Journal of

Advanced Psychology

(JAPSY)

Artificial Intelligence and Multimodal Technologies in Psychology: A Scoping Review of Speech, Language, Gesture, Communication, and Human-Computer Interaction Research: (2021- 2026)

CARI Journals

Journal of Advanced Psychology
ISSN: 2791-3244 (Online)
Vol.8, Issue No. 2, pp 87 - 106, 2026
www.carijournals.org

Artificial Intelligence and Multimodal Technologies in Psychology: A Scoping Review of Speech, Language, Gesture, Communication, and Human-Computer Interaction Research: (2021- 2026)

[1]Tabinda Khawaja Sarfraz, [2]*Abbas Rashid Butt

[1]Special Education Teacher, Hollifield Station Elementary School, Howard County Public School System, Maryland, USA

[2]Research Fellow, Lahore Institute for Research and Analysis, University of Lahore, Pakistan

https://orcid.org/0000-0002-1768-7051

Accepted: 7th July, 2026, Received in Revised Form: 20th July, 2026, Published: 29th July, 2026

Abstract

Purpose: This scoping review systematically captures and describes the existing research in psychology how AI and multimodal technologies are used for speech, language, gesture, communication, and human - computer interaction (HCI). It also analyzes the research on these issues published between January 2021 and June 2026.

Methodology: The review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) and the Joanna Briggs Institute (JBI) Population–Concept–Context (PCC) framework. The electronic search was initiated in the Scopus database mainly focusing on identifying Scopus-indexed peer-reviewed journal articles with Psychology and related cross-disciplinary areas.

Findings: The findings reveal that a great number of interdisciplinary research projects focusing on AI-enabled communication technologies were conducted during the period under the review. Multimodal learning, deep learning, machine learning, computer vision, natural language processing, transformer architectures, large language models, affective computing, speech recognition, gesture recognition, emotion recognition, conversational AI, and adaptive intelligent systems were the most commonly mentioned technologies in the literature.

Unique contribution to theory, practice and policy: The study gives policymakers the insight that standardized benchmarking of the communication AI systems, open and transparent AI governance systems, data privacy measures, inclusive and equitable datasets, and regulations of Ethical AI development and deployment should be made priority in policy guidelines on the use of AI communication systems in communication. This scoping review has one of its primary contributions to the literature through the provision of an excellent map of AI multimodal communication technology evidence between 2021 and 2026. The study, from a theoretical point of view, enriches the existing knowledge by bringing AI-related work together with psychological communication educational, health rehabilitation accessibility, and HCI evidence in a singular system.

Keywords: *Artificial Intelligence, Multimodal Artificial Intelligence, Human-Computer Interface, Recognition of Speech, Technologies for Assisting People with Disabilities*

1. INTRODUCTION

Artificial intelligence (AI) has emerged as one of the most transformative technological inventions of the twenty-first century, fundamentally changing how humans communicate, collaborate, learn and interact with digital system (Yang, 2022; Ma & Liu, 2023; Feng et al., 2025; Liu et al., 2026). Recent developments in Machine Learning (ML), Computer Vision (CV), and Deep Learning (DL) have empowered machines to perform tasks which used to be solely done by human beings (Zhu et al., 2020; Górriz et al. 2023; Jiang et al., 2024). These machines can understand, analyze, and respond to human communication with greater precision and depth than ever before. Besides, the rise of large language models (LLMs) like ChatGPT and GPT-4 further enables the generation of responses in a natural human language (Feng et al., 2025; Liu et al., 2026). Contrary to traditional computers that operate only based on rigid logical rules and predefined algorithms, AI nowadays processes a vast range of inputs simultaneously and draws information from them in the same way that humans perceive and reason through different sensory channels simultaneously (Hassani et al., 2024; Jiang et al., 2024; Giberga et al., 2025). That means, AI systems are now able to understand and interpret not only words but also spoken language, facial expressions, and gestures etc.

In fact, this ability has brought about such a dramatic shift that many of human needs and desires can now be addressed by machines without human involvement (Chaturvedi et al., 2026; Cheng, 2026). In short terms, this means that people don't have to rely on other people to get certain types of services anymore like health care, education or even entertainment, etc. In sum, the emergence of multimodal AI is a major turning point since human communication, in its essence, is multichannel. When we interact with others, we combine spoken words, vocal pitch, facial expressions, bodily movement, gesturing, and so on to be understood as we wish (Hassani et al., 2024; Jiang et al., 2024; Giberga et al., 2025; Seaborn et al., 2025). Recent achievements in multimodal machine learning, deep networks for cross modal representations, multimodal models, and vision-language models have given machines the ability to understand and create more accurate semantic contexts and react better to users than those that work in one modality only (Zhu et al. 2020; Liu et al. 2026; Feng et al., 2025; Liu et al. 2026; Xiao et al., 2025; Quan et al., 2026; Kristina et al., 2026). As a result, such machines provide a great deal of support in facilitating natural conversations, adaptive interaction, one-on-one communication, etc. not mentioning emotional and mental aspects of communication (Seaborn et al. 2025; McStay & Bakir, 2025).

