. Shi, P. Tomasello, A. Babu, S. Kundu, A. Elkahky, Z. Ni, A. Vyas, M. Fazel-Zarandi, A. Baevski, Y. Adi, X. Zhang, W. Hsu, A. Conneau, M. Auli, “Scaling speech technology to 1,000+ languages,” *Journal of Machine Learning Research*, vol. 25, pp. 1-52, 2024.

[11] N. Reimers and I, Gurevych, “Sentence-BERT: Sentence embeddings using siamese BERT-networks,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language (EMNLP-IJCNLP 2019), Hong Kong, China, Nov. 3-7, 2019, pp. 3980-3990.

[12] C. Cheng, “Taiwan’s largest heritage language,” Global Taiwan Brief, vol. 10, no. 18, pp. 1-5, 2025.