Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #47576

AI-Driven Multimodal System for Enhancing Non-Verbal Communication in Public Speaking

Dr. Brainerd Prince, Plaksha University

Brainerd Prince is the Associate Professor of Practice and the Director of the Center for Thinking, Language and Communication at Plaksha University. He teaches courses such as Reimagining Technology and Society, Ethics of Technological Innovation, and Art of Thinking for undergraduate engineering students and Research Design for PhD scholars. He completed his PhD on Sri Aurobindo's Integral Philosophy from OCMS, Oxford – Middlesex University, London. He was formerly a Research Tutor at OCMS, Oxford, and formerly a Research Fellow at the Oxford Centre for Hindu Studies, a Recognized Independent Centre of Oxford University. He is also the Founding Director of Samvada International Research Institute which offers consultancy services to institutions of research and higher education around the world on designing research tracks, research teaching and research projects. His first book The Integral Philosophy of Aurobindo: Hermeneutics and the Study of Religion was published by Routledge, Oxon in 2017. For more information, please visit: https://plaksha.edu.in/faculty-details/dr-brainerd-prince

Siddharth Siddharth, Plaksha University

Dr. Siddharth is the Director of Human-Technology Interaction Lab at Plaksha University. He recently came to Plaksha from the Human Factors Center of Excellence at Microsoft USA where he worked as Senior Human Factors Researcher. He is passionate about working on research themes with practical applications in the technology industry.

At Plaksha, Dr. Siddharth is investigating and modeling human physiological responses to improve cognitive well-being, and enhance human safety, productivity, and creativity. By integrating domain-specific knowledge from the fields of signal processing, computational neuroscience, computer vision, and machine learning, his research group is utilizing human physiological responses to track various cognitive states such as emotions, stress, awareness, mental load, and more. His research and recommendations have practical applications in the technology industry for the development of safer and more user-centered designs that promote cognitive well-being.

In the past, Dr. Siddharth led the Microsoft Human Factors Center of Excellence's efforts to design personalized user experiences through neurophysiological insights into users' preferences enabling them to be more productive and establish meaningful connections with each other. He has also worked at Facebook Reality Labs, Samsung Research America, French National Center for Scientific Research (CNRS), and National University of Singapore. His research has been awarded fellowships and research grants from various institutions including the National Science Foundation, Army Research Lab, and the Kavli Institute of Brain and Mind. It has also been published in many international conferences and reputed journals including IEEE Transactions and Nature Scientific Reports and has spawned multiple patents.

Dr. Siddharth received his Ph.D. from the University of California San Diego where he also served as the Graduate Fellow Editor of the Prospect Journal of International Affairs and Graduate Student Member of the South Asia Initiative. His first book, Founding Generations: Democracy's Origins and Parallels in America and India, was released recently. For more up-to-date information about him, please visit https://ssiddharth.in/about-me

Subham Jalan
Hibah Ihsan Muhammad, Plaksha University, Punjab
Chaitanya Modi

AI-Driven Multimodal System for Enhancing Nonverbal Communication in Public Speaking

Abstract

This work-in-progress paper introduces an AI coach to train engineering students on nonverbal communication. Engineering students often find it difficult to combine verbal and nonverbal cues in their interactions to communicate their ideas effectively. Conventional pedagogical approaches in engineering place emphasis on spoken communication while ignoring the crucial function of nonverbal cues, including head movements, body language, and facial emotions. This study proposes a personalized AI-driven coaching system designed to teach nonverbal communication skills to bridge the communication gap in the engineering community. Leveraging state-of-the-art multimodal AI models, our methodology integrates a comprehensive pipeline to evaluate nonverbal cues—namely, head pose, eye contact, and facial expressions, as well as verbal cues—namely, speech transcript. For head pose analysis, we utilize the MediaPipe library to classify head orientations. Facial expression evaluation relies on models trained on the FER-2013 dataset, with dominant emotions extracted and normalized from video frames. Verbal communication is assessed using OpenAI's Whisper for transcription and Meta's Llama3 for context-aware evaluation. Unlike traditional classroom settings, this approach integrates advanced AI to provide personalized, real-time feedback, bridging the gap between technical and interpersonal competencies. Our AI system's performance was evaluated against human evaluators for assessing students' communication skills and providing personalized feedback. On the Likert scale, our AI system outperformed human evaluators by 24.25% in clarity, comprehensiveness, personalization, and specificity of the feedback.