Communications in humans is done majorly by speech language gesture, and non-verbal cues. That means, research has been done on topics that include automatic speech recognition, speech synthesis, language understanding, emotion recognition, gesture recognition, sign language interpretation, facial expression analysis, and multimodal communication systems that are capable of interpreting human behaviors (Ma & Liu, 2023; Terkourafi & Hansen, 2026). These technologies are now an ever-present part of our lives, from providing us with healthcare-related support to educational evaluation and even assistive communication. The technologies include

Speech and Language Therapy, Educational Assessment Rehabilitation Assistive Communication, Tele-Health, Accessibility Services, Conversational Agents, Social Robotics, Virtual Assistants, Augmented Reality (AR), Virtual Reality (VR), and Extended Reality (XR) Learning Environments (Liu et al. 2026; Alsaleh, 2026; Wu et al. 2025; Kayis et al., 2026). A few studies have been devoted to analyzing human communication through multimodal sensors such as speech, gestures, facial expressions, eye tracking, physiological monitoring, and wearable sensors to give intelligent systems better understanding of human intentions and thereby provide customized assistance (Dwivedi et al. 2022; Hassani et al., 2024; Jiang et al., 2024; Giberga et al., 2025).

Multimodal technologies incorporating AI that provide human-computer interaction (HCI) have undergone a major change. Before, input devices like keyboard, mouse, and graphical user interface were the main means of interaction in HCI. However, contemporary systems that incorporate human gestures, speech, and facial expressions bio signals wearable devices, and context awareness for natural user interaction have increasingly taken the lead. Jiang et al. (2024) and Seaborn et al. (2025), and others have highlighted that multimodal user input is the current trend in HCI. Several new applications of multi-modal technologies have also emerged, such as Voice assistants, intelligent tutoring systems, conversational agents, collaborative robots, mixed reality, immersive XR platforms etc. (McStay & Bakir, 2025; Liu et al. 2026).

The idea behind emotional Artificial Intelligence (Affective AI) is to make computers not only respond functionally but also show emotions (like sadness, happiness, etc.) to users which makes humans to feel understood. A very recent study in the field of conversational AI has shown emotional responses of these systems as users' emotions (Quan et al., 2026). These capabilities of emotional response from AI go a long way in helping us get the emotional support from AI systems while also having to face challenges in the work environment with the use of computers. At the same time, AI has become one of the major driving factors for innovation in assistive and inclusive communication systems.

2. LITERATURE REVIEW

2.1 Conversational Intelligence and Language Technology

Recently, there has also been a dramatic expansion of foundation models and generative AI through which conversational intelligence and language technology have changed completely. Big models able to create text that are contextually adequate fluent in multiple languages, and also combining text/images are at the core of virtual assistant programs, AI-powered educators, AI scribes in medical contexts, academic activities translation programming, AI friends, and content generation of all kinds (Xiao et al., 2025; Quan et al., 2026; Kristina et al., 2026; Feng et al., 2025; Liu et al., 2026). With all that, new powers are added through methods like RL (reinforcement learning), transfer learning, self-supervised learning, joint multimodal pretraining, and foundation model designs that can take advantage of say a speech recording, a written message as well as a photo, a video, or just a set of contextual cues (Zhu et al., 2020; Górriz et al. 2023; Liu et al. 2026). This rapid progression has broadened AI's involvement in those domains dominated or at least heavily involving communication while opening up areas of innovation like customized learning, smart support, accessibility services, and dynamically adjustable interactions. On a positive note of these technological achievements, research nowadays is pretty much scattered among a bunch of areas like AI, machine learning, speech and language, pathology, psychology, education, rehabilitation sciences, HCI, cognitive science, accessibility studies, linguistics ethics and technology governance (Dwivedi et al. 2022).

Many existing research projects focus only on single modalities of communication, isolated techniques of AI, or single domains to a given study thereby preventing a deep understanding of the ever-changing picture of AI working with multimodal input (Hassani et al., 2024 ; Jiang et al., 2024; Giberga et al., 2025). Meanwhile, a greater part of scholarly work lately has dealt with the subject of regulation, trust in AI openness interpretation of algorithms, privacy, protection of rights, and international legislation at a level of abstraction that indicates a desire to see the development of new technologies being coupled with a strong ethical check (McStay & Bakir, 2025; Cheng, 2026; Oruma, 2026). Also, in educational aspects, a lot of the work focuses on the use of a combination of different modes of communication, the idea of trans language, and ways to cater a wide variety of students through the method of teaching. Though, in most instances, they do not rely on the use of AI for the assessment of the students (Bailey et al. 2025; Turner & Tour, 2025).

2.2 Use of Multimodal AI

Hospital and school were not the only sectors that have found the use of the multimodal AI. Now, it plays a role also in areas like customized learning, computer-assisted teaching, immersive XR (such as VR and AR) scenarios, technologies for making things accessible to people of all abilities and ways to visualize information. It has turned out in these domains too that one of the most prominent ways of AI implementation now is the combination of human machine interaction technologies with AI, for example, the development of the voice interfaces, the gesture recognition systems with speech, the contextual adaptation. Of course, in different cases, some of the methods are better or worse, and there is no common line across different disciplines (Liu et al. 2026). Parallel research into how humans communicate has looked into probabilistic reasoning, understanding of text and images, communication intent recognition, psychometric assessment of voice interaction, and others. So, a very broad variety of algorithmic approaches to intelligent communication are being developed nowadays from the computational perspective (Elzer et al. 2011; Seaborn et al. 2025). A multidimensional effort combining technological insights, educational practices, medical and clinical understanding, social implications as well as regulatory structures to create a clear, comprehensive, and unified map of multimodal communication technologies, is essential after the emergence of the AI age, if not before.

