

DOI: 10.5281/zenodo.12511043

DEVELOPMENT AND VALIDATION OF EMOTIONBRIDGE: A GPT-BASED DIAGNOSTIC SYSTEM FOR EMPATHIC COMMUNICATION SKILLS

Yuna Lee¹, AiJin Xiao², HyeRan Lee³, Sang-Soo Lee^{4*}

¹*Institute of Educational Policy Research, Future Education Institute, Uiryeong, South Korea.*

^{1,2,3}*Department of Education, Pusan National University, Busan, South Korea.*

Received: 01/12/2025

Accepted: 02/01/2026

Corresponding author: Sang-Soo Lee
(soolee@pusan.ac.kr)

ABSTRACT

This study developed and validated EmotionBridge, a GPT-based diagnostic tool for assessing empathic communication through real-time, scenario-based interactions. Using a Design and Development Research (DDR) approach, the system was grounded in a theoretical framework with two domains: empathic expression and empathic response. The prototype, built via GPT Builder, incorporated rubric-based evaluations. Expert reviews demonstrated high content validity (CVI = .95 and .97), and concurrent validity was supported by correlations between expert and self-report scores. Usability testing with 35 participants yielded a strong SUS score (79.64), and qualitative feedback informed iterative refinements. The results demonstrate the promise of generative AI in delivering scalable and personalized diagnostics in social-emotional learning, offering a replicable design model that links large language models with structured assessment principles.

KEYWORDS: Empathic Communication, GPT-based system, AI-assisted assessment, usability evaluation, social-emotional learning.

1. INTRODUCTION

In an increasingly complex and emotionally demanding world, empathic communication has become an essential skill for facilitating meaningful interpersonal relationships and fostering social cohesion. Empathic communication is defined as the ability to accurately recognize, understand, and respond to the emotional states of others with sensitivity and appropriateness (Bernhardt & Singer, 2012; Rogers, 1975). This ability contributes not only to social adaptation and relationship success but also to individual emotional well-being and quality of life (Jang et al., 2012; Qin et al., 2024; Sze et al., 2012).

In educational contexts, empathic dialogue plays a critical role in providing emotional support, enhancing students' sense of self-esteem and belonging (Wachs et al., 2023), and creating psychologically safe learning environments that foster motivation and engagement (Ge et al., 2021; Xiang et al., 2022; Waxman, 1983). However, many individuals experience difficulties in effectively expressing empathy (Vachon & Lynam, 2016), and there is a lack of practical tools for diagnosing and improving empathic communication skills in authentic settings.

Current assessment methods—including self-report questionnaires and expert-based interviews or observations—are limited in their objectivity, scalability, and ecological validity. Self-report measures are vulnerable to social desirability bias and subjectivity (Melchers et al., 2015; Murphy & Lilienfeld, 2019; Sunahara et al., 2022), while expert evaluations require substantial time and resources and may be affected by observer bias and reactivity (Brown et al., 2022; Fuller et al., 2021). These limitations underscore the need for an interactive and scalable diagnostic tool capable of assessing empathic communication in real-time contexts.

Recent advances in generative artificial intelligence (AI), particularly large language models such as GPT (Generative Pre-Trained Transformer), offer new opportunities to address this need. GPT technologies demonstrate the capacity to analyse and interpret complex emotional expressions and conversational contexts with increasing precision (Akpan et al., 2024; Elyoseph et al., 2023; Kolomaznik et al., 2024; Mamina & Piraynen, 2023). These advancements open new possibilities for evaluating empathic communication through AI-driven dialogue systems.

In response, this study introduces EmotionBridge, a GPT-based diagnostic system designed to assess empathic communication skills through structured, scenario-based conversations. The system evaluates

two core dimensions of empathic communication—empathic expression and empathic response—and includes subcomponents such as empathetic clarification, adaptive phrasing, acknowledgment, deep empathic inquiry, and emotional support provision. Built upon the principles of Nonviolent Communication (NVC) and the Barrett-Lennard empathy cycle, EmotionBridge provides both real-time diagnostic scoring and qualitative feedback to support individual development.

By offering an intuitive and accessible platform that overcomes the limitations of traditional assessment methods, EmotionBridge enables users to reflect on and improve their empathic communication anytime and anywhere. The system's potential extends beyond education to counselling, workplace training, and other applied domains where empathic interaction is essential.

The purpose of this study is twofold: (1) to describe the theoretical foundation, system design, and validation process of the EmotionBridge system, and (2) to empirically verify its reliability, validity, and usability through expert reviews, correlation analyses, and user feedback. This study contributes to the growing field of AI-assisted social-emotional learning by offering a design-valid, scalable, and practical diagnostic tool for empathic communication.

2. LITERATURE REVIEW

2.1 Theoretical Foundations and Emerging AI-Based Approaches to Empathic Communication Assessment

Empathic communication refers to the ability to understand and respect others' emotional states and perspectives, and to respond both verbally and nonverbally in a sensitive and appropriate manner (Fuller et al., 2021; Rogers, 1975). Rogers (1975) described empathic listening as "entering the other's perceptual world without prejudice" in order to understand them deeply. Wolvin (1988) identified three key components of empathic speech: attending to the speaker's experiential world, recognizing both emotional and verbal content, and responding with behavioural feedback. Expanding on these foundations, Ioannidou, Konstantikaki, (2006) highlighted emotional sharing, respectful language, emotional attunement, receptive attitude, active engagement, and courteous expression as central elements of empathic communication.

In this study, empathic communication is operationalized into two diagnostic domains: empathic expression, which refers to the ability to articulate one's own emotions and needs clearly, and

empathic response, which refers to the ability to understand and respond appropriately to the emotional messages of others.

Traditionally, empathic communication has been assessed through self-report scales (Davis, 1980; Lawrence et al., 2004) and expert evaluations (Truax & Carkhuff, 1967). However, such methods often fail to capture language behaviour in authentic, real-time interactions. Recently, researchers have shifted toward interaction-based, scenario-driven approaches. For instance, Kim et al. (2023) analysed empathic response types in adolescent counselling scenarios and found significant differences in empathy scale scores depending on response quality. Park (2020) developed an empathy assessment framework that integrates empathic listening, emotional expression, and feedback based on Barrett-Lennard's (1981) empathy cycle. These developments reflect an increasing demand for diagnostic models that can integrate both qualitative and quantitative analysis, and for tools that can evaluate real-time communication more authentically and objectively.

Recent advances in generative AI, particularly GPT-based large language models, offer new possibilities for developing such tools. GPT enables human-like conversation through real-time simulations and has been applied in educational settings for dialogue training and formative feedback (Maurya, 2024). In the context of empathy training, GPT models have been shown to support example-based modelling and immediate feedback, thus supporting skill acquisition (Ayers et al., 2023). Chang et al. (2023) emphasized the importance of goal-setting, feedback provision, and personalization when using GPT-based chatbots for self-regulated learning—principles that align with the design approach of the present study. Furthermore, GPT Builder enables role-based customization and scenario configuration, making it a viable platform for building interactive systems that diagnose and support empathic communication skills.

2.2 Theoretical Foundations for Emotion Bridge

In designing a diagnostic system for empathic communication, it is essential to consider both the principles for constructing dialogue scenarios between the learner and the GPT agent, and the diagnostic principles for evaluating empathic abilities based on the content of these interactions. The former addresses how the system can deliver an immersive and pedagogically effective conversational experience, while the latter concerns the criteria and algorithms by which the learner's empathic competencies are assessed.

In the following sections, the key theoretical foundations applied to the EmotionBridge system are

examined in two parts: system design principles, which address dialogue scenario construction and customization, and diagnostic design principles, which outline the algorithmic framework for assessing empathic communication.