Keywords

AI-driven Coaching, Nonverbal Communication, Pedagogy, Machine Learning

Introduction

The value of effective communication cannot be emphasized enough in the rapidly changing landscape of engineering education. In addition to solving problems, engineers are creative thinkers and leaders who need to express their thoughts clearly and firmly. However, a recurring issue undermines this ideal: many engineering students struggle to effectively synthesize verbal and nonverbal forms of communication. People's inability to communicate their ideas, despite their technical proficiency, stems from the systematic neglect of traditional pedagogical frameworks for nonverbal communication curricula within engineering programs.

It is plausible that a final-year undergraduate student presenting their research may have exemplary technical content, and yet, their delivery may lack coherence. For instance, the technical expertise of engineering students is overshadowed by their inability to communicate effectively. The speaker may fail to engage with the audience by avoiding eye contact, making minimal gestures, and holding a rigid posture. Despite being accurate, the verbal information might not be supported by nonverbal clues such as kinesics and voice modulation. This gap can make the presentation seem disjointed and unconvincing.

Body language and facial expressions are examples of nonverbal communication that are crucial in enhancing spoken material. Studies conducted by Schneider & Aburumman, talk about improving audience engagement, building credibility, and improving message retention [1] [2]. Yet, engineering curricula predominantly focus on verbal articulation—structured arguments, technical jargon, and precise language—while relegating nonverbal elements to the background. This pedagogical gap leaves students unprepared for professional environments where technical expertise must be complemented by persuasive communication.

Our solution to this gap is to create a coaching system powered by artificial intelligence (AI) that teaches engineering students nonverbal communication techniques. Our method assesses crucial non-verbal characteristics like speech patterns, head posture, and facial expressions by utilizing state-of-the-art multimodal technology. Through personalized feedback, it helps users to refine their communication skills in a manner that aligns with their technical orientation. Unlike traditional classroom methods, this AI-based solution offers an innovative, scalable, and engaging alternative that resonates with the digital-first mindset of engineering students.

This paper presents the development and evaluation of our AI system as a potential pedagogical aid, contributing to a broader rethinking of communication pedagogy within technical education. We analyze the interaction between verbal and nonverbal communication and illustrate how AI-driven solutions can help bridge the communication gap faced by engineering students. The subsequent section provides a comprehensive literature review to contextualize this approach.

Literature Review

Communication is a process involving verbal, nonverbal, and written components, each playing a unique role. Verbal communication serves as the basis for information exchange by using language to express ideas and intentions. However, nonverbal communication, such as body language, endearing vocals, gestures, and facial expressions, enhances spoken messages by adding emotional and contextual richness. Together, these modalities create more effective and impactful interactions. A similar duality is presented by Buck and VanLear, in their study on verbal and nonverbal communication [3]. Cultural standards also mold nonverbal communication styles and determine what is considered suitable in social situations. For instance, Goffman described the ways in which people coordinate their actions to establish culturally expected norms for one another [4]. Further, in cross-cultural contexts, misreading these signs might lead to misunderstandings. Different cultures have unique ways of expressing respect, emotions, and intentions, which may not always align with the norms of another culture.

Gestures play a critical role in both verbal and nonverbal communication, enhancing and clarifying the meaning of messages For instance, Duncan observed that if a speaker held a gesture in midair while pausing, no change in speakership would occur, even when various relinquishing behaviors were exhibited—the "turn-suppressing" gesture in effect cancelling out the meaning or effect of the other behaviors [4]**.** When there are inconsistencies between attitudes communicated verbally and through body posture, the postural component should dominate in determining the total attitude that is inferred. A lot of information may be conveyed verbally but in a face-to-face conversation, body language and facial expressions can have an incredible impact on how that information is interpreted.

Unfortunately, students in engineering schools frequently prioritize technical knowledge over the subtleties of nonverbal communication [5]. This oversight may make it more difficult for them to work across disciplinary and cultural borders. Additionally, according to a study conducted by Vilasini and Paul, students from rural backgrounds face barriers such as lack of resources and limited access to communication coaches, which hinder their speaking skills and confidence [6].
Unfortunately, students in engineering schools frequently prioritize technical knowledge over the subtleties of nonverbal communication [5]. This oversight may make it more difficult for them to work across disciplinary and cultural borders. Additionally, according to a study conducted by Vilasini and Paul, students from rural backgrounds face barriers such as lack of resources and limited access to communication coaches, which hinder their speaking skills and confidence [6].