Journal of Advanced Psychology
ISSN: 2791-3244 (Online)
Vol.8, Issue No. 2, pp 87 - 106, 2026 www.carijournals.org

2.3 Scoping Review

Because of the great diversity and interdisciplinarity of this rapidly growing body of evidence, one might argue that scoping reviews are the right methodological strategy for a systematic and in-depth inquiry of heterogeneous research that can cover different kinds of study designs, different populations technologies communication channels, and application fields (Arksey & O'Malley, 2005; Peters et al. 2020; Tricco et al. 2018). While conventional systematic reviews focus mainly on the effectiveness of interventions, scoping reviews can offer a more general overview of the existing evidence, highlight emerging topics, define the key concepts, describe the methods, and point out the areas of knowledge that are not covered or are in an early stage of development over various scientific disciplines (Peters et al. 2020). Taking into account these methodological advantages, the current scoping review follows both the method structure of the Joanna Briggs Institute (JBI) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline tailored for scoping reviews (PRISMA-ScR). Also, an upgraded Population-Concept-Context (PCC) system is utilized to cover all AI-supported communication technologies reported in the scientific literature from January 2021 to June 2026, (Peters et al. 2020; Tricco et al. 2018).

This scoping review is intended to chart the current state of the art of AI-driven communication technologies across various applications such as healthcare, education, rehabilitation, accessibility, assistive technologies, human, computer interface, extended reality, intelligent tutoring systems, conversational AI, emotive computing, governance, and multilingual communication. By drawing together high quality research articles, it not only presents a synthesis of the AI techniques themselves but also explores the different ways of communication, the various domains where those technologies are being used, the computational approaches, the design of the studies, the type of individuals targeted by the studies, the technological features, the ways in which they are evaluated, the ethical dimension, the governance model, as well as the major directions for future research.

2.4 Objectives

Through conducting this scoping review, the authors will be able to map and synthesize the literature on AI and multimodal technologies as these refer to speech language gesture, communication, and human-computer interaction in the period between January 2021 and June 2026. The review exactly aims to:

1. To delineate the scope of artificial intelligence approaches that have been used in research related to the domains of speech, language, gesture, communication, and human-computer interaction, ranging from machine learning and deep learning to natural language processing, computer vision, multimodal learning, large language models, generative AI, reinforcement learning, and related intelligent systems.

2. To explore the multimodal communication modalities and interaction channels considered by researchers, among which speech, voice, text, gestures, facial expressions, eye motion, body movement, physiological signals images video, wearable sensors, and environmental sensors.

3. To outline the broad categories of application areas in which AI and multimodal technologies have been developed and deployed like healthcare, education, rehabilitation, accessibility, assistive technologies, human-computer interaction, extended reality (XR), intelligent tutoring systems, conversational agents, social robotics and governance systems.

On top of that, this review raises what comes next set of research questions, namely:

RQ1. What are the sorts of AI technologies that have been used in studies that dealt mostly with language, gesture, speech, communication and other aspects of human-computer interface, from 2021 down to 2026?

RQ2. What is the nature and combination of different types of multimodal communication channels and input streams like speech, text, gestures, facial expressions eye-gaze videos, physiological signals, and sensors across the studies in the sample?

RQ3. Which are the various fields of application into which AI-driven multimodal communication systems have been successfully integrated like healthcare, education, rehabilitation, accessibility, assistive technologies, human-computer interaction, extended reality, or other similar domains of human activity?

3. METHOD

3.1 PRISMA-ScR Setup

This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Scoping Review extension (or PRISMA-ScR) (Tricco et al., 2018). PRISMA-ScR is for identifying, screening, and summarizing research evidence about artificial intelligence (AI), multimodal communication technologies and human-computer interaction (HCI). In addition, the Joanna Briggs Institute (JBI) (Peters et al. 2020) was was. The review followed five sequential stages; defining review aim and questions; developing inclusion and exclusion criteria using PCC framework; conducting a systematic literature search; screening title, abstract and full texts; and synthesizing and interpreting the findings. The PRISMA-ScR flow framework guided the screening process for 92 potential records. Duplicate articles were discarded; 34 studies met all eligibility criteria and were included in narrative synthesize and qualitative assessment.

3.2 Population-Concept-Context (PCC) Framework

The Population-Concept-Context model was used plus guidelines of the Joanna Briggs Institute for developing inclusion and exclusion criteria (Peters et al. 2020).

3.2.1 Population

The population was intentionally defined to include children, adolescents, adults, and older people. Studies involving augmentative and alternative communication (AAC) or text-to-speech devices or other AI-assisted communication technologies were included. The review also included stakeholders involved in developing, implementing, or using AI-enabled communication systems, including students, teachers, health-care workers, speech-language pathologists, rehabilitation specialists, researchers, and technology developers, AI experts, professionals and the general public. This broad population reflects the expending application of AI across education, healthcare, accessibility, intelligent tutoring systems, chatbots, adaptive assistive technologies, and human-computer interaction.