2.3 System Design Principles

The system was designed based on three core principles to effectively diagnose and enhance empathic communication skills. First, the dialogue scenarios were constructed to reflect realistic and multi-layered social contexts, such as teacher-student interactions, peer conflict, and family communication. These authentic settings allow learners to safely practice diverse empathic responses and enhance their self-efficacy in real-life situations (Bialystok & Kukar, 2018; Nikendei et al., 2005). Second, the diagnostic algorithm evaluates empathic expression and response using natural language processing (NLP) technologies. Specifically, it analyses whether the learner articulates their emotions and needs clearly and non-judgmentally, and whether they accurately interpret and respectfully respond to the emotions and messages of others (Son et al., 2024). The algorithm conducts a multi-layered assessment encompassing emotional understanding, perspective-taking, and empathic attitude (Decker et al., 2014), and delivers both quantitative scores and detailed, qualitative feedback. Third, the system leverages the customization features of GPT Builder to ensure that the GPT agent maintains a consistent role and tone appropriate to each scenario. This allows users to engage in sustained, coherent dialogue contexts and explore empathic responses at varying levels of depth. Additionally, immediately after the conversation, the system provides constructive feedback based on the NVC [observation-feeling-need-request] framework and I-message structure, supporting learner reflection and strategic improvement (Muhyatun, 2022; Rosenberg, 2015).

The system design was informed by three core elements: scenario-based dialogue construction, rubric-based diagnostic logic, and GPT customization through configuration and prompt design. The system design principles proposed in this study offer practical guidelines for developing a GPT-based diagnostic system for empathic communication. A detailed breakdown of design principles and implementation examples is provided in Appendix A.

2.4 Diagnostic Design Principles

This study aimed to develop a GPT-based diagnostic algorithm for assessing empathic communication by systematizing a set of theoretical

principles grounded in empathic expression and empathic response. These principles were informed by Barrett-Lennard's (1981) empathy cycle model, Rosenberg's (2015) Nonviolent Communication (NVC) framework, I-message structures (Darrington & Brower, 2012; Muhyatun, 2022), and the empathic listening theories of Rogers (1975) and Covey (1989). Through this integration, two major diagnostic domains were identified, each with distinct subcomponents and criteria.

Empathic expression is defined as the speaker's ability to clearly and respectfully convey their emotions and needs. It consists of two subcomponents: empathetic clarification, which assesses clarity, directness, and non-judgmental articulation following the NVC's four-part structure (observation, feeling, need, request), and empathetic adaptability, which evaluates the speaker's ability to adjust word choice, tone, and conversational direction based on social context and listener characteristics.

Empathic response refers to how the listener receives and reacts to the speaker's emotional content. This includes three subcomponents: empathetic acknowledgment, which involves recognizing and paraphrasing the speaker's message, deep empathic inquiry, which evaluates the quality of open-ended, reflective questioning to deepen emotional understanding, and emotional support provision, which assesses the use of emotional validation, encouragement, and constructive suggestions to strengthen relational trust.

The diagnostic framework thus distinguishes two core domains with five corresponding subcomponents, each operationalized through specific behavioural indicators. The proposed

principles offer a structured approach for analysing empathic dialogue in AI-mediated interactions, serving as the foundation for the GPT-based diagnostic algorithm used in this study. A full outline of subdomains and evaluation criteria is provided in Appendix B.

3. METHODS

This study aimed to construct a theoretical framework for empathic communication diagnosis and to derive design principles for system development, adopting a Design and Development Research (DDR) methodology (Richey & Klein, 2007). The DDR approach was chosen for its systematic and iterative structure, supporting continuous refinement through empirical validation.

The research process consisted of five major phases, as illustrated in Figure 1: (1) conducting a literature review to establish GPT-based system design principles and empathic communication diagnostic principles, (2) validating the clarity, feasibility, technical consistency, and applicability of the design and diagnostic principles through expert review, (3) developing the initial diagnostic prototype using OpenAI's ChatGPT GPT Builder, integrating both configuration and knowledge modules, (4) conducting empirical validation, including concurrent validation through agreement analysis between expert and GPT scores, correlation analysis with self-report empathy scales, and usability testing using the SUS and UEQ, and (5) finalizing the EmotionBridge system based on findings from the validation phases, including iterative revisions informed by both expert and user feedback.

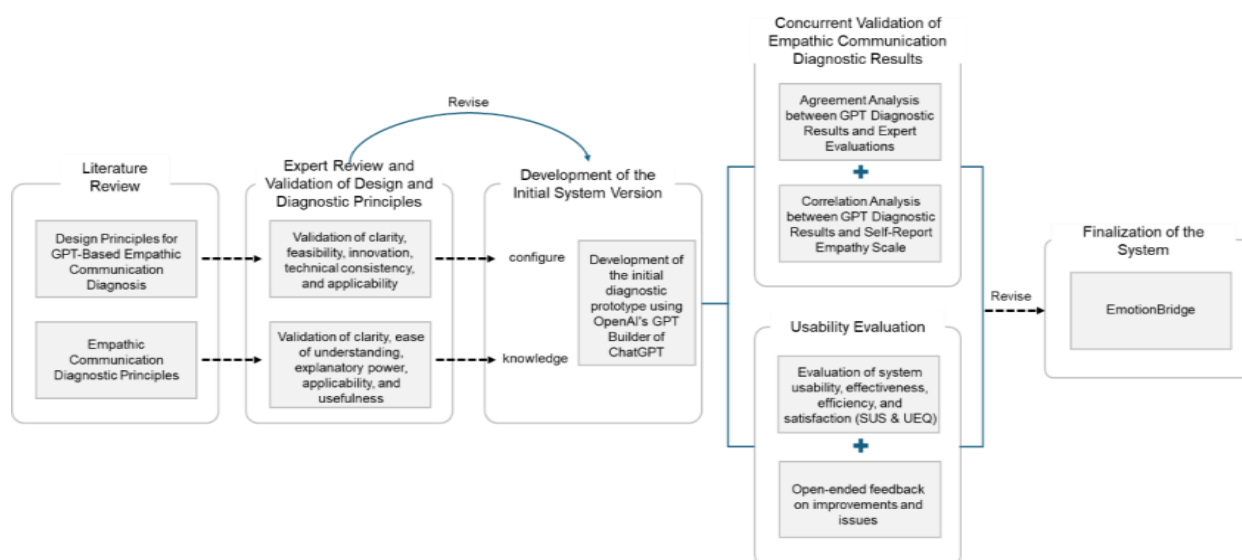


Figure 1: Development and Validation Process of the GPT-Based EmotionBridge System

3.1 Phase1: Literature Review and Conceptual Design

A comprehensive literature review was conducted to establish the theoretical foundation for developing a GPT-based diagnostic system for empathic communication. First, previous studies on the concept, components, and assessment principles of empathic communication were analysed. Through this review, core elements such as emotional recognition, empathic listening, emotional responsiveness, and reflective feedback were identified. These components informed the construction of evaluation criteria for the diagnostic system. The review then extended to studies related to diagnostic systems using artificial intelligence (AI) and natural language

processing (NLP) technologies. This included literature that explored the capacity of generative AI models – such as GPT – to interpret human language, infer emotional meaning, and respond based on conversational context. Particular attention was given to prior research on sentiment analysis, response generation strategies, and empathy-inducing algorithms in AI-based dialog systems. Furthermore, based on Rogers' (1975) theory of empathic listening, Rosenberg's (2015) Nonviolent Communication (NVC) model, and Barrett-Lennard's (1981) empathy cycle, this study defined two central diagnostic domains: empathic expression and empathic

response. These domains served as the conceptual backbone for the initial design specifications and rubric-based evaluation framework of the EmotionBridge system. The results of this literature review served as a basis for defining the core principles and design criteria necessary for diagnosing empathic communication using GPT. These findings directly informed subsequent phases of expert review and system design.

3.2 Phase 2: Expert Review and Content Validation

To verify the content validity of the GPT-based empathic communication diagnostic system, an expert panel was formed, consisting of individuals with recognized expertise in relevant fields. The panel included professionals with substantial theoretical and practical experience in empathic communication, artificial intelligence (AI), natural language processing (NLP), and the development of digital diagnostic systems. The criteria for selecting panel members were as follows: (1) individuals with theoretical or practical expertise in empathic communication, counselling, or psychology, (2) those with technical understanding of AI or GPT-based system development, and (3) individuals with experience in designing or reviewing educational or digital learning diagnostic tools. Based on these criteria, a total of ten experts were selected. The composition of the panel is presented in Table 1.

