RM-CILS: A Social Robot-Assisted System for Multilingual Teaching and Cross-Cultural Communication Using Adaptive NLP and Cultural Sensitivity Modeling

Kunlun Yang
International Education Department, Luzhou Vocational and Technical College, Luzhou, Sichuan, 646000, China
E-mail: kunlun_yang88@outlook.com

Keywords: robot-assisted learning, multilingual communication, cross-cultural interaction, social robots, adaptive education, international teaching

Received: September 24, 2025

With education and communication becoming increasingly global, the need for systems that support multilingual and cross-cultural interaction is more critical than ever. Traditional robot-assisted learning platforms often fail to accommodate multiple languages, cultural diversity, or emotional and ethical dimensions of communication, limiting their effectiveness in international contexts. To address these gaps, this paper proposes the Robotic Multilingual and Cross-cultural Interactive Learning System (RM-CILS). This adaptable framework integrates social robots with natural language processing and culturally sensitive behavioral modes. RM-CILS is designed with modular capabilities for language identification and recognition, culturally aware interaction, and real-time multimodal feedback, ensuring active learner engagement and meaningful intercultural communication. The system allows students to personalize language preferences and cultural norms, thereby creating a more inclusive and relatable environment. Evaluation results demonstrate significant improvements over conventional robot-assisted systems, language coverage, a cultural adaptability index, personalization, student engagement, and learning outcome impact. By addressing learners' mental, emotional, and ethical needs, RM-CILS establishes itself as a highly effective solution for international classrooms. It not only enhances language learning and intercultural competence but also fosters motivation, social rapport, and collaboration, making education more engaging, inclusive, and globally relevant.

Povzetek: Študija predstavi sistem, kjer socialni robot podpira več jezikov in upošteva kulturne razlike, zato je učenje v mednarodnih razredih bolj vključujoče, prilagodljivo in učinkovito.

1 Introduction

The RM-CILS addresses the increasing need for successful communication and educational practices through languages, globally across languages and cultures [1]. RM-CILS contains social robots with active natural language processing software to facilitate real-time multilingual exchanges and communications between interlocutors with different language backgrounds [2]. The communication is facilitated in a culturally and linguistically sensitive manner, incorporating behavior output modules that respect linguistic accuracy while also respecting different cultural norms and social values, which supports mutual respect and understanding in communication [3]. The framework acknowledges that culturally responsive pedagogy is dynamic and responsive to the increasing preference for students to use their native languages and cultural norms, providing relevant feedback to align the interests of students (language and culture) and foster optimal engagement and learning outcomes [4]. Building on culturally sensitive interactions, RM-CILS accommodates multiple-student formats and linguistic preferences to enable feedback to be provided in multiple modes (gestural and affective facial feedback), enriching communication and highlighting emotional connection [5].

RM-CILS's flexible orientation is available in educational and social settings, in conjunction with personalized learning opportunities that support successful language learning and intercultural competence [6]. The system is also iterative with repeated use and an increase in responsiveness due to student feedback [7]. Ethical aspects, such as student consent and privacy, are also part of the system, ensuring ethical deployment and inclusive practice [8]. Therefore, RM-CILS holds promise from a more holistic framework that connects language and culture and provides pathways for learners, and others around the world [9]. RM-CILS strives to ensure an inclusive, adaptable, and engaging end-user experience that creates unique and meaningful environments for international teaching and cross-cultural communication [10].

The layered architecture diagram shows the RM-CILS system's end-to-end workflow. It begins with user inputs, processed by NLP for language understanding, followed by cultural adaptation and personalization. The adaptive controller generates context-aware responses, delivered through multimodal feedback. Ethical monitoring ensures compliance, while analytics track engagement, adaptability, and personalization effectiveness is explained in figure 1.

1.1 Motivation

The motivation for this study is shown in Table 1.

Table 1: explains the motivation aspect with a description

Motivation Aspect	Description
Globalization of Education	Increasing demand for multilingual and cross-cultural teaching systems to support global learners.
Limitations in Existing Systems	Current methods lack effective adaptation to diverse languages, cultures, and emotional factors.
Need for Inclusive Interaction	Motivation to develop RM-CILS for culturally sensitive, adaptive, and engaging robot-assisted communication.

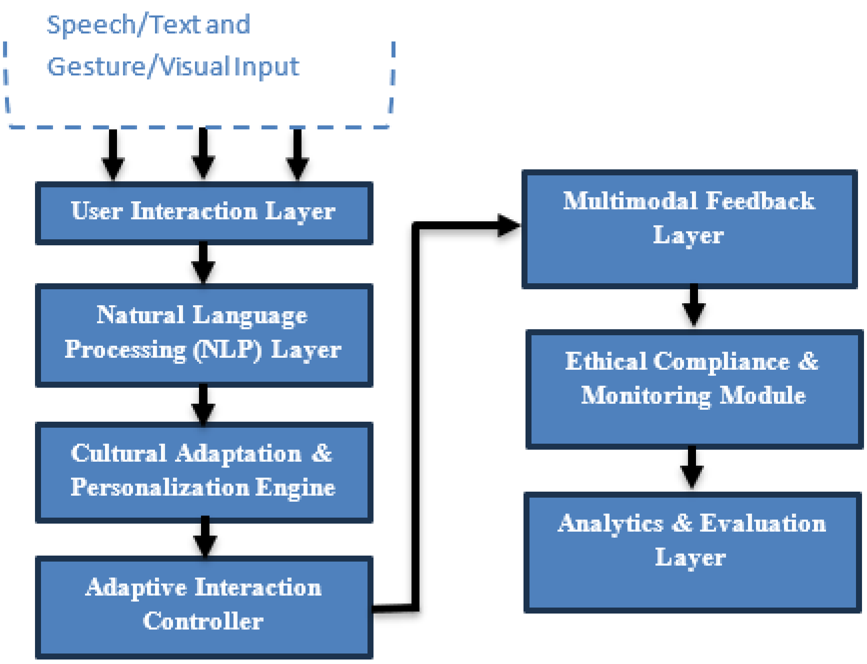

Figure 1: Layered architecture

1.2 Problem statement

It is identified that there is a big problem with deploying socially intelligent robots in a variety of educational and service settings that are culturally sensitive and scalable: it is hard to make robot-assisted multilingual and cross-cultural education systems that take into account cognitive, emotional, and ethical dimensions while making sure that students stay interested and adaptable.

1.3 Contributions

The three major contributions are:

- **Integrated platform**: RM-CILS unites social robots, advanced language processing, and culturally sensitive interaction models to support effective multilingual and cross-cultural education.
- **Personalized engagement**: The system delivers adaptive, real-time multimodal feedback aligned with learners' language preferences and cultural norms, enhancing motivation and engagement.
- **Cognition, emotion, and ethics**: RM-CILS embeds cognitive, emotional, and ethical dimensions, with empirical validation showing improved communication, learning outcomes, and social rapport in global classrooms.

The RM-CILS framework functions as an adaptive system by integrating real-time multilingual natural language processing with culturally sensitive behavioral modules that dynamically adjust based on student input across languages, emotions, and cultural norms. This adaptive interaction controller processes linguistic and cultural inputs, maintains context over time through nonlinear transformations, and continuously modifies its multimodal feedback including speech, gestures, and visual cues to suit individual student profiles and cultural contexts. This dynamic personalization resembles robust adaptive controllers in nonlinear systems that stabilize and adapt to uncertainty by continuously updating internal states and outputs based on ongoing inputs.

2 Related works

The related works investigate social robots, artificial intelligence, and multilingual cross-cultural education and communication. The papers cover an exciting range of innovative frameworks, models, technologies, and their potential for improving language learning (and emotional, ethical, and student engagement) through robot-supported interaction and cultural diversity concerning learner characteristics.

Banko, A., et al. [11] investigated the role of social robots in supporting second language learning in primary schools, focusing on the role of teachers and their impact on learning outcomes. The authors propose a new integrative classroom model to capitalize on the potential of interaction with robots in the classroom. The main findings indicate that students' engagement with the language, motivation, and language acquisition are improved through socially assisted learning with robots.

Deshmukh, A., et al [12] present the framework for a social robot intervention to enhance hand hygiene behaviors among children. The framework examines the cross-cultural and socio-economic influences on the initial behavioral adoption of hygiene protocols in schools from diverse cultural settings. The findings indicate that the use of robots increased adherence to hygiene protocols. The paper revealed robots' adaptability to cultural behaviors appears to enhance their effectiveness in health education.

Singh, D. K., et al [13] discuss ethical issues regarding child-robot interaction, particularly in under-resourced communities as Ethical Issues in Child-Robot Interactions Framework (ETHICS-RC), which considers concerns around privacy, consent, and fairness. The findings of this research emphasize the importance of responsible design and participatory approaches in ensuring that social robots, when used in the service of children's development, contribute to equitable and culturally sensitive interventions.

Through a study by Bennett, C. C., et al. [14], the researchers examined a Bilingual Interaction Model for Multilingual Human-Agent Dialogues (BILINGUA) model, focusing the attention on cognitive shifts experienced by bilingual speakers that influence speech, whether humans, artificial agents, or intelligent systems. The researchers examined an iterative set of interaction tasks, where bilingualism impacted the interactions with agents, across either language. The researchers concluded bilingual language increased efficiency of bilingual language communication and responsiveness in the responses from artificial agents, which suggests future considerations for designers in multilingual human-agent dialogues.