Furthermore, following English-language curriculum for non-native English speakers makes it difficult for them to understand the subtleties of verbal and non-verbal communication in the English language. As per Sarangi and Bai conventional classroom-based approaches to teaching communication skills often fall short, especially when it comes to addressing the difficulties that non-English speakers face. [7] [8]

Among the millions of engineering students that graduate from universities around the world every year, many struggle with communicating in English, especially when it is not their native language [7] [9]. Only a small percentage of these graduates possess the English language skills necessary for highly sought-after jobs in sectors like sales, consulting, and software development. Even when they can read and write in English, many students still struggle with fluency, proper pronunciation, syntax, and sentence structure when speaking the language. This gap in spoken English proficiency limits their employability in a competitive job environment where proficiency in effective communication is essential.

A study on 'Robust speech recognition via large-scale weak supervision' showed that traditional engineering pedagogy largely neglects nonverbal communication, creating a gap in students' holistic skill development [10]. Engineering curricula prioritizes technical subjects, leaving students with limited language and communication training. Moreover, students from rural areas often face greater communication challenges than those in metropolitan areas or premier institutions. Classroom methods focus on written language, overlooking nonverbal cues. Students rarely experience scenarios like interviews or presentations, leaving them unprepared for professional demands.

AI offers a powerful solution to bridge the above communication gaps for engineering students, particularly in areas of nonverbal communication that are often neglected in traditional education. Bai in his research on 'application of artificial intelligence in communication network' highlights AI's potential to enhance communication networks, a concept that can be adapted for skill development [8]. Other researchers have also emphasized how AI can provide personalized feedback and foster a willingness to communicate, which is especially beneficial for non-native speakers [11] [12]. A study by Cruickshank et al. stresses the need for flexible, adaptable support, which AI can effectively provide [13]. AI-driven coaching systems utilizing multimodal analysis offer tailored, real-time feedback that reduces learning anxiety and fosters a more engaging and productive learning environment [14] [15]. Additionally, students view AI-based systems as both engaging and effective, particularly for offering feedback that aligns with their individual needs [16]. However, addressing concerns related to anxiety and trust in AI is crucial to maximize its potential, boosting user confidence, and ultimately improving communication skills for engineering students [17].

In the last few years, tools powered by AI have demonstrated potential in aiding neurodivergent individuals, especially those with Autism Spectrum Disorder (ASD), develop effective communication skills. For instance, a virtual conversational agent, assists adolescents with ASD by providing a safe space to rehearse social interactions and gives feedback on verbal and nonverbal cues [18] . Another example is the VOISS [19] platform from the University of Kansas, which uses virtual reality to create authentic school-based scenarios for students to practice social skills, these tools are useful for users who may struggle with sensory overload or anxiety in traditional settings because they can work through drills at their own pace.

To address these concerns and train engineering students effective verbal and non-verbal communication, we built an end-to-end AI system. The AI system was presented to student participants in an experiment in which they received personalized feedback about their communication skills. This feedback was then compared to the one received by trained human evaluators. Below we present the methodology for building and evaluating our AI system.

Methodology

State-of-the-art AI models of different audio-video domains were utilized and aggregated to create our AI system for nonverbal and verbal communication evaluation.