3.2.2 Concept

The review focused on Artificial Intelligence (AI) and multimodal technologies that support communication and human interaction across diverse domains. These concepts included Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), Large Language Models (LLMs), Generative AI, Computer Vision, Multimodal and AI including texts, speech, facial expressions, body movements, gestures, and other visual information. The review also examined technologies supporting speech recognition, text-to-speech (TTS), gesture recognition, sign language translation, facial expression analysis emotion recognition, gaze tracking and multimodal data fusion. In addition, studies invloving Virtual Reality (VR), Augmented Reality (AR), Extended Reality (XR), wearable AI technologies, intelligent tutoring systems and smart environments were included. Ethical dimension such as Explainable artificial intelligence (XAI), Responsible AI, AI governance, transparency, fairness, accountability, privacy.

3.2.3 Context

The review adopted a broad interdisciplinary context to reflect the diverse applications of AI and multimodal communication technologies. Studies conducted in human-computer interaction (HCI), psychology, healthcare, speech therapy, education, accessibility, and assistive technologies were included. The review also considered research conducted in digital communication environments, virtual and extended reality (VR/XR) platforms, intelligent robotics, smart environments, and home or community-based systems supporting personalized interaction and adaptive communication. Finally, studies addressing AI ethics and governance, transparency, fairness, accountability, privacy, explainability, regulatory structure, and responsible AI development were included.

3.3 Search Strategy

A thorough review of published literature was undertaken focusing on studies published between January 2021 and June 2026. The search strategy was designed to comprehensively

Journal of Advanced Psychology
ISSN: 2791-3244 (Online)
Vol.8, Issue No. 2, pp 87 - 106, 2026 www.carijournals.org

identify relevant studies while ensuring scientific rigor and reproducibility. Controlled vocabulary and free-text keywords were combined to capture the rapidly evolving literature on AI and multimodal communication technologies.

3.4 Information Sources

The literature search was conducted exclusively in the Scopus database because of its extensive coverage of peer-reviewed journals, rigorous indexing standards, international reach, and multidisciplinary scope. To ensure methodological and disciplinary relevance, the search was restricted to peer-reviewed journal articles indexed in the Scopus Psychology subject areas. This approach ensured coverage of psychological, cognitive, communicative, behavioral, and human-centered dimensions of AI across education, rehabilitation, healthcare, accessibility, and human-computer interaction while maintain a transparent and reproducible screening process.

3.5 Eligibility Criteria

Selection criteria were established prior to the literature search.

3.5.1 Inclusion Criteria

Studies meeting PCC-based eligibility criteria were included. Only peer-reviewed journal articles published between January 2021 and June 2026 were considered eligible.

3.5.2 Exclusion Criteria

Articles failing to satisfy the predefined criteria were excluded. Conference papers, proceedings, editorials, letters, commentaries, book chapters, books, dissertations, theses, patents, technical reports, and unpublished manuscripts were excluded. Likewise, studies not indexed in Scopus, published outside the review period, or written in languages other than English were excluded.

3.6 Screening Process

The study selection process followed PRISMA-ScR guidelines to ensure transparency, systematic screening and reproducibility. After searching the Scopus database, 92 records were identified and imported into references management software to detect duplicates. Eight duplicates' records were removed, leaving 84 unique studies for screening. Screening proceeded in three stages, titles and abstracts of 84 records were screened against the PCC-based eligibility criteria. Fifty-eight studies were excluded, leaving 46 articles for full-text assessment. During full-test screening, methodological quality, conceptual relevance, and application domain were assessed. Twelve studies were excluded because they lacked sufficient methodological or conceptual detail, did not adequately address AI-based communication or multimodal interaction, fell outside the review scope, or failed to satisfy the inclusion criteria. Ultimately, 34 Scopus-indexed peer-reviewed journal articles satisfied all eligibility criteria were included in the qualitative synthesis.

3.7 Data Extraction

Before data extraction, a standardized data extraction form data was developed (see Appendix A), and applied consistently across all included studies to ensure consistency, accuracy, and transparency. The extraction form collected descriptive and methodological information including; Authors, year of publication, country of study, study objectives, study design, participant characteristics, sample size, AL methodology (ML, DL, NLP, LLMs, Computer vision, Generative AI), Multimodal inputs (speech, text, gesture, facial expression, eye gaze, video, sensors), communication or assessment task, application domain, dataset used, performance measures, and key findings.

4. RESULTS

4.1 PRISMA Flow Diagram

A complete search of Scopus identified 92 records between January 2021 and June 2026. The duplicate records were removed and the remaining studies were screened by title and the abstract using the predefined Population-Concept-Context (PCC) eligibility criteria (Figure 1). Studies unrelated to AI-assisted communication, multimodal communication technologies, speech, language, gesture communication, accessibility or human-computer interaction were excluded. The remaining articles underwent full-text assessment for eligibility.

4.2 Characteristics of the Included Studies (n = 34)

The final data set comprised 34 Scopus-indexed journals articles published between 2021 and June 2026. The studies represented diverse disciplines, including artificial intelligence, multimodal technologies, communication, psychology, education, healthcare, and human-computer interaction. Research designed included experimental studies, observational research, computational model development, qualitative research, mixed-methods studies, systematic scoping reviews, narrative reviews, theoretical papers, conceptual analyses, and AI governance studies.