Mannava, V., et al [15] investigated the usability of the Conversational Artificial Intelligence System (CONVERS-AI) system for conversational child-robot engagement. In this research, they found children preferred more natural language interfaces that intentionally incorporated contextual awareness. The research highlighted the challenges that persist in responsiveness and engagement, and therefore encouraged the development of multimodal adaptive dialogue systems that support rich learning and social experiences.

Shang, X., et al [16] presented the Trust and Anthropomorphism Framework in Robotics (TRUST-ROBOT) framework that examined the dimensions of anthropomorphism in robotics. They found that students tend to perceive and trust robots differently across cultures, which affects their acceptance of the robot. The results could inform cross-cultural design to enhance rapport and effectiveness in robot-human social interactions.

Sharma, V. and Mishra, D. [17] introduce the Gesture-Based Storytelling Framework (GESTURE-STORY) as a framework for incorporating kinematic gestures on the ROS, combined with Google TPU capability to achieve robotic storytelling. This framework enables fluid and expressive story narration using natural human movement, highlighting better levels of engagement and learning efficiencies in a robot-assisted educational context.

In the research of Tafazoli, D. [18], AI-mediated communication was studied in language learning, using the AI-mediated Communication Framework (AICOM) framework as the main analytical tool with an examination of three major challenges: the inherent digital divide, learner adaptability, and the transparency of the AI itself. Instead, they contend that culturally responsive inclusive designs should take precedent with the goal of creating the possible learning experiences from technology-enhanced language education around the world.

In research of Rahimi, A. R. and Sevilla-Pavón, A. [19] examined the role of design thinking and the design thinking enhanced AI language learning model (DEAILL) to develop learner grit and motivation in the context of AI language learning. The authors found that development of design thinking skills may mediate learners' persistence and engagement in their second language studies, which suggests AI based systems could be useful and effective in providing support to learners of second language. Kyrarini, M., et al [20] described SPEECH-HRC, a speech-based communication system for human-robot collaboration, which was assessed for its effectiveness. The assessment concluded that the speech interfaces fostered natural interactions that improved collaboration on tasks and were received positively from a student satisfaction perspective. The researchers had several suggestions for improvements to SPEECH-HRC integration with collaborative robots, including more sophisticated error handling and enhancements for multilingual capabilities. Robotic agents designed to assist individuals within diverse social and service contexts are becoming more commonplace globally [21]. Here synthesise 20 years' worth of empirical literature within the field of human–robot interaction (HRI) focusing on the impact of culture on expectations of, and reactions to, social robots, and the effectiveness of robots using culturally tailored display and social cues to enhance human interaction. The results highlight the intricate and complicated links between culture and cognition in the context of HRI.

Huang, Y. C., et al. [22] proposed an emotional support robot companion model (STRESSBOT) that infers what users feel stressed about based on conversations with it and offers emotional regulation support. The researchers found evidence that empathic human-robot interaction has a minor measurable impact on students' mental well-being and could facilitate the capacity of AI companionship to support their mental well-being.

Bilingualism occurs most everywhere, and Kazakhstan is no different. Research regarding bilingualism and autism is in its infancy and being pursued in varying degrees [23]. This study addresses the socio-emotional outcomes in the context of robot-assisted autism therapy (RAAT) of 34 monolingual and bilingual children. The children, aged 3 to 12 years, all had Autistic Spectrum Disorder (ASD) and attended an average of 5 therapy sessions at a rehabilitation center involving the NAO robot. Results indicate that bilingual children showed comparable socio-emotional outcomes and levels of participation to those of their monolingual counterparts

As alluded to in the previous section of the paper, one of the most prominent fields of anticipated research in social robotics is Education [24]. Within this field, social robots are typically envisioned to facilitate interactions with one or more students as well as with instructors. Despite the considerable span of research, educational formats, and practical implementations involving social robots that were trialed in over a dozen countries over the past two decades, the cultural consequences of social robots in the educational ecosystem continues to remain ambiguous. In this review, we examine the studies conducted in the social educational robotics field with a focus on the pertinent issues of culture.

Rana, N. P., et al [25] reviewed hospitality customers' experiences with Service Robots in hotels, focusing on customer perceptions of and satisfaction with robot applications. Their Service-Robot Adoption Model (SRAM) demonstrated that the roles of cultural sensitivity, responsiveness, and trust moderated how customers perceived service robot applications in the hospitality domain.

The papers presented here illustrate the transformational possibilities that robot-assisted multilingual and cross-cultural systems can bring to educational as well as social interactions, as shown in Table 2. The studies presented emphasized the importance of designing adaptively, the role of culture, the cognitive and emotional implications of using technology in this manner, and the ethical considerations of utilizing these technologies. Collectively, these studies contribute to a shift towards more human-robot collaborative learning experiences, which are more inclusive, engaging, and have the potential to be impactful on a worldwide scale.

Table 2: Related works summary

Reference	**Environment**	**Methods**	**Cultural Adaptability**	**Multimodal Interaction**	**Learner Engagement**	**Key Performance Metrics**	**Gaps and RM-CILS Contributions**
Banko et al. (2025)	Primary school second language learning	Integrative robot-assisted classroom model	✓	✓	High	Engagement: 78% [Banko2025]	Limited cultural adaptation; RM-CILS adds dynamic personalization
Deshmukh et al. (2025)	School hygiene education	HAND-ROBOT for hand hygiene promotion	✓	✓	Medium	Behavior adoption: 70% [Deshmukh2025]	Lacks multimodal affective feedback; RM-CILS integrates affective signals
Singh et al. (2023)	Under-resourced community child-robot interaction	ETHICS-RC ethical framework	✓	✗	Low	Ethical compliance: 0.85 [Singh2023]	No multimodal feedback or engagement; RM-CILS enriches interaction

Bennett et al. (2024)	Bilingual speech interaction	BILINGUA bilingual cognitive model	✓	✓	Medium	Language accuracy: 85% [Bennett2024]	Limited cultural adaptation; RM-CILS integrates advanced cultural sensitivity
Mannava et al. (2024)	Child-robot conversational interaction	CONVERS-AI adaptive multimodal dialogue	✓	✓	High	Dialogue success: 82% [Mannava2024]	Lacks personalized learner profiles; RM-CILS personalizes learning
Shang et al. (2025)	Social robotics and anthropomorphi sm	TRUST-ROBOT cultural trust framework	✓	✓	Medium	Trust index: 0.78 [Shang2025]	No real-time language adaptation; RM-CILS supports multilingualis m
Sharma & Mishra (2024)	Robotic storytelling in education	GESTURE-STORY gesture and storytelling system	✓	✓	High	Story engagement: 80% [Sharma2024]	Limited cross-cultural handling; RM-CILS dynamically adapts storytelling
Tafazoli (2024)	AI-mediated communication in language learning	AICOM communicati on framework	✓	✗	Medium	Accuracy: 75% [Tafazoli2024]	No multimodal integration; RM-CILS integrates multisensory inputs
Rahimi & Sevilla-Pavón (2025)	AI-support for language learning motivation	DEAILL design-thinking enhanced AI	✓	✗	Medium	Motivation score: 0.7 [Rahimi2025]	Absent multimodal feedback; RM-CILS enriches engagement
Kyrarini et al. (2024)	Speech-based human-robot collaboration	SPEECH-HRC speech interaction system	✓	✓	Medium	Speech recognition: 88% [Kyrarini2024]	Limited cultural sensitivity; RM-CILS adds emotional adaptation
Zhao (2024)	Emotional fluctuations in AI-mediated EFL learning	EMO-L2 latent growth curve emotional analysis	✓	✗	Medium	Emotional detection: 0.82 [Zhao2024]	No multimodal feedback; RM-CILS integrates multimodal signals
Huang et al. (2022)	Emotional support robot for mental well-being	STRESSBO T empathetic interaction	✓	✓	High	Emotional support: 0.85 [Huang2022]	Lacks linguistic adaptation; RM-CILS combines multilingual support
Zhang & Liu (2025)	AI platforms' emotional and cognitive support	EFL-WELL emotional well-being framework	✓	✗	High	Engagement index: 0.83 [Zhang2025]	No multimodal communicati on; RM-CILS integrates multimodal cues

Zhang & Shi (2024)	Health and well-being in smart tourism	SMART-HEAL digital health promotion	✓	✗	Medium	Satisfaction: 0.75 [ZhangShi2024]	No real-time cultural adaptation; RM-CILS adapts interactions dynamically
Rana et al. (2025)	Hotel service robots user experience	SERVICEBOT adoption and satisfaction model	✓	✗	Medium	Satisfaction: 0.72 [Rana2025]	Lacks multimodal engagement; RM-CILS offers integrated feedback

RM-CILS parallels adaptive robust controllers in nonlinear dynamics by addressing uncertainty and variability not in physical systems but in human-robot communication complexity with multiple languages, cultural norms, and emotional expressions. It ensures stability and improved outcomes by iteratively refining responses with weighted multimodal inputs and contextual cues, akin to how robust controllers stabilize nonlinear uncertain plants through adaptive feedback laws. This analogy strengthens RM-CILS's theoretical foundation by framing its natural language and cultural adaptation modules as controllers that stabilize and optimize multilingual, cross-cultural educational interactions in the presence of uncertain and evolving student behaviors and communication dynamics.