Table 1: Sample table counting some elements of a manuscript

Panel Type	Expertise Area	Affiliation Position	Experience [Years]
Academic Expert	AI in Education, Empathy-based Instructional Design	Professor, P University	28
	Instructional Design, Educational Technology	Professor, D University	23
	AI Systems, Natural Language Processing	Professor, Computer Engineering	15
	Educational Measurement and Evaluation	Professor, P University	20
	AI Systems, AI Ethics	CEO, Y Company	14
	Educational Psychology	Professor, Educational Psychology	10
	Language & Communication Education	Professor, Language Education	12
Field Practitioner	Counseling Psychology	Professor, Counseling Psychology	13
	School-based Counseling Practice	Director, Private Counseling Center	12
	Empathy Training Program Implementation	Supervisory Officer, Provincial Office of Education	15

To ensure the design validity of the GPT-based empathic communication diagnostic system, expert review was conducted from two perspectives. First, the system design principles were evaluated based on the evaluation standards of design and development research (Plomp & Nieveen, 2013; Richey & Klein, 2007). This included examining the clarity of the design principles, feasibility of implementation, appropriateness of GPT Builder utilization, innovativeness, and technical consistency (Holmes et al., 2019; Luckin, 2017; Zawacki-Richter et al., 2019). Second, the diagnostic principles for empathic

communication were reviewed to determine whether the GPT-based system could effectively assess the essential components of empathic interaction. The evaluation focused on the universality, comprehensibility, explanatory clarity, applicability, and overall usefulness of the principles (Gonsalves, & Pearson, 2023; Merrill, 2020; Reigeluth, 1999). Each item was evaluated using both a 4-point Likert-type scale (1 = not appropriate at all, 4 = highly appropriate) and open-ended items for qualitative feedback and suggestions. For quantitative analysis, the Content Validity Index (CVI) was calculated, representing the

proportion of experts who rated an item as either "appropriate (3)" or "highly appropriate (4)" (Davis, 1992; Lynn, 1986). Items with a CVI of 0.78 or higher were considered to have acceptable content validity.

Additionally, inter-rater agreement was assessed using the within-group agreement index (rWG), with a value of 0.70 or higher interpreted as indicating a high level of consensus among experts (James et al., 1984; LeBreton & Senter, 2008). Based on the review results, qualitative suggestions from experts were actively incorporated to revise and improve both the system design principles and the diagnostic framework for empathic communication. These revisions directly informed the development of the initial version of the GPT-based diagnostic system.

3.3 Phase 3: System Development Using GPT Builder

To construct the GPT-based empathic communication diagnostic system, a customized model was developed using OpenAI's GPT Builder. Among the available modes – create and configure –

the configure mode was selected for its capacity to enable detailed customization of system parameters, which was essential for implementing the system's core design principles. These principles included scenario-based dialogue, diagnostic algorithms for empathic communication, and customizable interaction settings.

As illustrated in Figure 2, the system was configured using two key components: The Configure and Knowledge functions. The Configure section defined conversational roles, scenario types, instructional prompts, and output formatting, while the Knowledge module was used to upload diagnostic rubrics based on the system's empathic communication framework. These rubrics trained the GPT model to analyse user utterances and evaluate empathic expression and response in real time. Five iterative cycles of micro-adjustments were carried out to refine conversational tone, prompt sequencing, response formats, and feedback delivery, resulting in the finalized prototype.

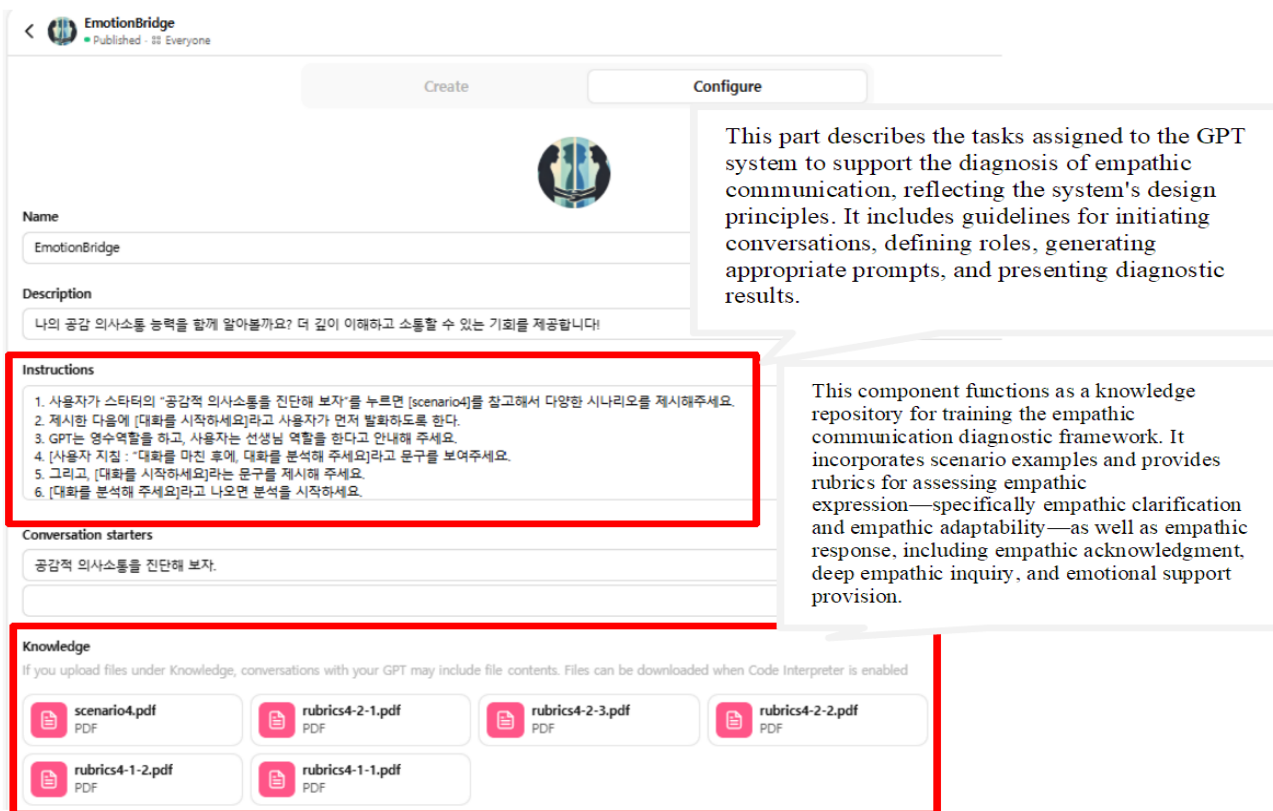


Figure 2: GPT builder Configure and Knowledge

As shown in Figure 3, the prototype interface includes a clearly stated purpose – “Shall we explore your empathic communication skills together? This system offers an opportunity for deeper understanding and interaction.” – which serves to engage users and set a supportive tone for the

diagnostic experience. A single conversation starter prompt – “Let’s begin diagnosing your empathic communication skills.” – was designed to initiate the structured interaction. These features together provide a user-friendly, immersive entry point into the assessment process.



Figure 3: Prototype of the GPT-Based EmotionBridge

3.4 Phase 4: Concurrent Validation and Usability Testing

3.4.1 Phase 4.1 Concurrent Validity

To establish the validity of the EmotionBridge system, two concurrent validation strategies were implemented: comparison with expert ratings and correlation with a validated empathy scale.

Phase 4.1.1. Agreement between GPT and Expert Ratings.

The expert panel participating in this analysis consisted of five individuals who had also contributed to the prior content validity review. These experts held specialized knowledge in empathic communication, education, and psychological assessment, and independently evaluated the same diagnostic scenarios. To assess the actual interaction between users and the GPT-based system, the experts reviewed transcripts of conversations involving 35 participants who had engaged with the system during the usability evaluation phase. A diagnostic rubric developed by the research team was used to guide evaluation. The rubric reflected the system's diagnostic framework, including empathic expression [empathetic clarification and empathetic adaptability] and empathic response [empathetic acknowledgment, deep empathic inquiry, and emotional support provision]. Each criterion was rated on a 5-point scale, ranging from 1 ("not at all present") to 5 ("very well demonstrated"), and all experts independently scored the GPT-generated diagnostic responses using the same evaluation standards. To quantitatively examine the agreement between the GPT-based diagnostic scores and the expert evaluations, the Intraclass Correlation Coefficient (ICC) was employed (McGraw & Wong, 1996).