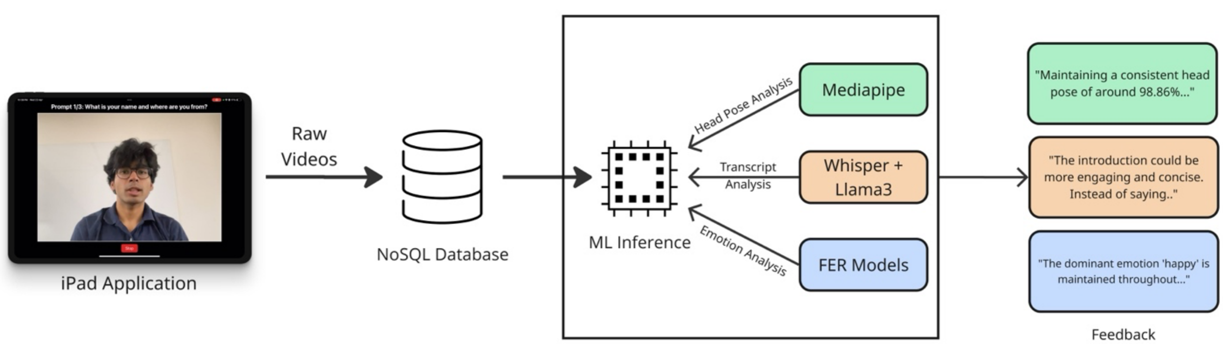


Fig.1. The system uses an ML-integrated pipeline for communication evaluation.

For extracting information about the head pose, we are using MediaPipe [20]. It gives access to

478 face landmarks (as shown in Fig. 1. below). We extract the landmarks corresponding to the nose tip, inner and outer eye corners, and mouth corners. Perspective-n-Point computation was used to estimate the head pose by mapping 3D landmark coordinates onto their 2D projections. We computed the camera matrix and used it to calculate the translation and rotation vectors representing the orientation of the face. Furthermore, pre-defined thresholds were used to classify every frame in the corresponding head pose. After classifying the frames for the entire video, the values are normalized to get the head pose proportions.

For transcript evaluation, OpenAI's Whisper [10] was used to extract the transcript from the audio file. Post that, a carefully engineered prompt was utilized to get evaluation results from the Large Language Model (LLM). Owing to its small size and competitive results, Meta's Llama3 [21] was used as LLM. Moreover, the evaluation task does not require perplexing NLP computations.

Fig.2. Facial landmarks (in white) and associated emotions (top-right) provided by the respective AI models. Participant gave their consent to use their image for the manuscript.

Finally, for evaluating facial expressions, models trained on the FER-2013 dataset were used. The state-of-the-art models in this field have an accuracy of about 73%. This motivated us to consider video averages instead of individual frames in emotion analysis. To extract the emotion displayed in videos at an interval of 100 milliseconds, frequency of each emotion was normalized to get the two main dominant emotions displayed in the self-introduction video. After compiling the individual results of each of these components, the result is parsed and formatted to be presented to the user (see Fig. 1). The project was developed and deployed on an iPad to get users' self-introduction videos and provide feedback. All the computation is run on an ML Inference server where pre-defined thresholds are used to determine communication statistics. The server integrates the outputs of different ML models and uses pre-defined thresholds to determine communication statistics. The feedback is sent to the users via e-email. Additionally, as per Kapp in his study 'gamification of Learning and Instruction' gamification techniques such as progress tracking and reward systems can significantly enhance user engagement, making learning more interactive and motivating for students [22].

Results

The evaluation of our AI-driven multimodal system for enhancing nonverbal communication skills was conducted using data collected from 11 engineering students aged 18-22, comprising 5 women and 6 men. Each participant recorded a video in response to prompts designed to simulate an interview setting, which included three self-introduction questions:

1. What is your name and where are you from?
2. Please share about what you're professionally doing now.
3. What are your hobbies and interests?