The reviewed studies investigated AI-supported communication technologies which range from machine learning to deep learning, natural language processing, computer vision, multimodal AI, conversational agents, large language models (LLM), affective computing, intelligent virtual assistants (IVAs), assistive technologies, virtual and augmented reality (VR/AR), and AI governance models. Target populations included children with developmental disorders, multilingual learners, adults, clinicians, educators, rehabilitation engineers, AI users, teachers, healthcare professionals, and members of the general public. Several Studies focused exclusively on computational models or conceptual frameworks without human participants, reflecting the rapid evolution of AI-assisted communication technologies.

Journal of Advanced Psychology
ISSN: 2791-3244 (Online)
Vol.8, Issue No. 2, pp 87 - 106, 2026

4.3 Publication Trends (2021-2026)

The trend of academic publishing for this period (2021-2026) reveals the growing interest in AI-enabled communication technologies. A notable surge in publication activity during 2023 and onwards with the highest number of publications released in 2025 and 2026. These results suggest that there is great and expanding research interest in multimodal artificial intelligence, large language models, conversational systems, affective computing, accessibility technologies, extended reality, and responsible AI.

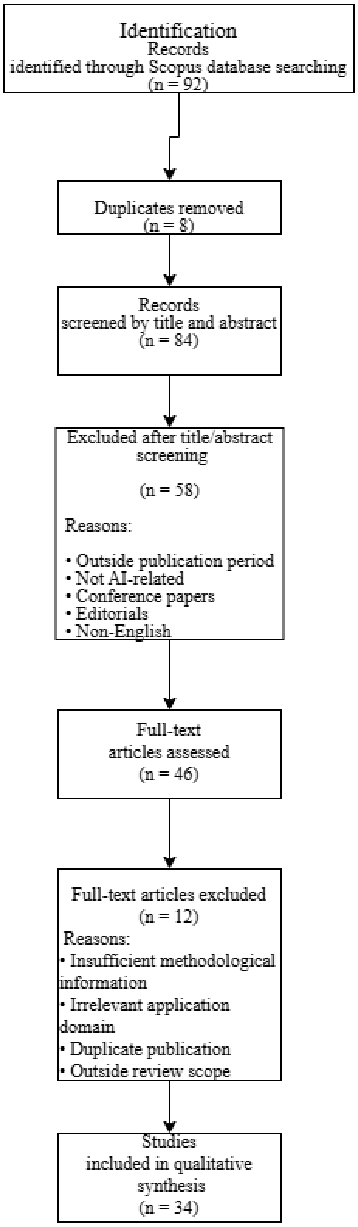

Figure 1. PRISMA-ScR Flow Diagram

4.4 Country-wise Distribution

The collected works have covered different continents showing globally spread interests on research for AI. Among them America accounted for most of papers, China, UK, Germany, Japan, Australia, Saudi Arabia, Canada, Spain, Italy, Netherlands, and some multinational collaborations followed. Publications by authors from North America and Europe constitute a main part of total publications. Asian countries also demonstrated increasing research activity in multimodal AI, intelligent educational technologies, assistive communication, and conversational systems. International collaboration was common, particularly studies addressing AI ethics and human-centered-AI.

4.5 AI Techniques

The reviewed studies demonstrated considerable diversity in AI techniques. Machine learning and deep learning were identified as two main approaches, particularly for speech recognition, gesture analysis, computer vision, and multimodal data fusion. Natural language emerged as another major research area. Its growth reflects the widespread adoption of conversational agents, intelligent tutoring systems, dialogue systems, language-based AI applications. Recent studies increasingly incorporated large language models (LMs), GPT-based systems, multimodal foundation models, and generative AI to support conversational interaction, adaptive, learning, intelligent communication. Computer vision research focused on facial expression recognition, sign language translation, gesture recognition, emotion detection, and visual scene understanding. Some studies also incorporated affective computing, eye tracking, reinforcement learning, explainable AI, wearable sensing, and multimodal transformer models to improve contextual understanding and adaptive communication. Theoretical and policy-oriented studies emphasized responsible AI, transparency, explainability, accountability, algorithmic auditing, ethical governance, and user trust rather than algorithm development.

4.6 Multimodal Inputs

The reviewed studies showed that AI systems increasingly integrate multiple communication modalities to improve understanding, interpretation, and response to complex human behavior. Rather than rely on a single input, theses systems combined multiple data sources to enhance contextual understanding, interaction quality, and decision accuracy.

The most common inputs included speech, voice, natural language text, facial expressions, hand and body gestures, posture, eye-gaze, images videos, sign language, motion capture, wearable sensor data, physiological signals, environmental sensing, touch- interaction, and contextual behavioral information. Their integration enabled more accurate and human-centered communication. Speech and voice were the most frequently used modalities particularly in conversational AI, voice assistants, speech recognition, intelligent tutoring systems, and communication support technologies. Spoken language remained the primary communication channel across most application domains. Natural language text supported dialogue management, text generation, and conversation through large language models, while visual information, including facial expressions, body gestures, and eye movements, enhanced emotions recognition, behavioral analysis, accessibility, and human-computer integration. Several studies adopted multimodal sensing by combining images, video data, wearable devices, physiological measures (e.g. heart rate, skin conductance), environmental sensing, and contextual user behavior. These approached were widely applied in healthcare, rehabilitation, emotion recognition, VR/XR environments, assistive technologies, education systems, and accessibility solutions.