The analysis was conducted using SPSS version 26.0 and applied a two-way mixed model. Interpretation of ICC values followed the guidelines proposed by Koo and Li (2016), where values $\geq .75$ indicate excellent reliability, values between .50 and .74 indicate moderate reliability, and values below .50 are considered to indicate poor reliability.

Phase 4.1.2. Correlation with Self-Report Empathy Scale.

Participants also completed a 52-item self-report scale, which was adapted from existing instruments developed by Kim (2011), Sohn(2022), Huh (2003), and Kang (2018). The scale was reconstructed to align with the diagnostic framework of the GPT-based empathic communication system. It comprised two major domains: empathic expression (including empathetic clarification and empathetic adaptability) and empathic response (including empathetic acknowledgment, deep empathic inquiry, and emotional support provision). To examine the validity and reliability of the scale, data were collected from a heterogeneous sample of 216 participants – including teachers, students, graduate students, and university faculty – from the Busan and Gyeongnam regions in South Korea, using a convenience sampling approach. A confirmatory factor analysis(CFA) was conducted to assess structural validity. The results indicated acceptable model fit ($\chi^2/df = 1.76$, RMSEA = .059, RMR = .041). While the TLI (.798) and CFI (.812) were slightly below conventional thresholds, they were deemed acceptable for scales measuring complex psychological constructs. The internal consistency of the overall scale was high (Cronbach's $\alpha = .952$), and the reliability coefficients for each subdomain were also strong: empathetic clarification (.813), empathetic adaptability (.832), empathetic

acknowledgment (.858), deep empathic inquiry (.832), and emotional support provision (.875). To examine the relationship between GPT-generated diagnostic scores and the self-report scale results, a Pearson’s correlation analysis was conducted. Correlation coefficients (*r*) and their statistical significance (*p*-values) were calculated to determine whether the diagnostic outcomes produced by the GPT-based system were meaningfully associated with scores from the self-report empathy scale. The results demonstrated strong and statistically significant correlations ($r = .77-.95, p < .01$) across all five diagnostic subdomains, thereby supporting the concurrent validity of the EmotionBridge system.

3.4.2 Phase 4.2 Usability Evaluations

This phase aimed to assess the practical applicability of the GPT-based empathic communication diagnostic system by evaluating users’ actual experiences with the system and identifying areas for improvement based on user feedback, thereby enhancing the overall quality of the final product.

Phase 4.2.1 Participants. The participants involved in the usability evaluation were the same as those who participated in the concurrent validity testing. The use of an identical sample for both evaluations was a strategic decision intended to maintain consistency in research design and enhance the reliability of the findings. Utilizing the same participant group also maximized the efficiency of the research process by eliminating the need for additional recruitment, thereby minimizing potential sampling bias and reducing the risk of introducing new confounding variables. This approach helped ensure methodological coherence and contributed to the overall credibility of the evaluation outcomes.

Phase 4.2.2 Instruments and Data Collection. For the quantitative evaluation, the System Usability Scale (SUS) and the User Experience Questionnaire (UEQ) were utilized. The questionnaire included ten items measuring perceived ease of use, effectiveness, efficiency, and overall satisfaction with the system. Each item was rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). For the qualitative evaluation, participants were asked to

describe any usability issues or areas for improvement encountered while interacting with the system. In addition, they were instructed to submit usage logs and screenshots to document their experience with the EmotionBridge interface. Open-ended questions were also included to capture users’ perceptions of the system’s strengths, limitations, and suggestions for enhancement. These data were collected to generate deeper insights into the user experience and inform system refinement.

Phase 4.2.3 Data Analysis. Quantitative data were collected via an online survey and analysed using descriptive statistics, including mean (*M*) and standard deviation (*SD*), for each item to assess the system’s overall usability. SUS scores were converted to a 0–100 scale, where scores above 68 were interpreted as indicating above-average usability and scores above 85 as excellent usability (Bangor et al., 2008; Brooke, 1996).

Qualitative data were analysed using content analysis. Participants’ responses to open-ended questions were thematically categorized to identify user-reported strengths, usability challenges, and suggested improvements. Based on this analysis, actionable insights were derived to guide further refinement of the GPT-based empathic communication system.

4. RESULTS

4.1 Content Validity of the System

To assess the theoretical and structural validity of the EmotionBridge system, expert review was conducted in two domains: (1) GPT-based system design principles and (2) empathic communication diagnostic framework.

4.1.1 GPT-Based System Design Principles

Ten experts evaluated eight items related to the system’s architecture, usability, instructional logic, and technical appropriateness. CVI ranged from .80 to 1.00, with a mean CVI of .95. The inter-rater agreement (rWG) ranged from .76 to 1.00 (mean rWG = .83), indicating strong consensus among experts. Table 2 presents expert recommendations and subsequent revisions for the GPT-based system design principles.

Table 2: Expert Recommendations and Revisions for the GPT-Based System Design Principles

Evaluation Criteria	Expert Recommendation	Reflected	Implementation Detail
Clarity	The GPT design principles should be organized into a practical and logical sequence.	Applied	Added scenario structure and flow diagrams to clarify the diagnostic process.
Feasibility	“Include specific scenario flow examples for practical use.”	Applied	Included sample prompts and guidance for using key system features in educational settings.

4.1.2 GPT-Based System Design Principles

Seven items assessing the clarity, universality, interpretability, and instructional relevance of the

diagnostic principles were reviewed by a subset of experts specializing in communication, education, and psychology. The mean CVI was .97, and the

average rWG was .81, demonstrating high content validity and consistency across evaluators. Table 3 summarizes the expert recommendations and the

corresponding revisions made to enhance the Empathic Communication Diagnostic Framework.

Table 3: Expert Recommendations and Revisions for the Empathic Communication Diagnostic Framework

Evaluation Criteria	Expert Recommendation	Reflected	Implementation Detail
Item Specificity	Beyond recognizing emotional expressions, learners should demonstrate contextual interpretation of meaning.	Applied	Added "contextual meaning interpretation" to rubric items.
Clarity of Explanation	Diagnostic criteria should be communicated clearly and understandably to learners.	Applied	Included GPT feedback templates with example phrasing.
Practical Applicability in Schools	It would be helpful to include application cases in real school contexts.	To be addressed in subsequent research	Development of classroom-based application scenarios is planned in follow-up research.

4.2 Concurrent Validity

4.2.1 Agreement between GPT and Expert Ratings

To evaluate the reliability of the GPT-based empathic communication diagnostic system, five experts (F, G, H, I, J) who had previously participated in the content validity review independently assessed

actual conversation samples conducted between the GPT system and 35 participants. Based on these independent expert ratings, intraclass correlation coefficients (ICCs) were calculated to examine the degree of agreement between the system-generated diagnostic scores and the expert evaluations. The results are presented as follows.

Table 4: Agreement Between Expert Ratings and GPT-Based Diagnoses

Subdomain	Single ICC(95% CI)	Mean ICC(95% CI)	Interpretation
Empathic Clarification	.81(.72~.87), $p < .001$.91(.87~.95), $p < .001$	Excellent
Empathic Adaptability	.76(.66~.83), $p < .001$.87(.82~.92), $p < .001$	Excellent
Empathic Acknowledgment	.74(.63~.81), $p < .001$.86(.80~.91), $p < .001$	Good
Deep Empathic Inquiry	.69(.57~.78), $p < .001$.82(.75~.88), $p < .001$	Moderate to Good
Emotional Support Provision	.77(.67~.84), $p < .001$.81(.84~.93), $p < .001$	Excellent
Overall Average	.75(.71~.79)	.87(.85~.90)	Excellent

The results of the ICC analysis indicated moderate to high agreement between the GPT-based diagnostic system and expert evaluations across the diagnostic categories. In particular, the domains of empathetic clarification (ICC = .81) and emotional support provision (ICC = .77) demonstrated high ICC values, suggesting strong consistency between the GPT-generated diagnostic results and human expert judgments. However, the domain of deep empathic inquiry showed a comparatively lower ICC value, indicating that further refinement may be needed in this area. Experts generally evaluated the system as having a strong capacity for diagnosing empathic communication, with sufficient reliability for potential application in educational and counselling contexts. Specifically, in the empathetic clarification domain, GPT was noted to accurately recognize users' emotions and deliver appropriate empathic

responses. Expert J commented, "GPT expressed emotions with such naturalness and accuracy that it could be a useful support tool even in real counselling sessions." On the other hand, several experts noted that in the deep empathic inquiry domain, the system tended to produce formulaic and surface-level empathic expressions, particularly when addressing complex or nuanced emotional contexts. Expert H suggested, "In order for GPT to capture emotional depth and subtle contextual cues more effectively, further improvements are needed in prompt design and data training processes."