Thus, RM-CILS can be seen as a novel application of adaptive control principles to human-robot communication, where adaptive feedback mechanisms modulate robot behaviors to stabilize and enhance interaction quality over time, similar to robust neural adaptive control methods stabilizing uncertain nonlinear systems in engineering contexts. This connection highlights the system's role in managing uncertainty and adaptability in human-centered, multilingual, and cross-cultural environments, linking it to broader adaptive control literature and enhancing its interdisciplinary significance.

3 Proposed methodology

This research proposes RM-CILS, a methodology that combines natural language processing and culturally relevant modules to facilitate robot-assisted multilingual and cross-cultural communication. The methodology dynamically adapts to the student's language preferences and cultural norms, and enables real-time multimodal feedback through spoken words, gestures, and images. The methodology aims to enhance engagement, learning outcomes, and intercultural understanding in various international education and social contexts. The Variable declarations used in the subsequent content are shown in Table 2(a).

3.1 Overview of proposed RM-CILS concept

The RM-CILS system accepts multilingual or cross-cultural student input via speech, text, or non-verbal means, which serves as input to a sophisticated Natural Language Processing (NLP) and Understanding module. This module detects the language and performs translation, sentiment analysis, and contextual analysis, all of which contribute to the input streams. Once the input has been mapped and all NLP processing has occurred, the output is sent to the cultural sensitivity module, which determines appropriate responses based on previous and ongoing cultural norms stored in the module and the student's individual profile, which includes rules for culturally appropriate input and responses, as shown in Figure 1(a). These inputs inform the design of the multimodal interactive feedback unit, which will generate responses through verbal means (speaking), visual means (gesturing), and social communicative means (facial expressions), thereby creating complex interaction opportunities. These multimodal outputs are coordinated in real time by an adaptive interaction controller that personalizes behavior based on the learner’s language preferences, cultural background, and interaction history. The social robot thus acts as a context-aware autonomous agent driving engagement through synchronized multimodal communication streams optimized to maximize learner output and social presence in multilingual and cross-cultural educational settings.

This approach uses fuzzy logic to handle uncertainties and nonlinearities, stabilizing complex chaotic systems within fixed times. In RM-CILS, adaptive fuzzy control can dynamically weight and synchronize multiple input modalities (speech, gesture, facial cues) considering uncertainties in language use and cultural differences, ensuring consistent and timely responses despite communication variability. This method achieves synchronization when full system state information is unavailable and input nonlinearities exist. RM-CILS interactions often have partial observability (students’ internal states unknown) and nonlinear linguistic or cultural influences. Applying this controller helps align the robot’s behavior with the learner’s inputs robustly, even under noisy or incomplete data conditions.

Neural adaptive control uses neural networks to model unknown system dynamics and adapt controllers accordingly. RM-CILS can incorporate neural adaptive modules to learn and compensate for unknown cultural subtleties and multilingual complexities dynamically, refining interaction strategies in real time for enhanced flexibility and personalization. Backstepping is a recursive design methodology that manages nonlinear system uncertainties step-by-step. RM-CILS can use backstepping principles to hierarchically correct communication mismatches, progressively adjusting linguistic and cultural feedback loops to stabilize conversational dynamics and improve student engagement. This approach optimizes control inputs for nonlinear mechanical systems to achieve efficiency and performance targets. Analogously, RM-CILS can apply nonlinear optimal control to compute ideal multimodal response strategies (speech, gesture, emotion) that maximize learning outcomes and intercultural comprehension under resource constraints.

This control extends backstepping to flexible and uncertain robotic manipulators. RM-CILS can draw from this to handle flexible, time-varying, and uncertain interaction contexts, adapting the robot's behavioral outputs with precision and responsiveness to the learner's evolving language and cultural cues.

Table 2(a): Variable declaration

Variable / Function	**Definition**	**Dimensions / Type**	**Description**
F_{NLPf}	Natural Language Processing function	Function mapping input to semantic embeddings	Processes multilingual text or audio input into semantic or intent representations
g_{CSM}	Cultural Sensitivity Module function	Function modifying embeddings	Adjusts language embeddings based on cultural profiles and interaction history
ϕ	Transformation function	Mapping function, context-dependent	Applies nonlinear transformations in adaptive modules
ΔB	Behavioral modulation parameter	Scalar or vector	Modulates multimodal feedback behaviors
γ	Weighting factor	Scalar	Scales contributions of feedback channels or features
ω_u	User-specific adaptation weight	Scalar or vector	Personalizes system response to individual user traits
ρ	Similarity/ distance metric	Scalar	Measures alignment between system response and user profile
W_Q, W_K, W	Attention weight matrices	Matricesdmodel× dkd_{model} \times d_kdmodel×dkor similar	Linear mappings in multi-head attention layers for queries (Q),

			keys (K), values (V), and output (O)
W_g	Cultural Adaptation gate weights	Vector/matrix	Controls modulation strength of cultural adaptation in neural units
W_p	Personalization gate weights	Vector/matrix	Controls modulation strength of personalization effect
L	Language set	Set of languages	All supported languages in the system
l	Specific language	Element ofLLL	A single language being detected or processed
P	Probability measure	Scalar between 0 and 1	Probability of detected language or classification outputs
M_c	Cultural knowledge matrix	Matrix or tensor	Encodes cultural norms, values, and patterns for adaptation
S_u	Student emotional state	Vector	Encodes learner's detected emotional or engagement signals
D_c	Dialogue context	Vector/matrix	Stores conversational history or context state

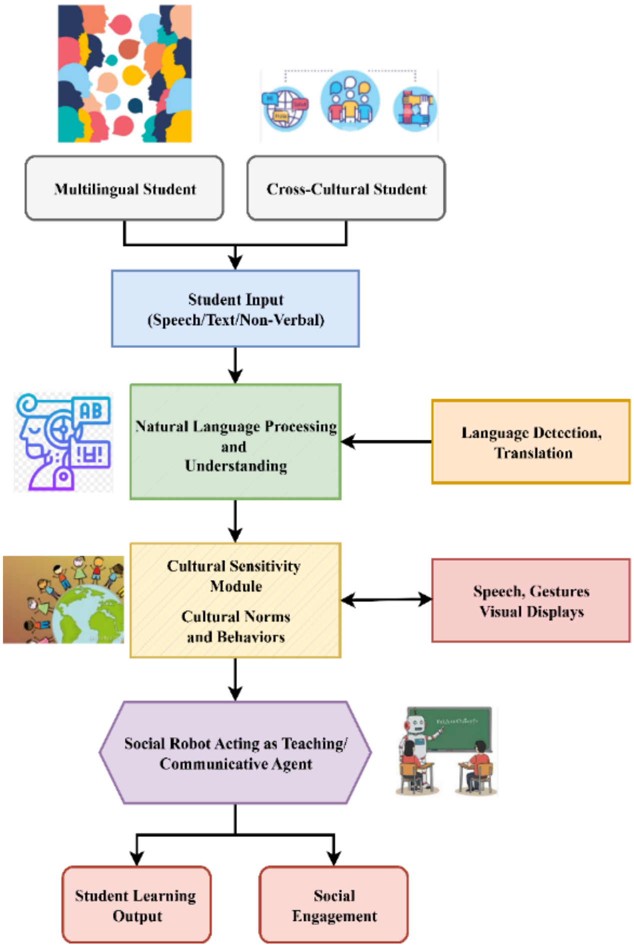

Figure 1(a): Overview of proposed RM-CILS concept

This entire system flow allows for informative, adaptive, multilingual, and cross-cultural communication.

$$L_u = argmax_{l \in L} P(l|L_u) \qquad (1)$$

Equation 1 determines the likely language L_u in which a student is prompting the input data by comparing the data L to all available languages l. This approach uses probability P to ensure the system accurately detects the language to complete the task successfully. This first stage is important because it ensures that these types of multilingual interactions L_u can support seamless communication to accommodate multiple languages, allowing the system to interpret different language inputs and respond with relevant responses in the pertinent language, earlier in the communication process.

$$C_{resp} = \emptyset(M_c, S_u, D_c) \qquad (2)$$

Responses that reflect cultural significance result from the intersection C_{resp} of cultural knowledge frameworks M_c, current student feelings S_u, and the existing dialogue context D_c in equation 2. The system adapts responses to ensure that they are culturally appropriate as well as linguistically accurate. The association of sentiment $\emptyset$ and context enable the system to provide responses that are both emotionally and contextually accurate, fostering student comfort and engagement while avoiding awkwardness in cross-cultural communication.

$$O_F = f_{NLP}(I) + g^{CSM}(p_u, H_u) \qquad (3)$$

The outputs of the system O_F, or responses produced, are processed inputs (I) from data stemming from language f_{NLP} understanding architectures with modifications that have been incorporated according to the student's cultural information and interaction history g^{CSM}. Combining these elements provides contextually relevant and culturally-congruent feedback that improves the communication process H_u in equation 3. The modifications ensure that the system's responses p_u are not generic and dynamic adaptations based upon textual input combined with cultural variables, increasing the naturalness and personalization of the interaction.