Recent studies increasingly combined speech, text, facial expressions, gesture recognition, eye tracking, and contextual sensing to develop intelligent conversational agents, adaptive learning environments, immersive XR applications, assistive communication technologies, and human-computer interaction systems. This transition from single-modal to multimodal processing represents a major technological advancement, enabling AI systems to become increasingly context-aware, adaptive, personalized, and aligned with human-centered design principles.

4.7 Application Domains

The included studies showed that AI-enabled multimodal communication technologies which are widely applied across communication, education, healthcare, and digital interaction. The leading application area was Human-Computer Interaction (HCI) where AI enhanced human-machines interfaces through speech, language, gestures, facial expressions and eye gaze. Conversational AI and voice user interfaces were among the most frequently investigated topics. Studies examined virtual assistants, dialogue systems, chatbots, and voice-based interaction to improve communication efficiency, user satisfaction, and inclusive design.

Healthcare was another major application area, including clinical communication, speech technologies, rehabilitation, assistive communication devices, and AI-assisted diagnosis and therapy. Many studies explored multimodal AI for communication assessment, patient engagement, personalized rehabilitation, emotion recognition, clinical decision support. Education also emerged as a prominent domain, with research focusing on intelligent tutoring systems, AI-assisted language learning, and personalized educational environments integrating speech, language, and non-verbal communication to enhance learning outcomes. Numerous studies addressed accessibility and assistive technologies. AI supported speech-to-text, text-to-speech, sign language recognition, eye-gaze interfaces, wearable assistive devices, and multimodal communication platforms improved independence, digital inclusion and social participation for individuals with communication, sensory, cognitive, or physical disabilities.

Studies on Extended Reality (XR), Virtual Reality (VR), and Augmented Reality (AR) integrated AI with multimodal interfaces to create adaptive and immersive learning and interaction environments. Several studies also addressed smart environment, human-robot interaction, emotional AI, AL governance, ethics, transparency, explainability, privacy and responsible AI.

Emotion-aware AI emerged as growing research are aimed at enabling more natural, empathetic, and personalized human-AI interaction. Overall, healthcare, education, assistive technologies, and human-machine interaction were dominant application domains, demonstrating growing role of AI in developing accessible, adaptive, and human-centered communication systems.

4.8 Target populations

The reviewed studies focused on children and adolescents particularly those with Developmental Language Disorder (DLD) and Autism Spectrum Disorder (ASD), using AI for communication, assessment, language development, behavioral analysis, and personalized intervention. Other studies investigated multilingual learners, school-aged children university students using intelligent tutoring systems, voice interfaces, conversational AI, and human-computer interaction (HCI) technologies. Teachers, clinicians, speech-language therapists, rehabilitation specialists, health professionals, and researchers were also frequently included to evaluate the implementation, usability, acceptance, and effectiveness of AI-based communication technologies. Adults represented another major participant group, particularly in the studies of conversational agents, voice user interfaces, multimodal interaction, affective computing, accessibility technologies, and human-computer interaction systems. A few studies also aimed at older adults for assessing AI-based communication technologies that promote independence, accessibility, and digital inclusion.

4.9 Performance Measures

The review studies employed diverse performance measures reflecting the multidisciplinary nature of AI. Experimental and computational studies primarily used traditional AI evaluation metrics including accuracy, recall, F1-score, Area Under the ROC (AUC), classification accuracy, and confusion matrices to evaluate speech recognition, language processing, gesture recognition, language processing, gesture recognition, emotion detection, and multimodal classification. Studies also assessed communication effectiveness using response time, system latency, gesture recognition, accuracy, and speech recognition performance. Psychometric studies emphasized measurement validity and reliability through Cronbach’s Alpha, Composite Reliability (CR), Average Variance Extracted (AVE), Confirmatory Factor Analysis (CFA), RMSEA, CFI, TLI, and SRMR. These measures were primarily appled in studies evaluating user acceptance, voice quality, communication assessment, and other human-centered outcomes. Qualitative studies emphasized participants’ experience, perception trust, engagement, usability, ethical concerns, and implementation challenges, providing valuable insights into the real-world adoption of AI-enables communication technologies.

Journal of Advanced Psychology
ISSN: 2791-3244 (Online)
Vol.8, Issue No. 2, pp 87 - 106, 2026
www.carijournals.org

5. DISCUSSION

5.1 Evolution of AI from Single-Modality to Multimodal Intelligence

Earlier AI technologies relied on speech recognition or image analysis, limiting human natural communication, which is inherently multimodal. Human communication integrates gestures, verbal language, vocal cues, facial expressions, eye contact, body movements (Levy & McNeill, 1992). Consequently, modern AI systems increasing employ multi-dimensional machine learning models that integrate multi-sensory inputs improve human-machine interaction. Studies indicate that multimodal AI systems are more adaptive because they cmbine multiple communication channels to interpret emotions and respond more accurately. Integrating speech recognition with facial expression, body movements, and vocal cues provide richer contextual information for emotions recognition. Similarly, multimodal transformer models and foundation models combine text, voice, vision behavioral to support higher level reasoning and learning (Radford et al. 2021). New multimodal foundation models extend these reasoning across text, speech, images and video. They demonstrate grater generalization, adaptability and contextual understanding. These advances support natural human-computer dialogue, context-based decision-making, personalized assistance, and adaptive learning across healthcare, education, accessibility, and immersive environments.