4.2.2 Correlation with Self-Report Empathy Scale

Pearson correlation analysis showed strong, statistically significant correlations between GPT-based diagnostic scores and the self-report empathy scale across all five subdomain

Table 5: Correlation Between GPT Diagnoses and Self-Report Empathy Scores

Subdomain	Correlation (r)	Significance (p)
Empathic Clarification	.868**	< .01
Empathic Adaptability	.868**	< .01
Empathic Acknowledgment	.930**	< .01
Deep Empathic Inquiry	.953**	< .01
Emotional Support Provision	.770**	< .01
Empathic Clarification	.868**	< .01

These findings support the concurrent validity of the EmotionBridge system as a diagnostic tool that aligns with established measures of empathic communication. The results of the correlation analysis suggest that the GPT-based EmotionBridge application can serve as an important tool for both the assessment and support of empathic communication skills. These findings provide a basis for the future integration of GPT models into empathy enhancement programs and emotional support systems. AI-based applications like EmotionBridge perform functions beyond simple conversation. They have the potential to sensitively detect users' emotional states and deliver empathic responses tailored to those emotions, thereby contributing to real-life emotional needs. These findings underscore the promise of using GPT-based systems as practical tools for promoting empathic competence and delivering emotionally responsive interactions in diverse settings.

4.3 Usability Evaluation Results

4.3.1 Quantitative Results

Quantitative analysis revealed a strong usability

rating for the EmotionBridge system. The average System Usability Scale (SUS) score was 79.64 ($SD = 9.24$), indicating high user satisfaction and ease of use. According to established interpretation guidelines (Bangor et al., 2008; Brooke, 1996), a SUS score above 68 is considered to reflect "good" usability, and a score above 85 is categorized as "excellent". Subdomain scores were also favorable. Among the ten items assessed using the SUS and the User Experience Questionnaire (UEQ), the highest ratings were found in overall satisfaction ($M = 4.29, SD = 0.95$) and ease of use ($M = 4.24, SD = 0.85$), demonstrating strong positive user perceptions in those areas. Meanwhile, the lowest score was observed in the effectiveness domain ($M = 4.18, SD = 0.91$), suggesting room for further enhancement of system effectiveness. These findings collectively support the conclusion that EmotionBridge offers a high level of usability and user satisfaction, meeting the practical requirements for implementation in educational or counselling contexts.

Table 6: Usability Evaluation Results for the EmotionBridge

Usability Dimension	Mean (M)	Standard Deviate(SD)
Ease of Use	4.24	0.85
Effectiveness	4.18	0.91
Efficiency	4.19	1.01
Overall Satisfaction	4.29	0.95
SUS	79.64	9.24

4.3.2 Qualitative User Feedback and System Improvements

To better understand user experiences with the GPT-based empathic communication diagnostic system (EmotionBridge), participants were asked to provide open-ended feedback regarding strengths, areas for improvement, and suggestions for additional features. Using content analysis, their responses were categorized into major thematic domains. Most users (82.9%) evaluated the system's naturalness of empathic feedback and appropriateness of emotional support as its primary strengths. More than half of the participants (57.1%) suggested adding nonverbal elements such as voice-based dialogue or emoticons to overcome the limitations of a purely text-based interface. A considerable portion of users (62.9%) also requested

greater scenario diversity and expanded role customization to enhance immersion during the diagnostic interaction. In addition, 51.4% of participants expressed a desire for more interactive engagement within the conversation flow. Some noted, "It would feel more natural if the system could initiate more questions," while others emphasized that "a stronger question-response mechanism is needed" to sustain meaningful dialogue. Moreover, 45.7% of respondents pointed out that the feedback was somewhat generic and recommended that the system offer more personalized and context-specific responses based on user dialogue. Finally, 42.9% noted difficulties in the initial stages of system use, suggesting that guidance for initiating conversations or automatic feedback analysis would improve accessibility.

Table 7: Results of the EmotionBridge Usability Evaluation for Empathic Communication

Theme	Example Participant	Frequency	Ratio
Naturalness of Empathic Feedback	"GPT's empathic responses felt natural, as if I were communicating with a real person."	29	82.90%
Need for Diverse Scenarios	"It would be easier to engage if there were more diverse scenarios (e.g., workplace, friends, etc.)."	22	62.90%
Request for Voice and Non-Verbal Features	"Text-based responses differ from real conversations. Adding voice could better convey emotions."	20	57.10%
Lack of Interactive Engagement	"It would feel more natural if the system could initiate more questions."	18	51.40%
Desire for Tailored Contextual Feedback	* "More precise and context-sensitive feedback suggestions would be helpful."	16	45.70%
Need for Easier Conversation Entry Points	* "There should be guidance for conversation starters and feedback initiation."	15	42.90%

In response to these suggestions, the research team immediately implemented the following

improvements to enhance the system's practical usability and user satisfaction.



The user is guided through a structured set of three questions to define the interpersonal relationship, the situation type, and the setting in which the conversation will occur.

Step 1 prompts the user to select a broad context, such as "daily life" or "school-related" situations.

Step 2 asks the user to define the interpersonal relationship and specify the topic of conversation. In this example, the user selects a "teacher-student" interaction involving a concentration issue during class.

Step 3 invites the user to choose a specific situational setting, such as a classroom, gym, or school club. Based on these selections, the system automatically generates a context-rich interaction scenario. In the displayed case, the student (Minsoo) is struggling to stay focused during lessons, and the user is assigned the role of the teacher (Mr. Park).

The interface provides emotionally supportive prompts encouraging the user to understand the student's emotional state and initiate a constructive conversation. This stepwise role- and context-driven approach enhances narrative realism and emotional engagement, allowing users to practice empathic-based dialogue in settings that closely mirror real-world challenges.

Figure 4: Enhanced Role and Scenario Selection Interface Reflecting System Improvements in EmotionBridge

As illustrated in Figure 4, the revised system interface introduces an enhanced scenario and role selection process aimed at increasing user immersion. Users are guided through a structured sequence of three questions to define the interpersonal relationship, the type of situation, and the contextual setting for the conversation. The interface presents a wider range of realistic scenarios, such as "daily life" versus "school-based" situations, including peer conflict, academic stress,

or teacher-student misunderstandings. The role assignment feature has also been refined to prompt users to explicitly identify their role in the conversation (e.g., "I am a teacher" or "I am a student"), ensuring narrative clarity and alignment with the selected context. This multi-phase flow, combined with guided prompts, helps establish emotional engagement and supports more authentic, context-sensitive empathic dialogue.



The first part of the figure shows how each subdomain includes up to six feedback statements directly linked to the user's utterances. These responses demonstrate real-time analysis of emotional sensitivity and communication quality. Specifically, the empathic acknowledgment section evaluates whether the user correctly identified, interpreted, and validated the speaker's emotional state, offering praise for paraphrasing and emotional mirroring. The deep empathic inquiry section assesses the depth and relevance of follow-up questions, emphasizing open-ended, reflective exploration. The emotional support provision section focuses on affirming language, practical suggestions, and emotional encouragement provided by the user.

The second part of the figure summarizes performance in each subdomain through numerical scores and provides additional feedback, including positive reinforcement, areas for improvement, and scenario-specific recommendations to guide future practice. This multi-layered feedback structure enables users to reflect more deeply on their empathic performance and receive developmentally targeted guidance, thereby supporting both diagnostic accuracy and formative growth.