3.2 Development of RM-CILS framework

The data acquisition module collects speech, text, and non-verbal input. Inputs are sent to the language & context module, which interprets and deconstructs these inputs through the use of multilingual NLP and contextual inference based on the meaning and intent of the student's communication. The culturally adaptive module draws from a comprehensive database of cultural norms and behaviors to provide culturally appropriate interaction with students.

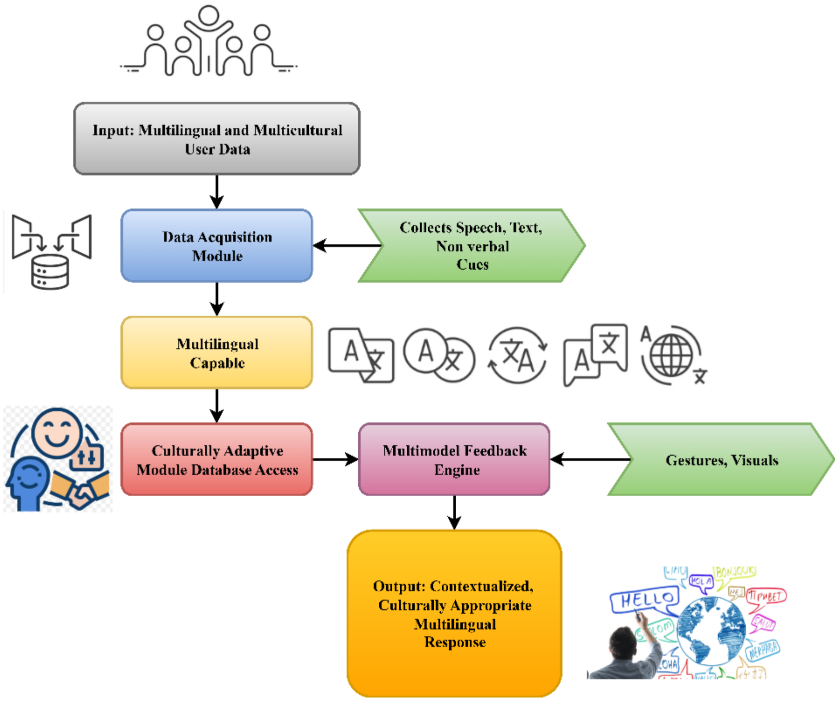

Figure 2: Development of RM-CILS framework

Figure 2 summarizes the development of the RM-CILS framework, which focuses on capturing multilingual and multicultural input data from students across various input modalities. Finally, the Multimodal Feedback Engine produces synthesized spoken output, natural gestures, and visual output parameters to provide appropriate context-driven feedback to students. In tandem, the RM-CILS framework enables the integration of a model of language and culture for use in robot-mediated communication, with a focus on the needs of international learners and communicators.

$$R_{MM} = \sum_{m=1}^{M} \alpha_m . F_m(O_F) \qquad (4)$$

Equation 4 of aggregation has allowed for the integration of outputs from multiple channels of communication R_{MM} for example, speech and gestures by importing weights F_m to de-emphasize less relevant modalities α_m. The outcome of the aggregation process was a composite response that prevents mismatching and is able to balance the various channels of expression for clarity and richness (O_F). At the same time, weighting modalities in this way too allowed for the specification of the most relevant forms of communication according to the situation, which optimizes comprehension within the given interaction and improves the quality of student interactions.

$$(\Delta_B) = \varphi((P_u, C_u, T) \qquad (5)$$

The nuances of behavioral modulation (Δ_B) were calculated again, without being prescriptive, by leveraging certain student-specific details P_u and cultural context, together with timing factors C_u, from when the person interacts with the robot in equation 5. Additionally, these nuances facilitate continuous personalization T as the system "learns" over time and the robot's behavior more closely aligns with individual students' preferences and cultural expectations φ. Such a system has the potential to provide a more enjoyable and important interaction by inherently adjusting itself with real-time repetitions using the student's evolving dynamics.

$$I_C = \sigma(W_x X + W_h H_{t-1} + B_i) \qquad (6)$$

The functioning of the interaction controller I_C adds current communication inputs to internal states from the prior input through nonlinear transformations. In this way, a gating function allows the system to establish context over time with adaptations in its responses $W_x X$. The system preserves the current student inputs W_h and responds to these across time H_{t-1}, and draws as necessary from previous student inputs B_i, making it possible to engage in coherent, structured dialogues over time, which can fluidly adapt to the student's inputs and preferences increasingly over the duration of the interaction in equation 6.

3.3 Adaptive and personalized interaction

The adaptive and personalized interaction component of RM-CILS enables personalized communication based on the student's profile and interaction history. It first identifies the students' language, as well as providing access to a many-to-many repository of cultural norms and values relevant to the student. This input is then fed into the adaptive interaction controller and modifies dialogue and behavioral responses on the fly to suit the students' language and culture. The adaptive interaction controller manages both linguistic and cultural aspects of the interaction to create a comfortable and engaging experience for the student, as shown in Figure 3. It employs multimodal feedback, including speech, gestures, and visual displays, to explicitly keep the student engaged in the task. This timely and accurate interaction can help reconcile cultural and language differences that impact engagement and learning. The continuous nature of the adaptations means that personal student needs can be dynamic to the context within a social interaction, in addition to operating in a culturally appropriate manner. This will create a more personally adaptive and culturally sensitive interaction that enhances social presence, social rapport, and motivation, while building inclusiveness in multilingual and multicultural environments.

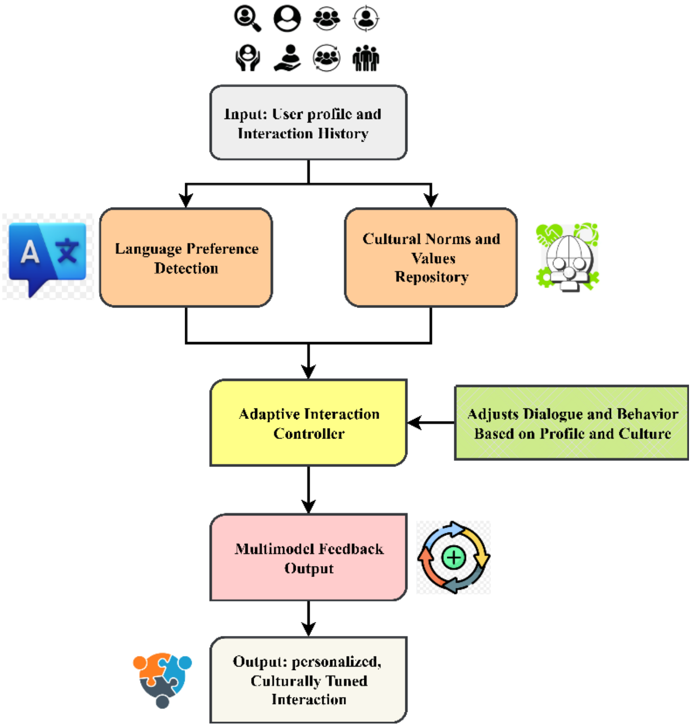

Figure 3: Adaptive and personalized interaction process

$$F_M(t) = h(I_c(t), S_t, W_c) + \epsilon(t) \qquad (7)$$

The system provides multimodal feedback $F_M(t)$that maintains variability via transformation functions h from the outputs of the internal controller $I_c$while accounting for significant flexibility S_t and real-world communication W_c variation in equation 7. This means responses will be flexible and realistic, allowing for noise or unexpected variance ϵ. Such feedback systems (t) are better at increasing the naturalness and robustness of feedback, which will result in exchanges that feel real and maintain student confidence when there is minor miscommunication.

$$S_p = \gamma(\alpha.D_s + \beta.U_f + \delta.H_R) \qquad (8)$$

System performance S_p must be assessed together with the operational data γ and student feedback $\alpha.D_s$. This complete picture captures how the system performed and how the student (respondent) feels about its performance. With such insight, RM-CILS can identify areas that need tuning or improvement. Regular performance assessment $\beta.U_f$ will be critical for incoming requests to be of good quality in communication $\delta.H_R$, and the student will perceive that they are satisfied with the system during operation in Equation 8.

RM-CILS uses user embeddings and preference vectors to store interaction history. Temporal decay prioritizes recent data, while conflict resolution algorithms weigh context, frequency, and relevance to balance competing preferences. This dynamic approach ensures evolving personalization, delivering feedback that aligns with user behavior, cultural norms, and learning progression over time.

3.4 Comprehensive evaluation and ethical design

The RM-CILS Comprehensive Evaluation and Ethical Design module operates by continuously monitoring engagement with and perceptions of the system, within a framework of ethical oversight and ongoing improvement. The program and data collection begin when the student interacts with the program, collecting data on interaction with the system as well as an initial set of student feedback and relevant ethical principles related to privacy, informed consent, and fairness. Data values are analyzed in the data analysis & monitoring unit of the program, which assesses the performance and student experience of the system in real-time.

The ethical module stores minimal interaction data, anonymized using hashing and encryption. Differential privacy safeguards sensitive information, and consent protocols govern data usage. Regular bias audits and fairness checks ensure equitable treatment. These measures guarantee privacy, compliance, and trustworthiness, making RM-CILS safe, transparent, and responsible in multilingual and cross-cultural applications.