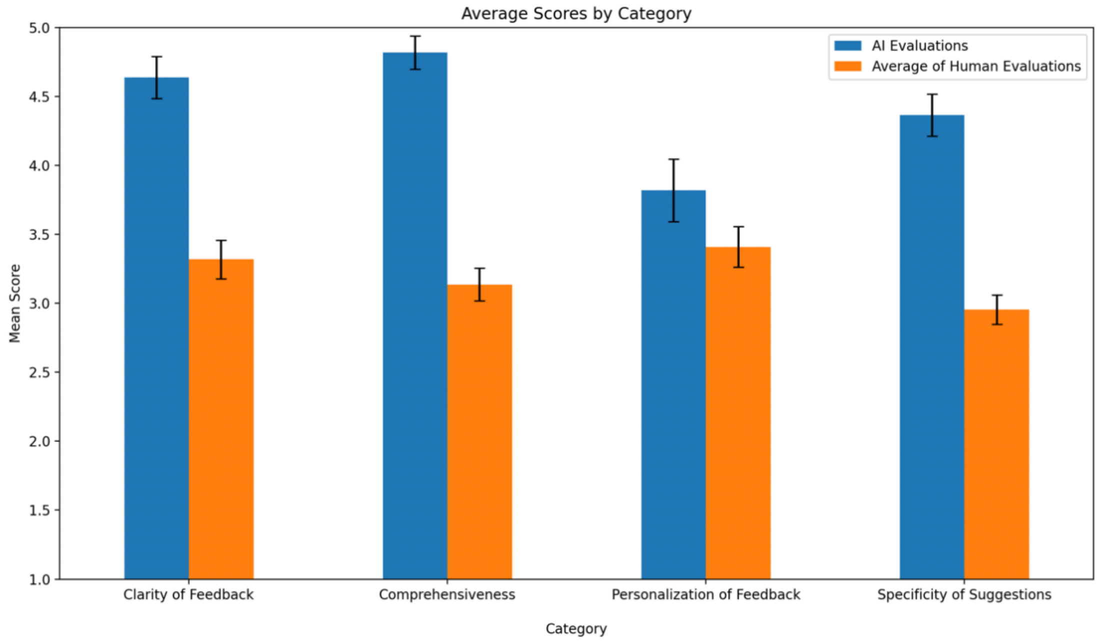
The recorded videos were then analyzed using our AI model, which provided feedback on their nonverbal communication skills. To validate the AI-generated feedback, we collaborated with two academic professors specializing in communication and language. They independently evaluated the same videos and provided feedback.Their assessments served as a benchmark for comparing the AI-generated feedback.
To objectively assess the feedback quality, we developed a structured rubric, evaluated on a Likert scale (1 to 5), across the following four dimensions:

1. Clarity: How easy it is to understand the feedback.
2. Specificity of Suggestions: How actionable and detailed the suggestions are.
3. Comprehensiveness: How well the feedback covers all relevant aspects of nonverbal communication.
4. Personalization of Feedback: How well the feedback adapts to the individual's performance.

Five independent evaluators rated the feedback blindly for each participant, unaware of whether it was generated by AI or human evaluators. Care was taken to remove percentages and numbers that would not be possible for a human evaluator to compute. The aggregated scores across the participants were used to compare the effectiveness of AI and human evaluation. Since there were two human evaluators, the average of their scores was used in our analysis. Across all four evaluation dimensions, AI consistently outperformed human evaluators. The average scores for AI and human evaluations are summarized below:

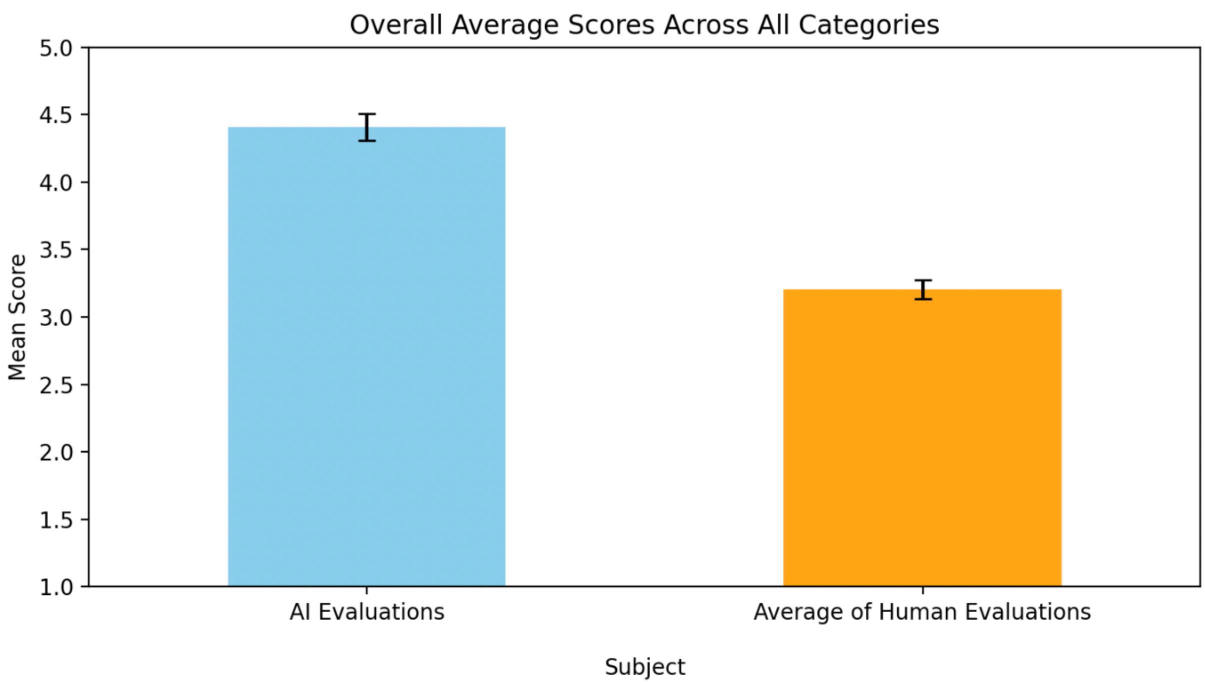
- Clarity: AI (4.64) vs. Humans (3.32)
- Specificity of Suggestions: AI (4.36) vs. Humans (2.95)
- Comprehensiveness: AI (4.82) vs. Humans (3.14)
- Personalization: AI (3.82) vs. Humans (3.41)