Figure 5 presents a detailed evaluation report for each subdomain of empathic response—empathetic acknowledgment, deep empathic inquiry, and emotional support provision—based on user-generated dialogue.

Figure 5: Enhanced Feedback Output for Empathic Response Evaluation in EmotionBridge

Second, as illustrated in Figure 5, the feedback mechanism in EmotionBridge was substantially refined to deliver more detailed, context-sensitive, and pedagogically meaningful responses based on user utterances. The updated system now provides distinct feedback for each subdomain of empathic response—namely, empathetic acknowledgment, deep empathic inquiry, and emotional support provision—with up to six individualized feedback statements per item. These feedback statements are directly tied to users’ real-time inputs, reflecting fine-grained analysis of emotional cues and conversational content. Moreover, the feedback is diversified across multiple dimensions, including praise, emotional validation, concrete guidance, and time-specific support strategies. This multi-layered feedback structure enables users not only to understand the effectiveness of their empathic responses but also to develop more adaptive and reflective communication strategies in future scenarios. Overall, the redesigned feedback

mechanism exemplifies a more personalized and instructionally rich approach to empathy diagnostics. It effectively integrates performance-based evaluation with formative developmental guidance, thereby aligning with contemporary goals in social-emotional learning and AI-assisted education.

Third, as presented in Figure 6, the interface was revised to include onboarding prompts and a structured question–response framework designed to improve initial accessibility for first-time users. The system now guides users through early-stage interactions by initiating the dialogue with simple, context-relevant questions that promote self-expression and emotional engagement. This guided questioning model reduces entry barriers and encourages active, self-driven participation. As shown in the figure 6, the interaction is structured to support empathic expression from the outset, helping users feel emotionally supported as they engage in the diagnostic process.

상황:
초등학교 1학년 민호는 새로운 환경에 적응하기 어려워하고 있습니다. 3학년 선배 유진은 민호를 돕기로 했습니다. 민호는 학교가 너무 커서 길을 잃을까 걱정하고, 친구들과 잘 지낼 수 있을지 두려워합니다.

[대화 먼저 시작하세요]

민호(후배) 입장에서 먼저 말해 보세요! 😊

선배님, 저 요즘 몇 가지 일 때문에 학교생활에 적응하기 어려워요

유진(선배):

"아, 그래? 민호야, 무슨 일 때문에 힘든지 말해 줄 수 있을까? 😊 나도 처음에는 적응하기 힘들었거든. 네가 어떤 부분이 힘든지 알려주면 내가 도와줄게!"

민호의 걱정이나 어려움을 좀 더 자세히 이야기해 볼래요? 😊

환경이 바뀌고 무엇을 해야 할 지 모르겠어요...

유진(선배):

"그렇구나, 민호야. 새로운 환경이 많이 낯설고 어색할 수 있어. 😊 내가 학교를 더 편하게 느낄 수 있도록 도와줄게! 혹시, 지금 가장 걱정되는 게 뭐야? 예를 들면, 교실을 찾는 거나 친구들과 어울리는 것 같은 거?"

이야기해 줘서 고마워! ❤️

교실을 찾는 거나 새로운 친구들과 어울리는 것이 어려워요

The sections enclosed in curly braces on both sides illustrate a revised dialogue interface in EmotionBridge, which was designed to improve initial accessibility for first-time users by incorporating onboarding prompts and a structured question–response framework

In this example, the system initiates the interaction with a conversational prompt that invites the user—to express their concerns first. The dialogue then proceeds with context-sensitive responses from a senior student (Yujin), who uses follow-up questions and emotionally supportive language to explore the junior’s difficulties adjusting to a new school environment. The interface provides guidance through simple and relevant question stems (e.g., “What are you struggling with the most?”, “Can you tell me more?”), which help users articulate their emotions and situational challenges. This guided questioning model reduces entry barriers and encourages self-directed participation, while also facilitating empathic engagement.

As shown in the figure, the structure supports empathic expression from the outset and helps users feel emotionally supported as they move through the diagnostic process.

Figure 6: Enhanced Onboarding and Question–Response Interface for Empathic Dialogue Initiation in EmotionBridge

Fourth, as presented in Figure 7, the interface was enhanced with onboarding prompts and automated emotional analysis features to improve initial accessibility for first-time users. The system now guides users through the early stages of the dialogue using structured prompts, while incorporating visual and textual cues—such as facial emojis and

descriptive phrases (e.g., “nodding,” “pausing”)—to simulate nonverbal expression and emotional tone. These design improvements aim to reduce entry barriers for novice users and increase emotional engagement from the outset, allowing users to more naturally express their feelings and initiate empathic interactions

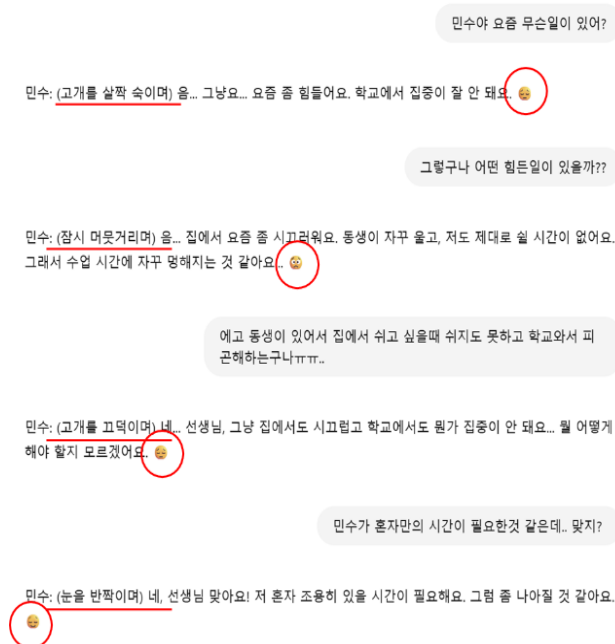


Figure 7 presents a sample dialogue from EmotionBridge that demonstrates the integration of nonverbal cues and emotional expression enhancements.

The underlined phrases (e.g., “lowering their head slightly,” “hesitating briefly,” “with a sparkle in their eyes”) simulate physical gestures and subtle emotional states through descriptive text. These cues are interwoven with verbal expressions to convey not only the speaker’s words, but also their underlying emotional posture, tone, and intention.

The circled elements in the image represent facial emojis placed at the end of user utterances. These serve to visually reinforce the speaker’s affective states—such as frustration, sadness, or reflection—enhancing emotional resonance within the interaction. Together, these multimodal features are designed to replicate nonverbal aspects of communication in a text-based environment, promoting richer emotional engagement and realism in simulated conversations.

Figure 7: Emotionally Augmented Interface with Nonverbal Cues and Guided Interaction in Emotion

5. DISCUSSION

This study aimed to develop and validate EmotionBridge, a GPT-based diagnostic system for empathic communication skills. The results provided strong empirical support for the system’s design validity, diagnostic accuracy, and practical usability, suggesting that AI-driven dialogue tools can serve as reliable alternatives to traditional methods for assessing social-emotional competencies.

5.1 Key Findings and System Evaluation

The findings of this study provide strong empirical support for the reliability, validity, and usability of EmotionBridge, a GPT-based diagnostic system for empathic communication. Expert review results confirmed high content validity for both the system design principles (CVI = .95) and the diagnostic framework (CVI = .97), with strong inter-rater agreement ($rWG > .80$). The integration of scenario-based interactions and rubric-aligned feedback mechanisms was found to be pedagogically appropriate and accessible for learners.

In addition, concurrent validity was demonstrated through two independent analyses. First, interclass correlation coefficients (ICCs) between GPT-generated diagnostic scores and expert evaluations showed moderate to high agreement across all diagnostic subdomains, with particularly strong alignment in empathetic clarification (ICC = .81) and emotional support provision (ICC = .77). Second, Pearson correlation analysis revealed significant correlations between the system’s diagnostic outputs and self-report empathy scale scores ($r = .77-.95$),

supporting the system’s alignment with established empathy measures.