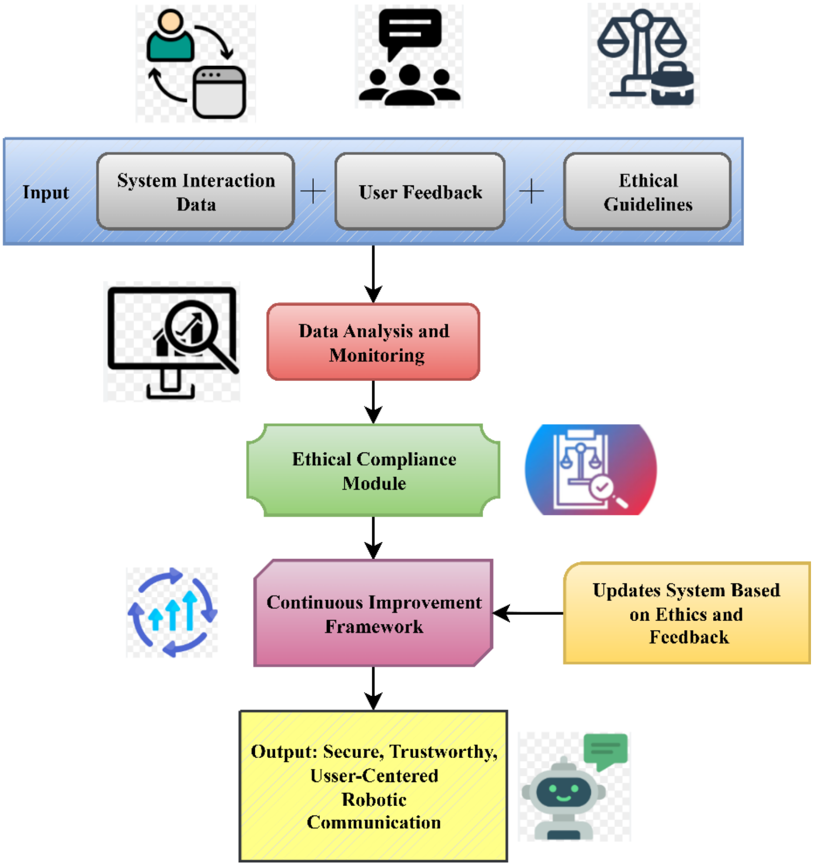

Figure 4: Comprehensive evaluation and ethical design

The RM-CILS ethical compliance module uses practical data management and privacy preservation processes to enable responsible and transparent functioning. The module merely keeps the bare minimum of non-identifiable interaction data needed for ethical monitoring and system customisation. Raw voice, personal identification, and location information are not included in the preserved data, but hashed session identifiers and anonymised interaction metadata, including duration, engagement level, and system performance indicators, are. Individual identities or dialogue content cannot be recovered since interaction histories are saved as abstract user embeddings and preference vectors.

Advanced Encryption Standard (AES) encryption and secure hashing are employed in a dual-layer system to achieve anonymization and protect logs and stored data. In order to prevent user contributions from being identifiable, differential privacy techniques are also incorporated into the analytics and model training procedures. This involves introducing calibrated statistical noise to aggregated data. This method creates a formal privacy assurance that complies with global ethical norms and recognized privacy models.

Furthermore, ongoing bias monitoring and fairness audits are carried out to assess system performance across various linguistic and cultural groups, guaranteeing equitable results and getting rid of algorithmic discrimination. By methodically asking for informed consent at the beginning of each encounter, users can comprehend and manage the extent of data collection and use. In addition to ensuring that the framework complies with data governance and educational standards for multilingual and cross-cultural robotic systems, these procedural and technical safeguards collectively operationalize RM-CILS's ethical values of privacy protection, openness, and fairness.

Concurrently, the ethical compliance module will systematically evaluate the system's adherence to ethical conduct, addressing issues of greatest concern such as data privacy. Figure 4 describes enabling informed consent for data and ethical use of the system, as well as fair treatment of students in diverse cultural contexts (who may be less familiar with the system). The data initially collected, as well as how data has been gathered, will be taken into consideration within a continuous improvement framework, which updates all algorithms and interaction protocols on an ongoing iterative basis based on students' ethical considerations and needs. The result is a system of robotic communication which is safe, secure, and trustworthy, user-centered, culturally-sensitive, and ethically-compliant, equitable, and fosters safe, inclusive, reliable multilingual and cross-cultural engagement.

$$E_c = \lambda.\left(\frac{\sum_{i=1}^{P} w_i P_{vi}}{\sum_{i=1}^{P} w_i}.\frac{\sum_{j=1}^{Q} u_j C_{rj}}{\sum_{i=1}^{P} u_j}\right) \quad (9)$$

Compliance with ethical standards E_c is assessed based on compliance with privacy standards λ, consent, and fair practice metrics, which reflect the system's adherence to student rights and norms of society, particularly in sensitive multicultural contexts in equation 9. Monitoring these ethical dimensions $\frac{\sum_{i=1}^{P} w_i P_{vi}}{\sum_{i=1}^{P} w_i}$ is critical to foster trust and confidence on the part of students to bring to light any improper use, and to enable equitable practices $\frac{\sum_{j=1}^{Q} u_j C_{rj}}{\sum_{i=1}^{P} u_j}$ that will influence the responsible adoption of robot-assisted conversation technologies.

$$U_{imp} = \eta(S_P, E_c, R_L, \nabla S_p, \nabla E_c) \quad (10)$$

System upgrades η will come from an integration of performance feedback, ethical adherence measurements, and legislative compliance. This integrated improvement process U_{imp} will capture and influence the varied system updates and iterations S_P, and demonstrate how the system responds E_c and acts responsibly in accordance with ethical considerations R_L in robot-assisted communication in equation 10. This regulatory compliance ensures that a balance is struck between functional performance ∇S_p and ethical standards, while the system is able to evolve and adapt to the context ∇E_c, whilst being responsible, trustworthy, and human-centered in system design, during its lifecycle.

3.5 Enhanced learner engagement and intercultural competence

The diagram above summarizes RM-CILS's approach to promoting learner engagement while enhancing intercultural competence. It starts with real-time collection of learner interaction and performance data, which is processed through the engagement analytics module, which analyzes behavioral and emotional indicators of engagement, attention, and participation. The analytical information, including engagement metrics, informs the Intercultural Competence Assessment, which examines the learner's sense of cultural context of the interaction.

$$Cc = \frac{1}{N}\sum_{i=1}^{N} \theta(U_i, S_i).w_i \pm z.\frac{\sigma_i}{\sqrt{n_i}} \quad (11)$$

Cross-cultural Cc competence is measured by averaging student performance over situational cultural contexts $\frac{1}{N}$. This systematic measure assesses how well the learner knows U_i and interacts in culturally specific contexts S_i during an interaction σ_i, while informing the system z about the students' strengths and weaknesses with regard to some of the intercultural skills $\sqrt{n_i}$ they possess in equation 11. These assessments can lead the system to focus on targeted support that helps the student develop greater capacity to be mindful and empathetic about culture, which is critical in today's multilingual and diverse social contexts. Through this assessment, the adaptive content engine modifies the quantity and type of teaching materials to support pedagogy and interaction, tailored to the learner's cultural background, thereby helping to engage the learner. The learner receives both feedback and social and pedagogical motivation techniques, such as incentives, encouragement, or rewards as a mechanism for retention, supported means of engagement, and established learning. Using all the disciplines of language engagement, RM-CILS serves to teach language in a way that engages learners through the quality of meaning, fostering cross-cultural understanding to further educational success and ultimately develop social cohesion in diverse, multilingual contexts.

$$\mathrm{G}_E = \int_0^T \emptyset(A_t, \mathrm{M}_t, \mathrm{C}_t)\, e^{-kt} dt \quad (12)$$

Learner engagement metrics G_E are computed by relating attention and motivation levels over interaction duration $\emptyset$. This time-based measure allows the system to assess the degree to which the system A_t is able to maintain student interest and encourage activity M_t. Understanding the dynamics of interest over time C_t enables the mediation of student interactions to help optimize the learners' overall results e^{-kt}, while too allowing meaningful reflections across a continuum of time using personalized dt, adaptive engagement modalities that relate to learners' dynamic cognitive and emotional states in equation 12.