5.2 Expansion of Human-Computer Interaction

Human-computer interaction has become one of the principal areas of AI-based communication technologies. Rather than relying on the traditional keyboard, graphic menu and command line systems, nowadays there is a tendency towards the use of conversational, multimodal and context-aware interaction models. This change is very much visible in areas like voice assistants, smart conversational agents, AI companions, tutors, social bots, fully immersive, multisensory and hybrid (XR) platforms, and multimodal interfaces. The studies show that several user-related parameters, like level of engagement, quality of interaction trust perceived intelligence, have been greatly improved through the use of multimodal interaction. Combination of voice interface and contextual awareness allows the users talk with intelligent systems more naturally. It helps to decrease the cognitive demand and increase interaction efficiency at the same time.

Gesture detection, gaze tracking, and emotion identification further enhance immersive technologies. Beyond technical performance, successful interaction depends on communication quality, user trust, transparency, emotional appropriateness, explainability, and ethical decision. Recent chatbot evaluations therefore, incorporate psychological dimensions, including trust, comfort, emotional connection, and user experience alongside conventional performance metrics. This trend reflects a broader transition from traditional HCL toward human-centered AI that integrates cognition, emotion, and social interaction (Shneiderman, 2022).

5.3 Advances in Speech, Language, & Communication Technologies

The reviewed studies identify speech and language technologies as among the most mature AI application areas. Continuous progress has been made in automatic speech recognition (ASR), speech generation, natural language understanding, conversational AI, multilingual language processing, and voice interaction through deep learning and transformer architectures. Large Language Models (LLMs) represent the most significant breakthroughs. Modern conversational systems generate coherent dialogues, summarizing information, answer complex questions, and support educational, clinical and professional applications. Integration with speech technologies enables seamless conversational interaction. Consequently, LLM-powered chatbots are increasing deployed in intelligent tutoring systems, digital assistants, accessibility applications, medical transcription, and communication support. Despite these advances, relatively few studies have focused on AI-assisted diagnosis clinical speech-language disorders, highlighting the need for reliable, interpretable and objective diagnostic tools for speech-language professionals and educators. Interdisciplinary integration of AI into clinical practice could substantially improve speech therapy by enabling automated performance monitoring and AI-assisted feedback for both therapist and patients.

5.4 Gesture Recognition, Computer Vision, and Affective Computing

Computer vision emerged as another major area. Studies show that speech and language technologies increasingly complemented by facial expression analysis, gesture recognition, reading emotions, body movement recognition, sign language translation, and scene understanding to support multimodal communication systems. The reviewed studies emphasized affective computing, which aims to develop AI systems capable of recognizing, interpreting, and responding to human emotions. Emotions recognition from facial expression, speech prosody, physiological signals, and gestures supports applications in mental healthcare, intelligent education, rehabilitation, customer service and socially assistive robotics. However, emotionally responsive AI also raises ethical concerns regarding privacy, personal autonomy, mental health interventions, and the protection of vulnerable populations. Gesture recognition has likewise expanded accessibility, interactive learning, sign language, virtual reality, rehabilitation, and human, robot interaction. Compared with speech recognition, gesture recognition remains methodologically less standardized because of inconsistence sensors, datasets, annotation protocols, and benchmarking procedures. Developing standardized multimodal gesture datasets should therefore be priority for future research.

5.5 AI in Healthcare, Education, Rehabilitation, and Accessibility

A major finding of this review is the expanding role of AI across healthcare, education, rehabilitation, and accessibility. AI systems have evoloved beyond analytical tools to support decision-making, communication, education, rehabilitation, and independent living. Healthcare studies focused on conversational AI, rehabilitation technologies, communication support systems, emotional support, and intelligent monitoring. Education research primarily examined personalized learning experiences, AI-powered tutoring, multilingual instruction, XR-supported learning, and AI-assisted assessment. Accessibility research emphasized speech to text systems, sign language recognition, eye-gaze interfaces, wearable assistive devices technologies, rehabilitation robots, and intelligent environments. These findings align with global efforts to promote inclusive and accessible AI by reducing communication barriers for individual with disabilities, multilingual learners, and other underserved populations, thereby improving educational opportunities, health care access, independent living, and social participation.

5.6 Emergence of Responsible AI and Governance

The reviewed literature demonstrates a clear a shift toward responsible AI governance, ethics, and regulatory frameworks. Earlier studies emphasized technological innovation, whereas recent research increasingly highlights transparency, accountability, fairness, explainability, privacy, human oversight, and legal compliance as essential requirements for trustworthy AI. Governance discussions frequently address empathic AI, conversational agents, emotion recognition, and multimodal surveillance systems. While these technologies substantial benefits healthcare, education, and accessibility, they also introduce challenges related to psychological dependence, manipulation, children's safety, algorithmic bias, informed consent, and data governance. Responsible AI has therefore emerged as an interdisciplinary field spanning computer science, psychology, communication, ethics, law and public policy, emphasizing that future AI systems must integrate ethical governance with human-centered design.