The usability evaluation further reinforced the system’s educational potential. Participants rated the system highly across key usability domains, with an average SUS score of 79.64, which falls within the “good” usability range (Bangor et al., 2008; Brooke, 1996). Subdomain scores were similarly favorable, especially in overall satisfaction ($M = 4.29$) and ease of use ($M = 4.24$). These results suggest that EmotionBridge offers an intuitive, engaging, and pedagogically effective user experience. Moreover, users highlighted the system’s strengths in delivering emotionally resonant feedback and personalized interaction, aligning with prior research on the educational applications of GPT in social-emotional learning (Ayers et al., 2023; Maurya, 2024).

5.2 Educational Contributions, Practical Implications, and Future Directions

This study contributes to the intersection of artificial intelligence and affective learning assessment. EmotionBridge demonstrates how generative AI can deliver real-time, rubric-based, and personalized diagnostic feedback, addressing the limitations of conventional tools such as self-report surveys and expert interviews. The research also advances design-based methodologies by integrating expert validation, iterative design refinement, and NLP-driven scoring systems to produce replicable, scalable models for social-emotional learning.

In terms of practical application, the system offers valuable use cases across educational, counseling, and workplace settings. For educators, it supports the

development of reflective communication and scaffolds empathic behavior. For counselors and trainers, it serves as an accessible and cost-effective tool for formative empathy assessment. For developers and designers, the study offers a blueprint for incorporating large language models (LLMs) into ethically aligned, socially intelligent learning environments.

Despite its strengths, the study acknowledges several limitations. First, expert feedback suggested that GPT's responses in the domain of deep empathic inquiry sometimes lacked nuance or appeared overly formulaic. Second, the current system supports only text-based interaction, excluding critical nonverbal communication cues such as tone and facial expressions. Third, the diagnostic framework, although grounded in universal theories, has not yet been adapted to account for diverse cultural and linguistic contexts.

Future research should focus on developing multimodal extensions (e.g., voice, gesture), expanding scenario coverage to include clinical and intercultural communication contexts, and conducting longitudinal validations in real-world educational and therapeutic settings. Overall, EmotionBridge demonstrates the potential of GPT-based systems to transform the assessment and development of empathic communication skills, offering a practical and scalable solution for AI-enhanced empathy education.

6. CONCLUSION

This study developed and validated EmotionBridge, a GPT-based diagnostic system designed to assess empathic communication skills through real-time, scenario-based interactions. Applying a Design and Development Research (DDR) methodology, the system was built through theoretical modeling, expert validation, and iterative implementation using GPT Builder.

The results demonstrated that EmotionBridge achieved high content validity, strong concurrent validity with both expert evaluations and self-report measures, and high usability as reported by users. These findings affirm the system's capacity to deliver accurate, scalable, and learner-centered diagnostics for empathic communication—an area that has

traditionally been difficult to assess objectively.

Importantly, this study presents a practical and pedagogically grounded model for integrating generative AI into diagnostic tools for social-emotional learning. By combining large language models with domain-specific rubrics and personalized feedback mechanisms, EmotionBridge bridges the gap between technical innovation and educational relevance.

Future research is encouraged to address remaining limitations, including the absence of non-verbal communication cues, limited cultural adaptability, and the need for more nuanced emotional sensitivity. Ongoing development should explore multimodal extensions (e.g., voice and gesture-based input), context-specific scenario diversification, and longitudinal validation in authentic educational and therapeutic environments.

Ultimately, EmotionBridge marks a meaningful step forward in AI-enhanced empathy assessment and offers actionable insights for educators, counselors, and learning technology designers seeking to advance empathic competence in diverse, real-world settings.

Abbreviations

AI: Artificial Intelligence.

DDR: Design and Development Research

Declarations: Funding this research did not receive funding.

Competing interests:

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Ethical approval was not required for this study according to institutional and national guidelines for social science research in the Republic of Korea. The study involved voluntary participation and collected anonymized data without any personally identifiable information. All procedures involving human participants were conducted in accordance with relevant ethical standards and regulations. Informed consent was obtained from all participants prior to participation.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to ethical and privacy considerations but are available from the corresponding author on reasonable request.

REFERENCES

- Akpan, I. J., Kobara, Y. M., Owolabi, J., Akpam, A., & Offodile, O. F. (2024). An investigation into the scientific landscape of conversational and generative artificial intelligence, and human–chatbot interaction in education and research. *arXiv*. <https://arxiv.org/abs/2407.12004>
- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, *183*(6), 589–596. <https://doi.org/10.1001/jamainternmed.2023.1838>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology*, *28*(2), 91–100. <https://doi.org/10.1037/0022-0167.28.2.91>
- Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience*, *35*(1), 1–23. <https://doi.org/10.1146/annurev-neuro-062111-150536>
- Bialystok, L., & Kukar, P. (2018). Authenticity and empathy in education. *Theory and Research in Education*, *16*(1), 23–39. <https://doi.org/10.1177/147787851774664>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Brown, F., Smith, J., & Lee, A. (2022). Sleep problems and antisocial behavior in adolescents: The role of intervention strategies. *Journal of Adolescent Health*, *70*(3), 315–322.
- Chang, D. H., Lin, M. P. C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, *15*(17), 12921. <https://doi.org/10.3390/su151712921>
- Covey, S. R. (1989). *The 7 habits of highly effective people: Powerful lessons in personal change*. Free Press.
- Darrington, J., & Brower, N. (2012). *Effective communication skills: "I" messages and beyond*. All Current Publications.
- Davis, L. L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research*, *5*(4), 194–197. [https://doi.org/10.1016/S0897-1897\(05\)80008-4](https://doi.org/10.1016/S0897-1897(05)80008-4)
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
- Decker, S. E., Nich, C., Carroll, K. M., & Martino, S. (2014). Development of the Therapist Empathy Scale. *Behavioural and Cognitive Psychotherapy*, *42*(3), 339–354. <https://doi.org/10.1017/S1352465813000039>
- Elyoseph, Z., & Levkovich, I. (2023). Beyond human expertise: The promise and limitations of ChatGPT in suicide risk assessment. *Frontiers in Psychiatry*, *14*, 12131410. <https://doi.org/10.3389/fpsy.2023.1213141>
- Fuller, M., Kamans, E., van Vuuren, M., Wolfensberger, M., & de Jong, M. D. T. (2021). Conceptualizing empathy: A professional communication perspective. *Journal of Business and Technical Communication*, *35*(3), 333–368. <https://doi.org/10.1177/105065192110011>
- Ge, Y., Li, W., Chen, F., Kayani, S., & Qin, G. (2021). The theories of the development of students: A factor to shape teacher empathy from the perspective of motivation. *Frontiers in Psychology*, *12*, 736656. <https://doi.org/10.3389/fpsyg.2021.736656>
- Gonsalves, C., & Pearson, J. (Eds.). (2023). *Improving learning through assessment rubrics: Student awareness of what and how they learn*. IGI Global.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Huh, G. H. (2003). Construction and validation of a global interpersonal communication competence scale. *Korean Journal of Journalism & Communication Studies*, *47*(6), 380–408.
- Ioannidou, F., & Konstantikaki, V. (2008). Empathy and emotional intelligence: What is it really about? *International Journal of Caring Sciences*, *1*(3), 118–123.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*(1), 85–98. <https://doi.org/10.1037/0021-9010.69.1.85>