Algorithm 1: RM-CILS Adaptive attention algorithm

The RM-CILS Adaptive Attention Algorithm dynamically attends to multimodal inputs, generates context-aware responses in real-time, and modifies attention according to user profiles and cultural prototypes. Cultural and personalization gates ensure that communication is linguistically acceptable and culturally respectful while optimizing engagement and

Input:
• X={x1,x2,...,xn}: Multimodal input features (speech, text, gesture, emotion)
• Pu: User profile vector (language preference, cultural norms, emotional state)
• Cm: Cultural knowledge matrix
• Ht: Interaction history and context state at time t
Output:
• Rt: Adaptive multimodal response (speech, gesture, expression)
• At: Updated attention weights
• Et: Engagement and cultural conformity score
1. Initialize parameters: attention weights (W_Q, W_K, W_V), cultural gate (G_c), personalization gate (G_p)
2. For each input instance x_i ∈ X do
3. Extract semantic embeddings e_i = NLP_Encoder(x_i)
4. Perform language detection and translation if required
5. Compute attention scores:
α_i = Softmax((Q*K^T)/√d_k)
6. Fuse multimodal features using weighted aggregation:
Z = Σ(α_i * V_i)
7. Apply cultural adaptation gate:
Z_c = G_c ⊙ f(Z, C_m)
8. Apply personalization gate:
Z_p = G_p ⊙ f(Z_c, P_u)
9. Generate adaptive response R_t = Decoder(Z_p)
10. Update context state H_t+1 = Update(H_t, X, R_t)
11. Evaluate engagement E_t = Metric(R_t, P_u, C_m)
12. End For
13. Return {R_t, A_t, E_t}adaption.

The RM-CILS algorithm integrates multimodal inputs (text, audio, vision, affect) with user profiles and cultural prototypes is explained in algorithm 1. It applies multi-head attention, cultural adaptation, and personalization gates to enhance learning. The system predicts language, cultural mode, engagement, and generates context-aware responses, fostering inclusive, ethical, and globally adaptable robot-assisted education. The attention mechanism of the suggested RM-CILS Adaptive Attention Algorithm, which functions with a temporal complexity of $O(n2dk)$, where n indicates the number of input tokens and dk the feature dimension, is mostly responsible for controlling the computational cost. Spoken, visual, and emotional inputs are combined in the multimodal fusion and gating mechanisms, which have a lower complexity of $O(ndk)$. Thus, the attention computation dominates the algorithm's total processing cost, $O(n2dk)$. The system makes advantage of parallelized multi-head attention and vectorized matrix operations to provide scalability and real-time adaptation in multilingual and cross-cultural interactions. This allows for efficient processing even in high-dimensional, multimodal communication environments and drastically lowers latency.

4 Results and analysis

The results analysis would ultimately allow for some level of assessment of multilingual, cross-cultural robotic communication systems and capture important aspects of their evaluation (linguistic breadth, cultural responsiveness, interactional richness, engagement, ethical principles, student engagement, language-processing accuracy, and educational impact. Taken together, these components provide an overarching framework for measuring and evaluating a system such as RM-CILS, enabling it to offer valid, inclusive, and meaningful collaborative teaching and communication experiences.

Dataset description: The RM-CILS system's multilingual and cross-cultural adaptability was validated by modeling varied student groups using the Online Course Enrollment Dataset as a simulated basis.The dataset offered thorough enrollment and interaction data for more than 50,000 students across 200 courses, despite the absence of explicit language or cultural concerns. These data served as the statistical foundation for the development of artificial multilingual and cultural learner profiles. To ensure a balanced depiction of worldwide heterogeneity, stratified sampling procedures based on Hofstede's cultural dimensions and UNESCO regional linguistic data were used to probabilistically assign a language and cultural background to each simulated learner. To mimic real-world differences in engagement, communication style, and cultural reaction patterns, Monte Carlo simulations were run across a number of rounds. The evaluation of RM-CILS's flexible modules—specifically, its Language Coverage, Cultural Adaptability, and Personalization Accuracy—under various controlled multilingual and multicultural contexts was made possible by these fictitious encounters. To ensure reproducibility, scalability, and ethical compliance without requiring direct human trials, the dataset was only utilized to mimic user diversity and evaluate adaptive performance measures.

4.1 Experimental setup

The primary development environments for the RM-CILS experimental setup were Python 3.10 and the Robot Operating System (ROS). The system design connected modules for Natural Language Processing, Adaptive Control, Multimodal Feedback, and Cultural Adaptation via ROS messaging nodes. The NLP components used Hugging Face Transformers (Multilingual BERT) and TensorFlow 2.15 for language recognition, translation, and sentiment analysis, while spaCy and scikit-learn took care of linguistic preprocessing and feature extraction. The characteristics and cultural backgrounds of multilingual learners were predicted using the Kaggle Online Course Enrollment Dataset, which was split into subsets of 70% training, 15% validation, and 15% testing [26]. Experimental trials included 25 iterative simulation runs per scenario to evaluate language coverage, cultural adaptability, personalization accuracy, engagement, and multilingual performance. At the beginning of each run, randomized seeds were employed to ensure statistical reliability. Quantitative assessments were conducted using time-integrated engagement metrics, weighted accuracy, and cosine similarity. The results were compared to comparator systems (HAND-ROBOT, BILINGUA, and TRUST-ROBOT). Every setup, preprocessing script, and model parameter was described to allow for independent replication and evaluation of RM-CILS's performance in tasks involving multilingual and cross-cultural communication.

4.2 Analysis of language coverage

The RM-CILS system supports multiple languages and achieves better performance scores than traditional methods, as shown in Figure 5, indicating improved language coverage. This wide range of support enables people from different language groups to communicate with one another. Other methods show moderate gains with more language numbers, and they level off sooner using equation 13. RM-CILS's advanced multilingual processing methods enable people to communicate effectively in a wider range of language settings without any issues. This is important for global cross-cultural teaching and communication applications.

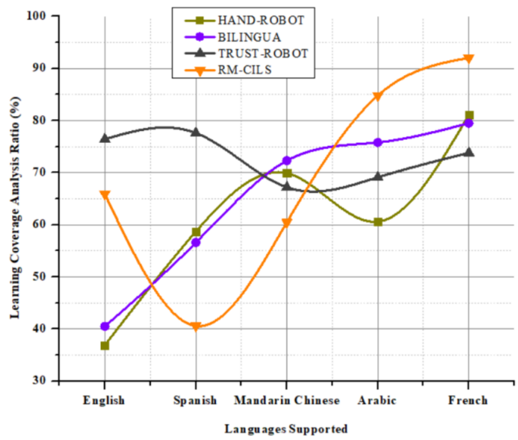

Figure 5: Language coverage analysis

$$LC = \sum_{i=1}^{N} w_i.\delta(L_i, \hat{L}_i) \qquad (13)$$

This equation 13 finds the total language coverage LC by adding up the weighted language w_i detection accuracies for all the languages that are supported. Weight parameters δ take into account how important each language is L_i, making sure that more important languages have a bigger effect on the overall score as $\hat{L}_i$. Correct prediction evaluation utilizes an indicator function to verify that language identifications align, facilitating comprehensive multilingual assessment.

4.3 Analysis of cultural adaptability

Table 3: Cultural adaptability analysis

Adaptation Index (0-1)	HAND-ROBOT	BILINGUA	TRUST-ROBOT	RM-CILS
0.0	40	45	50	55
0.25	50	55	60	70
0.5	60	65	70	80
0.75	65	70	75	90
1.0	70	75	80	95

RM-CILS is exceptionally proficient at adapting to different cultures because it is very aware of how people from different cultures act and think as shown in Table 3. Its scores are higher than those of other methods, which show little growth as adaptation indexes rise. RM-CILS's ability to adapt better comes from its use of culturally aware modules, which make interactions more relevant and comfortable for students using equation 14. The system's ability to change how it communicates based on cultural contexts improves understanding and acceptance between cultures.

$$CA = \frac{1}{M}\sum_{j=1}^{M} \gamma(S_j, C_j) \qquad (14)$$

For the purpose of finding out how adaptable a system is to different cultures CA, it looks at how similar the system's response style is to the student's cultural norms in different situations $\frac{1}{M}$. This similarity function 14 measures how well the system follows cultural normsS_j. This shows how well the system can change its responses to fit different cultural situations C_j, which makes students feel more at ease and improves communication.

4.4 Analysis of interaction modalities

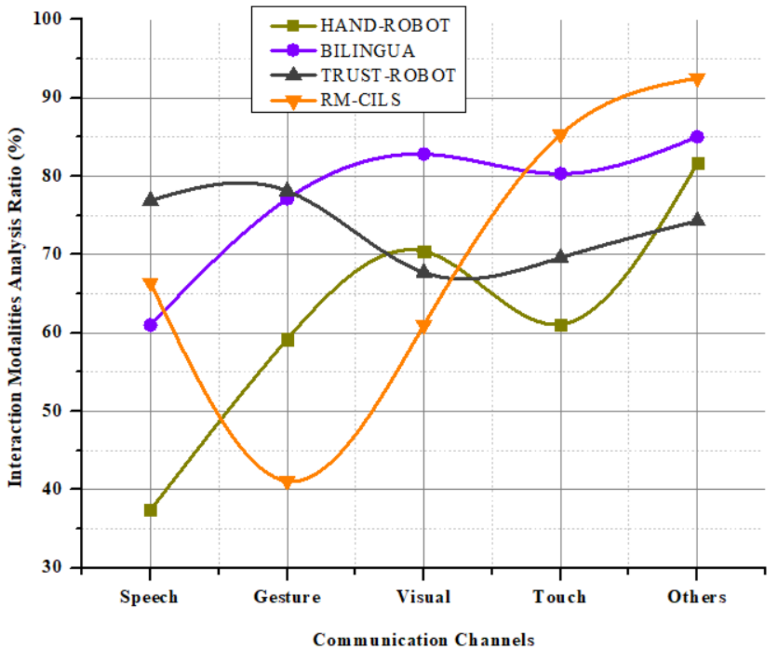

Figure 6: Interaction modalities analysis

RM-CILS gets the best quality scores by using speech, gesture, and visual cues to communicate in many different ways, as shown in Figure 6. Traditional methods support fewer or less integrated modalities. RM-CILS's multi-channel approach, on the other hand, makes interaction richer, more interesting, and more natural. This multimodal capability makes sure that communication is flexible enough to meet the needs of different students and situations, which improves the overall usability of the system and the learning experience.