Graph 1: The bar chart below illustrates the average scores for each category.



The overall average scores across all categories reinforce the trend.

- AI Evaluations: 4.41
- Human Evaluations: 3.20



Graph 2: A comparison of the overall scores highlights the significant advantage of AI-driven feedback.

Potential reasons why AI performed better than human evaluators

- Provides objective feedback without confirmation bias.
- Delivers detailed insights through actionable feedback on nonverbal cues.
- Offers fair assessments by evaluating performance against a consistent baseline, not comparing users.
- Has better memory retention compared to humans.

Discussions

Our AI-powered coaching system demonstrates significant potential to improve nonverbal communication skills among engineering students. According to the evaluation results, the AI system performed 24.25% better than human evaluators in terms of giving detailed, individualized, precise, and understandable feedback. This discussion focuses on the limitations of our approach and suggests areas for future research.

Limitations and Challenges

Despite the promising results, several important limitations must be acknowledged:

- *Limited Sample Size:* The evaluation was conducted with a small sample size of 11 engineering students. This may not be representative of the diverse backgrounds and communication styles among the engineering students generally.
- *Risk of Social Isolation:* If AI-based communication training is used excessively, it may result in social isolation. For engineering students, who already struggle in social communication situations, a technology-mediated communication style devoid of human interaction may unintentionally limit prospects for genuine social learning.
- *Lack of Longitudinal Assessment:* The study does not track how effectively students implement the AI-generated feedback over time, making it difficult to assess the long-term impact of the system on communication skill development.

Future Work

Building on our findings and addressing the limitations identified, we propose several promising directions for future research:

- *Additional Communication Modalities:* We aim to incorporate more modalities such as voice modulation, tone analysis, and gesture recognition to provide a more comprehensive evaluation of communication skills. These additional data points will create a more holistic assessment of nonverbal communication.
- *Accessible Interface for Neurodivergent Users:* The system will be enhanced to be more inclusive by making the interface intuitive and easy to understand for neurodivergent people. This includes implementing customizable sensory settings (adjustable visual

contrasts, sound levels), clear and concise instructions with visual aids, alternative feedback formats (visual, audio), and progress visualization tools that accommodate different cognitive processing styles. These adaptations will ensure that the AI coaching system is accessible and beneficial to people across the neurodiversity spectrum, catering to the communication challenges they may face in both academic and professional settings.

Conclusion

Effective communication includes both verbal and nonverbal components. However, engineering students frequently find it difficult to combine these aspects. Our system leverages state-of-the-art tools to give users personalized feedback by combining insights from different modalities. This creative strategy fits with engineering students' preference for digital learning and provides a scalable and interesting substitute for traditional classroom-based approaches. Therefore, this research underscores the potential of AI in bridging the communication gap within the engineering community. By integrating technical expertise with interpersonal effectiveness, the system equips students to excel in collaborative and professional settings. Future work will focus on training a unified model to combine head pose, emotion, and transcript data for feedback, enhancing the system's generalization and impact. Additionally, more modalities such as voice modulation and tone will be incorporated.

References

[1] S. Schneider, F. Krieglstein, M. Beege and G. D. Rey, "The impact of video lecturers' nonverbal communication on learning – An experiment on gestures and facial expressions of pedagogical agents," vol. 176, p. 104350, 1 January 2022.