- Jang, H., Kim, E. J., & Reeve, J. (2012). Longitudinal test of self-determination theory's motivation mediation model in a naturally occurring classroom context. *Journal of Educational Psychology, 104*(4), 1175–1188. <https://doi.org/10.1037/a0028089>
- Kang, J. (2018). *Development and validity of communication competence scale of middle manager* (Doctoral dissertation, Kwangwoon University).
- Kim, M. (2011). *A study on the relativity between empathic communication and personality* (Master's thesis, Sun Moon University).
- Kim, H. S., & Ha, H. S. (2023). The effect of empathic ability on gender egalitarianism of high school students. *Journal of Future Oriented Youth Society, 20*(3), 51–76. <https://doi.org/10.34244/JFOYS.2023.20.3.3>
- Kolomaznik, M., Petrik, V., Slama, M., & Jurik, V. (2024). The role of socio-emotional attributes in enhancing human-AI collaboration. *Frontiers in Psychology, 15*. <https://doi.org/10.3389/fpsyg.2024.1369957>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine, 34*(5), 911–920. <https://doi.org/10.1017/S0033291703001624>
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815–852. <https://doi.org/10.1177/10944281062966>
- Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–386.
- Mamina, R. I., & Piraynen, E. V. (2023). Emotional artificial intelligence as a tool for human-machine interaction. *Discourse, 9*(2), 35–51. <https://doi.org/10.32603/2412-8562-2023-9-2-35-51>
- Maurya, R. K. (2024). Using AI-based chatbot ChatGPT for practicing counseling skills through role-play. *Journal of Creativity in Mental Health, 19*(4), 513–528. <https://doi.org/10.1080/15401383.2023.2297857>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Melchers, M., Montag, C., Markett, S., & Reuter, M. (2015). Assessment of empathy via self-report and behavioural paradigms: Data on convergent and discriminant validity. *Cognitive Neuropsychiatry, 20*(2), 157–171. <https://doi.org/10.1080/13546805.2014.991781>
- Merrill, M. D. (2020). *First principles of instruction*. Association for Educational Communications and Technology.
- Muhyatun, M. (2022). Transforming stress to happiness: Implementation of the I-message technique for establishing positive communication patterns in the family. *Indonesian Journal of Islamic Psychology, 4*(2), 18–35.
- Murphy, B. A., & Lilienfeld, S. O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? *Psychological Assessment, 31*(8), 1062.
- Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223–250. <https://doi.org/10.1080/15298860309027>
- Nikendei, C., Zeuch, A., Dieckmann, P., Roth, C., Schäfer, S., Völkl, M., & Jünger, J. (2005). Role-playing for a more realistic technical skills training. *Medical Teacher, 27*(2), 122–126. <https://doi.org/10.1080/01421590400019484>
- Park, J. (2020). A study on the evaluation methods of empathic communication competence. *Journal of CheongRam Korean Language Education, 73*, 7–31.
- Plomp, T., & Nieveen, N. (2013). *Educational design research*. Netherlands Institute for Curriculum Development (SLO).
- Qin, T., Wei, P., & Zhu, C. (2024). Research on the influencing factors of subjective well-being of Chinese college students based on panel model. *Frontiers in Psychology, 15*, 1366765. <https://doi.org/10.3389/fpsyg.2024.1366765>
- Richey, R. C., & Klein, J. D. (2007). *Design and development research*. Lawrence Erlbaum Associates.
- Reigeluth, C. M. (Ed.). (1999). *Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2). Lawrence Erlbaum Associates.
- Rogers, C. R. (1975). Empathic: An unappreciated way of being. *The Counseling Psychologist, 5*(2), 2–10. <https://doi.org/10.1177/001100007500500202>

- Rosenberg, M. B., & Chopra, D. (2015). *Nonviolent communication: A language of life* (3rd ed.). PuddleDancer Press.
- Sohn, S. (2022). *Development and validation of the nonviolent communication scale* (Doctoral dissertation, Sahmyook University).
- Son, W. S., Chae, M. K., Hwang, D. U., Nah, K., Kim, M., Kim, J. H., & Lee, J. (2025). Social contact patterns in South Korea: An analysis of a survey conducted in 2023–2024. *BMC Infectious Diseases*, *25*, 295. <https://doi.org/10.1186/s12879-025-10706-y>
- Sunahara, C. S., Rosenfield, D., Alvi, T., Wallmark, Z., Lee, J., Fulford, D., & Tabak, B. A. (2022). Revisiting the association between self-reported empathy and behavioral assessments of social cognition. *Journal of Experimental Psychology: General*, *151*(12), 3304–3322.
- Sze, J. A., Gyurak, A., Goodkind, M. S., & Levenson, R. W. (2012). Greater emotional empathy and prosocial behavior in late life. *Emotion*, *12*(5), 1129–1140.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Aldine Publishing Company.
- Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*(2), 135–149. <https://doi.org/10.1177/10731911145679>
- Wachs, S., Castellanos, M., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2023). Associations between classroom climate, empathy, self-efficacy, and countering hate speech among adolescents: A multilevel mediation analysis. *Journal of Interpersonal Violence*, *38*(5–6), 5067–5091. <https://doi.org/10.1177/08862605221120905>
- Waxman, S. R. (1983). Nonverbal bases of social categorization in preschool children. *Child Development*, *54*(6), 1541–1548. <https://doi.org/10.2307/1129824>
- Wolvin, A. D., & Coakley, C. G. (1988). *Listening* (3rd ed.). William C. Brown Publishers.
- Xiang, D., Qin, G., & Zheng, X. (2022). The influence of student–teacher relationship on school-age children’s empathy: The mediating role of emotional intelligence. *Psychology Research and Behavior Management*, *15*, 2735–2744. <https://doi.org/10.2147/PRBM.S380689>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, *16*, 39. <https://doi.org/10.1186/s41239-019-0171-0>

APPENDIX A SYSTEM DESIGN PRINCIPLES FOR EMOTIONBRIDGE

The following table outlines the core design elements, corresponding principles, and implementation examples used to develop the EmotionBridge system using GPT Builder.

Design Element	Design Principle	Example of Configure Implementation
Scenario-Based Dialogue	<input type="checkbox"/> Generate realistic and multi-layered scenario prompt templates.	- You are [a counselor/teacher/friend, etc.].
	<input type="checkbox"/> Clearly define roles in the prompt so that both the user and GPT engage in role-specific dialogue.	- The other person is experiencing [sadness/anxiety/conflict].
Diagnostic Algorithm	<input type="checkbox"/> Structure multi-layered prompts to assess empathy elements such as emotional recognition, feedback appropriateness, and goal achievement in dialogue.	- Use follow-up prompts to deepen the user’s emotional elaboration:
	<input type="checkbox"/> Analyze user input – including emotions, keywords, and sentence patterns – to evaluate empathic components.	“Could you explain more about why you felt that way?”
	<input type="checkbox"/> Use GPT Builder’s follow-up question feature to continue deeper dialogue.	“Can you describe your feeling more specifically?”
	<input type="checkbox"/> Fine-tune GPT Builder with datasets related to empathic communication to improve diagnostic precision.	- Display example: Empathic acknowledgment: 18/24 points
	<input type="checkbox"/> Set evaluation criteria and scoring rules for GPT-generated responses.	- Provide explanation for each score and offer at least three improvement suggestions based on the rubric.
	<input type="checkbox"/> Use GPT Builder’s summary/report function to generate automated feedback.	
Dialogue Customization (GPT Builder)	<input type="checkbox"/> Present diagnostic outcomes as both quantitative scores and qualitative feedback.	
	<input type="checkbox"/> Use the persistence function to maintain conversational context and continuity in empathic responses.	- Maintain consistency by recalling prior user input:
	<input type="checkbox"/> Set tone and mood in the GPT Builder to reflect various empathic communication scenarios.	“Earlier, you mentioned [user’s feeling]. Let’s continue building empathy based on that.”

APPENDIX B DIAGNOSTIC PRINCIPLES FOR EMPATHIC COMMUNICATION IN EMOTIONBRIDGE

The table below summarizes the theoretical diagnostic framework that guides the evaluation of empathic communication in the system. The framework distinguishes between empathic expression and empathic response, with detailed criteria for each subdomain.

Empathic Process	Subdomain	Definition	Indicators and Evaluation Criteria
Empathic Expression	Empathetic Clarification	The ability to clearly and accurately deliver one’s message.	- Clear purpose
			- Concise and direct language
	Empathetic Adaptability	The ability to flexibly adjust one’s expression according to the listener and the situation.	- Emotion-Need-Request structured expression
			- Appropriate word choice
Empathic Response	Empathetic Acknowledgment	The ability to accept and explicitly acknowledge the speaker’s emotions through summarizing or confirmation	- Adjusting conversational flow
			- Response-based adaptation
			- Explicit expressions of understanding
	Deep Empathic Inquiry	The ability to ask exploratory questions to better understand the speaker’s emotional experience.	- Emotional reflection
			- Justified acknowledgment
			- Use of open-ended, exploratory, and reflective questions
Emotional Support Provision	The ability to offer emotional stability and support to the speaker.	- Positive and affirming responses	
		- Expressions of encouragement	
		- Concrete support suggestions	