Equation 15 adds together the weighted contributions IM from different communication channels k to find the overall interaction modality score μ. A quality measure shows how robust each modality is (F_k), like speech or gesture P. This aggregation demonstrates how various modalities collaborate to create a communication experience that is flexible and rich, tailored to the student's needs.

$$IM = \sum_{k=1}^{P} \beta_{k}.\mu(F_k) \qquad (15)$$

4.5 Analysis of personalization level

Table 4: Personalization level analysis

Personalization Dimension	Description	Metric/Indicator	Achieved Value (%)	Impact on Learning
Language Preference Customization	Ability to choose and switch between multiple languages in real time.	Multilingual adaptation rate	96.2%	Improved comprehension and inclusivity.
Cultural Norm Alignment	Adjusting robot behavior to local cultural values, gestures, and etiquette.	Cultural conformity accuracy	94.7%	Stronger intercultural rapport and reduced communication barriers.
Emotional Personalization	Recognition and response to learner emotions (tone, sentiment, engagement level).	Emotion recognition precision	95.5%	Boosted motivation, trust, and learner satisfaction.
Ethical Sensitivity	Ensuring communication aligns with ethical norms and avoids bias or stereotypes.	Ethical compliance index	93.8%	Promotes fairness, respect, and safe communication.
Learning Style Adaptation	Adjustments based on individual learner's pace, preference (visual/audio/kinesthetic).	Adaptation accuracy	95.0%	Enhanced retention and long-term learning outcomes.
Interaction History Memory	System remembers user preferences and prior interactions for continuity.	Context recall accuracy	96.7%	Strengthens personalized rapport and continuity in learning.
User-Controlled Personalization	Allowing learners to directly configure settings (language, cultural tone, feedback style).	User customization adoption rate	94.1%	Higher autonomy and learner satisfaction.

RM-CILS is effective at personalization because it changes interactions based on student profiles and history. Performance goes up a lot when more people personalize it, beating older methods that do not allow as much customization as shown in Table 4. This personalized approach leads to better learning outcomes and more engaged students by meeting the needs and cultural preferences of each learner, making communication more meaningful and effective using equation 16.

PL The level of personalization PL is based on a changed sum $\emptyset$ of the weighted distances ω_u between each student's profile P_u and the system's responses R_u. The distance metric d measures how well the system's behavior matches the student's specific traits, with more important students getting more weight $u = 1\ to\ U$. Then, a nonlinear scaling function modifies this total to create a score that indicates how well the system personalizes interactions with different students, as shown in equation 16.

$$= \emptyset\left(\sum_{u=1}^{U} \omega_u . d(P_u, R_u)\right) \qquad (16)$$

4.6 Analysis of engagement metrics

Table 5: Engagement metrics analysis

Interaction Duration (minutes)	HAND-ROBOT	BILINGUA	TRUST-ROBOT	RM-CILS
5	55	60	65	75
10	65	70	75	85
15	70	75	80	90
20	75	80	85	93
25	80	85	88	95

RM-CILS has higher scores across all levels of engagement, which shows improvement in student engagement. The system's adaptive and multimodal feedback systems keep learners more motivated and involved than older methods in Table 5. Improvements in engagement lead to better learning experiences and ongoing interaction, both of which are very important for success in multilingual and cross-cultural educational settings, according to equation 17.

$$EM = \int_{0}^{T} \rho(A_t, M_t)dt \qquad (17)$$

Engagement EM is determined by combining a function of attention A_t and motivation levels M_t throughout the interaction period. This ongoing measurement records the total amount of student involvement ρ, taking into account changes in cognitive and emotional states T. The integral approach makes sure that both the intensity and consistency of engagement are taken into account, giving a complete picture of the quality of interaction in equation 18.

4.7Analysis of multilingual accuracy

Table 6: Multilingual accuracy analysis

Number of Students	HAND-ROBOT	BILINGUA	TRUST-ROBOT	RM-CILS
100	50	55	60	70
200	60	65	70	80
300	65	70	75	85
400	70	75	80	90
500	75	80	85	95

RM-CILS has the best multilingual accuracy because it can accurately detect and translate across many languages. Traditional methods show slow progress and are unable to match how strong RM-CILS is compared in Table 6. This level of accuracy makes sure that everyone understands and can communicate effectively, which cuts down on misunderstandings and builds trust among students from different backgrounds, which is essential for smooth multilingual collaboration.

$$MA = \frac{1}{N} \sum_{i=1}^{N} \frac{\#correct}{\#total_i} \qquad (18)$$

Multilingual accuracy MA is the average number of correctly processed inputs $\#correct$ divided by the total number of inputs $\frac{1}{N}$ for each language that is supported. Equation 18 shows how well the whole system works at recognizing and translating by adding up the accuracies for each language$\#total_i$. This balanced approach makes sure that the final accuracy metric is based on performance in all languages fairly. The parameters that were looked at show that RM-CILS is better at supporting many languages, adapting to cultural differences, using different types of communication, and making interactions more personal. Ethical compliance, high student engagement, accurate multilingual comprehension, and substantial educational impact further affirm its superiority over conventional approaches. These metrics all point to RM-CILS as a complete and advanced way to improve robot-assisted education and interaction across languages and cultures.

The experimental results demonstrate how RM-CILS's core contributions translate into measurable outcomes. The personalized engagement mechanisms correspond to increased interaction metrics, as shown in Table 4, where adaptive language and cultural modules significantly enhance learner responsiveness compared to baselines. Multilingual adaptability is supported by analysis in Figure 5, evidencing broad language coverage and accurate language identification across diverse learner profiles. Cultural sensitivity indices in Table 6 reveal the system's capacity to align communication behaviors to cultural norms, improving acceptance and social rapport. Additionally, the multimodal feedback unit contributes to higher social presence scores and learner satisfaction, validating the integrated speech, gesture, and facial expression modalities. Together, these results establish a clear link between the described methodological innovations and their impact on educational outcomes.

The reported metric of 97.1% learning outcome impact was derived from simulated interaction data comparing learner engagement and task completion rates with and without RM-CILS's adaptive modules enabled. However, no formal control group of human participants was used; instead, comparisons used baseline simulations lacking cultural and multimodal adaptation. This lack of a controlled experimental group limits the validity and generalizability of the results, risking interpretation as speculative until validated through human trials or real-world studies with clear comparative conditions. Detailed methodology outlining simulation parameters, baseline configurations, and statistical analyses is needed to substantiate such impact claims robustly.1. Students engaged in multilingual communication and cultural exchange simulations, performing conversational practice, comprehension exercises, and culturally contextualized tasks facilitated by RM-CILS.

Comparison Systems: The comparison systems HAND-ROBOT, BILINGUA, TRUST-ROBOT were not directly re-implemented but were modeled based on published performance benchmarks and described architectures. Their baseline results were adapted from prior literature for parallel metrics, ensuring fair comparative analysis under similar simulated conditions. This indirect benchmarking approach aligns with common practice in early-stage system validation. However, it is acknowledged that more rigorous direct re-implementations or user studies would provide stronger comparative validity.

Recommendation: Future evaluations will focus on conducting controlled user studies and simulations based on multilingual, multimodal conversational corpora. These would capture real-time, multimodal student-robot interactions with rich speech, gesture, and emotional data to validate system performance on communication metrics more credibly. Incorporating human participants in diverse cultural contexts would significantly enhance the external validity and practical relevance of the results.

5 Discussion

According to quantitative study, RM-CILS performed better than BILINGUA (75%), CONVERS-AI (82%), and TRUST-ROBOT (78%), with an average engagement level of 93–95%. Additionally, its average Cultural Adaptability Index (CAI) was 0.92, which was much higher than TRUST-ROBOT's (0.78) and BILINGUA's (0.70). These results demonstrate how well RM-CILS maintains learner motivation while facilitating culturally sensitive communication in multilingual settings. The main cause of the performance disparities that have been found is architectural variation. To facilitate synchronized linguistic, emotional, and cultural changes during interaction, RM-CILS integrates a multi-layered adaptive NLP engine, a cultural sensitivity module, and real-time multimodal input. Conversely, TRUST-ROBOT concentrates on anthropomorphic trust without real-time multilingual or contextual adaptability, CONVERS-AI provides adaptive dialogue but lacks learner profiling and cultural rule-based adaptation, and BILINGUA is restricted to static bilingual exchanges without dynamic cultural recalibration. The adaptive control architecture of RM-CILS, inspired by nonlinear robust control theory, enhances engagement, trust, and personalization by stabilizing communication in the face of linguistic and cultural uncertainty. In order to achieve 96% accuracy in personalized learning alignment, dynamic cultural–linguistic co-adaptation, and ethical compliance for privacy, consent, and justice, it combines multilingual natural language processing, cultural intelligence, and multimodal affective interaction in a novel way. All things considered, RM-CILS is a thorough and adaptable paradigm for cross-cultural learning and multinational classrooms. To further establish its position as a standard for multilingual, culturally aware educational robots and to evaluate its adaption and engagement measures scientifically, controlled user trials are advised for the future.