[2] N. Aburumman, M. Gillies, J. A. Ward and A. F. d. C. Hamilton, "Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions," vol. 164, p. 102819, 2022.

[3] R. Buck and C. A. VanLear, "Verbal and Nonverbal Communication: Distinguishing Symbolic, Spontaneous, and Pseudo-Spontaneous Nonverbal Behavior," vol. 52, no. 3, pp. 522-541, 1 September 2002.

[4] S. Jones and C. Lebaron, "Research on the Relationship Between Verbal and Nonverbal Communication: Emerging Integrations," *Journal of Communication,* vol. 52, pp. 499 - 521, September 2002.

[5] E. Gözüyeşil, "An Analysis of Engineering Students' English Language Needs," vol. 116, pp. 4182-4186, 21 February 2014.

[6] V. Ashokan and X. J. Paul, "ENHANCING SPEAKING SKILLS FOR ENGINEERING STUDENTS HAILING FROM RURAL BACKGROUND," *International Journal of Mechanical Engineering and Technology,* vol. 10, no. 1, p. 1002–1006, 2019.

[7] "LinkedIn," 5 February 2021. [Online]. Available: https://www.linkedin.com/pulse/english-really-problem-engineering-education-smruti-sarangi/.

[8] Z. Bai, "Research on Application of Artificial Intelligence in Communication Network," *Journal of Physics: Conference Series,* vol. 2209, p. 12014, February 2022.

[9] P. Verma and S. Bhattacharya, "How English-speaking skills are keeping Indian engineers away from their dream jobs," 7 August 2015. [Online]. Available: https://economictimes.indiatimes.com/jobs/how-english-speaking-skills-are-keeping-indian-

engineers-away-from-their-dream-jobs/articleshow/48384078.cms?from=mdr.

[10] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever, 6 December 2022. [Online]. Available: http://arxiv.org/abs/2212.04356.

[11] A. Dolunay, "The use of artificial intelligence in the field of communication: A research on the perspectives of communication academics," *Journal of Autonomous Intelligence ,* vol. 7, no. 5, p. 1610, 21 May 2024.

[12] F. Huang and B. Zou, "English speaking with artificial intelligence (AI): The roles of enjoyment, willingness to communicate with AI, and innovativeness," *Computers in Human Behavior,* vol. 159, p. 108355, 2024.

[13] K. CRUICKSHANK, S. NEWELL and S. COLE, "Meeting English Language Needs in Teacher Education: a flexible support model for non-English speaking background students," *Asia-Pacific Journal of Teacher Education,* vol. 31, no. 3, pp. 239-247, 2003.

[14] P. Shahidi, M. Amini Farsani and S. A. Mosayebi, Teaching Engineering Students English Using Artificial Intelligence (AI) in Comparison with Traditional Methods, 2023.

[15] S. Aniruddh and T. Ankit, "The Impact of Artificial Intelligence on Technical Communication in 2023," vol. 8, no. 12, pp. a944-a953, Decemeber 2023.

[16] T. T. U. Duong and M. H. Nguyen, "Students' Perspective on Applying AI in English Learning and Teaching – A Study at UEH," vol. 5, pp. 91-102, 12 September 2024.

[17] M. N and P. K. N. S, "Investigating ESL Learners' Perception and Problem towards Artificial Intelligence (AI) -Assisted English Language Learning and Teaching," *World Journal of English Language,* vol. 13, no. 5, p. 270, 7 April 2023.

[18] M. a. C. D. a. J. L. a. B. A. a. P. J. a. R. R. a. H. E. Ali, "LISSA — Live Interactive Social Skill Assistance," 2015, pp. 173-179.

[19] M. Mosher Ph.D., "The Social Validity and Efficacy of a Virtual Reality Intervention for Improving Middle School Students' Social Communication: A Randomized Controlled Study," 2023.

[20] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg and M. Grundmann, "MediaPipe: A Framework for Building Perception Pipelines," 14 June 2019. [Online]. Available: http://arxiv.org/abs/1906.08172.

[21] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle and A. Letman, 23 November 2024. [Online]. Available: http://arxiv.org/abs/2407.21783.

[22] K. Kapp, The gamification of learning and instruction: Game-based methods and strategies for training and education. San Francisco, CA: Pfeiffer, 2012.