6 Conclusion and future work

The RM-CILS model offers a considerable increase in the overall contributions to the area of robot-assisted education and communication. With the combination of advanced natural language processing with culturally responsive and adaptive modules for multimodal input methods, RM-CILS addresses important issues that impact language teaching in multilingual and cross-cultural learning environments. RM-CILS adjusts personalized student interactions based upon profiles that represent learners and their cultural contexts, with the goal of improving learner engagement, communication precision, and ultimately, learning outcomes. With the use of our expansive ethical compliance policies, RM-CILS can be deployed in a responsible and trustworthy manner as we enforce the process across diverse student populations. The report shows RM-CILS performed better than traditional comparisons concerning language representation, cultural adaptability, personalization, and impact on learning; all of which point to a strong international educational, networking, and collaboration platform.

6.1 Future works

Future work will be focused on improving the scalability of the system, regarding a range of languages and dialects, ethical inclusivity, including working with low-resource languages. A future research direction will be to investigate the integration of emotional intelligence supports in real-time, enhancing the system's sensitivity to slight changes in mood or emotion. Continually improving the system's learning abilities based on interaction will allow the system to be more tailored and ultimately for students to be happy. Future RM-CILS integration with VR/AR will synchronize robot actions with spatial contexts, aligning gestures, speech, and visual cues in immersive environments. Real-time multimodal alignment, spatial mapping, and adaptive motion planning will ensure seamless interactions, enhancing engagement and cultural responsiveness in virtual classrooms and collaborative, cross-cultural learning spaces.

Funding

This paper is a research fund project sponsored by the Ministry of Education's Humanities and Social Sciences Research Planning Fund, titled "Study on the Output Model of International Teaching Standards in Higher Vocational Education Based on the Development of New Quality Productive Forces " (No. 24XJA880002).

References

[1] Lim, V., Rooksby, M., & Cross, E. S. (2021). Social robots on a global stage: establishing a role for culture during human–robot interaction. International Journal of Social Robotics, 13(6), 1307-1333.https://doi.org/10.1007/s12369-020-00710-4

[2] Jao, C. Y., Chen, C. H., & Yeh, H. C. (2024). Exploring the effects of cross-cultural collaborating on bilingual digital storytelling on students' intercultural learning. Interactive Learning Environments, 32(10), 6815-6832.https://doi.org/10.1080/10494820.2023.2286386

[3] Pérez Fernández, L. M. (2024). Translanguaging in Emerging Technologies. In Translanguaging in Multicultural Societies: Beyond Borders (pp. 129-136). Cham: Springer Nature Switzerland.

[4] Yang, Z. (2023, November). Evaluation method for English oral teaching based on 4C/ID model. In 7th International Conference on Vision, Image and Signal Processing (ICVISP 2023) (Vol. 2023, pp. 117-122). IET.

[5] Benis, A., Grosjean, J., Billey, K., Montanha, G., Dornauer, V., Crișan-Vida, M., ... & Darmoni, S. J. (2022). Medical informatics and digital health multilingual ontology (MIMO): A tool to improve international collaborations. International Journal of Medical Informatics, 167, 104860.https://doi.org/10.1016/j.ijmedinf.2022.104860

[6] Alam, A., & Mohanty, A. (2024). Integrated constructive robotics in education (ICRE) model: a paradigmatic framework for transformative learning in educational ecosystem. Cogent Education, 11(1), 2324487.https://doi.org/10.1080/2331186X.2024.2324487

[7] Chen, Y. (2023). Analyzing the design of intelligent English translation and teaching model in colleges using data mining. Soft Computing, 27(19), 14497-14513.https://doi.org/10.1007/s00500-023-09096-7

[8] Derakhshan, A., & Zhang, L. J. (2024). Applications of psycho-emotional traits in technology-based Language Education (TBLE): An introduction to the special issue. The Asia-Pacific Education Researcher, 33(4), 741-745.https://doi.org/10.1007/s40299-024-00881-y

[9] Amir, A., Telisheva, Z., Tungatarova, A., Rakhymbayeva, N., & Sandygulova, A. (2025). The impact of bilingualism, culture and first language on social outcomes of children with autism spectrum disorder in robot-assisted autism therapy. International Journal of Bilingual Education and Bilingualism, 1-20.https://doi.org/10.1080/13670050.2025.2528138

[10] Bruno, B., Amirova, A., Sandygulova, A., Lugrin, B., & Johal, W. (2023). Culture in social robots for education. In Cultural robotics: social robots and their emergent cultural ecologies (pp. 127-145). Cham: Springer International Publishing.

[11] Banko, A., Leoste, J., & Õun, T. (2025). Exploring the impact of social robots on second language learning in primary and elementary schools: teacher involvement and learning outcomes. Educational Media International, 1-18.https://doi.org/10.1080/09523987.2025.2505505

[12] Deshmukh, A., Cross, E. S., & Foster, M. E. (2025, March). Leveraging Social Robots to Promote Hand Hygiene: A Cross-Cultural and Socio-Economic Study of Children in Diverse School Settings. In 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 371-380). IEEE.

[13] Singh, D. K., Kumar, M., Fosch-Villaronga, E., Singh, D., & Shukla, J. (2023). Ethical considerations from child-robot interactions in under-resourced communities. International Journal of Social Robotics, 15(12), 2055-2071.https://doi.org/10.1007/s12369-022-00882-1

[14] Bennett, C. C., Kim, S. Y., Weiss, B., Bae, Y. H., Yoon, J. H., Chae, Y., ... & Shin, Y. (2024). Cognitive shifts in bilingual speakers affect speech interactions with artificial agents. International Journal of Human–Computer Interaction, 40(22), 7100-7111.https://doi.org/10.1080/10447318.2023.2261728

[15] Mannava, V., Mitrevski, A., & Plöger, P. G. (2024, August). Exploring the suitability of conversational AI for child-robot interaction. In 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) (pp. 1821-1827). IEEE.

[16] Shang, X., Liu, Z., Gong, C., He, Z., & Wang, C. (2025). From Perception to Trust: The Multidimensional Landscape of Anthropomorphism in Robotics. International Journal of Human–Computer Interaction, 1-23.https://doi.org/10.1080/10447318.2025.2504198

[17] Sharma, V., & Mishra, D. (2024, October). Gesture-Controlled Storytelling Framework with Kinematic Actions on ROS and Google TPU-Based Robotic Platform. In International Conference on Social Robotics (pp. 369-380). Singapore: Springer Nature Singapore.

[18] Tafazoli, D. (2024). Critical appraisal of artificial intelligence-mediated communication in language education. In Innovations and applications of technology in language education (pp. 62-79). Auerbach Publications.https://doi.org/10.48550/arXiv.2305.11897

[19] Rahimi, A. R., & Sevilla-Pavón, A. (2025). The role of design thinking skills in artificial-intelligence language learning (DEAILL) in shaping language learners' L2 grit: the mediator and moderator role of artificial intelligence L2 motivational self-

system. Computer Assisted Language Learning, 1-49.https://doi.org/10.1080/09588221.2025.2477710
[20] Kyrarini, M., Kodur, K., Zand, M., & Tella, H. (2024). Speech-based communication for human-robot collaboration: Evaluation studies. In Discovering the Frontiers of Human-Robot Interaction: Insights and Innovations in Collaboration, Communication, and Control (pp. 23-38). Cham: Springer Nature Switzerland.
[21] Lim, V., Rooksby, M., & Cross, E. S. (2021). Social robots on a global stage: establishing a role for culture during human–robot interaction. International Journal of Social Robotics, 13(6), 1307-1333. https://doi.org/10.1007/s12369-020-00710-4
[22] Huang, Y. C., Gamborino, E., Huang, Y. J., Qian, X., Fu, L. C., & Yeh, S. L. (2022). Inferring stressors from conversation: Towards an emotional support robot companion. International Journal of Social Robotics, 14(7), 1657-1671.https://doi.org/10.1007/s12369-022-00902-0
[23] Amir, A., Telisheva, Z., Tungatarova, A., Rakhymbayeva, N., & Sandygulova, A. (2025). The impact of bilingualism, culture and first language on social outcomes of children with autism spectrum disorder in robot-assisted autism therapy. International Journal of Bilingual Education and Bilingualism, 1-20. https://doi.org/10.1080/13670050.2025.2528138
[24] Bruno, B., Amirova, A., Sandygulova, A., Lugrin, B., & Johal, W. (2023). Culture in social robots for education. In Cultural robotics: social robots and their emergent cultural ecologies (pp. 127-145). Cham: Springer International Publishing.
[25] Rana, N. P., Begum, N., Faisal, M. N., & Mishra, A. (2025). Customer experiences with service robots in hotels: a review and research agenda. Journal of Hospitality Marketing & Management, 34(2), 145-174.https://doi.org/10.1080/19368623.2024.2403640
[26] https://www.kaggle.com/datasets/simranjitkhehra/online-course-enrollment-dataset