6. CONCLUSION

This scoping review covered empirical, computational, qualitative systematic review, and governance studies across psychology, computer science, education, healthcare, rehabilitation, accessibility, linguistics, and human-computer interaction. AI research has evolved single-modality systems to multimodal intelligence. Major areas include speech recognition, natural language understanding, gesture recognition, facial expression analysis, emotion recognition, eye-tracking, conversational AI, intelligent tutoring systems, virtual and augmented reality (XR), intelligent robotics, and multimodal foundation models. Advances in deep learning, transformer architectures, self-supervised learning, generative AI, and multimodal large language models have fundamentally improved AI's ability to interpret relationship across multiple communication modalities, enabling more natural, adaptive, and personalized human-computer interaction.

AI applications now extend beyond speech and language technologies into education, assistive communication, accessibility, rehabilitation, healthcare, human-computer interaction, conversational agents, voice interfaces, immersive learning, intelligent tutoring systems, AI companions, governance, and responsible AI. These findings demonstrate the increasingly interdisciplinary nature of AI, integrating computational intelligence, behavioral sciences, communication studies education psychology, and healthcare. The review also highlights the growing importance of responsible AI, emphasizing privacy, transparency, algorithmic fairness, explainability, accountability, emotional manipulation, user autonomy, and data governance. As AI becomes increasingly integrated into education, healthcare, accessibility, and communication, robust ethical and regulatory frameworks are essential to safeguard human rights, particularly for vulnerable populations.

This review also emphasizes the shift toward human-centered and context-aware AI, where user experience, accessibility, trust, transparency, emotional intelligence, inclusivity, and ethical governance are as important as computational performance. AI should complement rather than replace human communication across educational, healthcare, professional, and social setting. Future AI research should balance technological innovation with ethical, human-centered design to develop intelligent systems that communication natural, adapt to divers user needs, and promote accessibility, inclusion, and social well-being across different populations and environments.

7. REFERENCES

Alsaleh, A. (2026). Alsaleh, A. (2026). The influence of artificial intelligence on individuals with disabilities. Acta Psychologica, 262, 106010.https://doi.org/10.1016/j.actpsy.2025.106010

Arksey, H., & O'malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

Bailey, A. L., Abraham, M. E., Dong, J. Y., Torres, I. M., Wang, Y. P., Zamora, E., & Zhang, X. (2025). Translanguaging : Conceptual underpinnings of equity-oriented instructional and assessment practices with adolescent multilingual learners. Linguistics and Education, 87(February), 101418. https://doi.org/10.1016/j.linged.2025.101418

Chaturvedi, R., Verma, S., Das, R., & Dwivedi, Y. K. (2026). Technological Forecasting & Social Change Social companionship with artificial intelligence : Recent trends and future avenues. Technological Forecasting & Social Change, 193(November 2022), 122634. https://doi.org/10.1016/j.techfore.2023.122634

Cheng, K. Y. (2026). Principles of consent and non-addiction in AI grief bots. Journal of Responsible Technology, 26(May), 100177. https://doi.org/10.1016/j.jrt.2026.100177

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-navarrete, S., Giannakis, M., Al-debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, C., Jebabli, I., … Fosso, S. (2022). International Journal of Information Management Metaverse beyond the hype : Multidisciplinary perspectives on emerging challenges , opportunities , and agenda for research , practice and policy. International Journal of Information Management, 66(July), 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Elzer, S., Carberry, S., & Zukerman, I. (2011). The automated understanding of simple bar charts. Artificial Intelligence, 175(2), 526–555. https://doi.org/10.1016/j.artint.2010.10.003

Feng, S., Zhang, H., & Ga, D. (2025). Computers and Education : Artificial Intelligence Mapping the evolution of AI in education : Toward a co-adaptive and human-centered paradigm. 9(December). https://doi.org/10.1016/j.caeai.2025.100513

Giberga, A., Guerra, E., Ahufinger, N., Igualada, A., Aguilera, M., & Esteve-gibert, N. (2025). Prosody and gestures help pragmatic processing in children with Developmental Language Disorder. Journal of Communication Disorders, 115(December 2024), 106525. https://doi.org/10.1016/j.jcomdis.2025.106525

Górriz, J. M., Arco, J. E., Atzmueller, M., Ballarini, F., Barakova, E., Bologna, G., Bonomini, P., Duro, R. J., Elizondo, D., Julian, V., Kotz, G., Kutt, K., Leming, M., Lope, J. De, Macas, B., Martinez, J. J., Mekyska, J., Nalepa, G. J., Novais, P., … Zhu, Z. (2023). Computational approaches to Explainable Artificial Intelligence : Advances in theory , applications and trends. Information Fusion, 100(July), 101945. https://doi.org/10.1016/j.inffus.2023.101945

Hassani, S., Dackermann, U., Mousavi, M., & Li, J. (2024). A systematic review of data fusion techniques for optimized structural health monitoring. Information Fusion, 103(November 2023), 102136. https://doi.org/10.1016/j.inffus.2023.102136

Jiang, T., Sun, Z., Fu, S., & Lv, Y. (2024). Human-AI interaction research agenda : A user-centered perspective. 8(February).

Kayıs, E., Sarı, S., Dogan, N., Mumcu, E., Arı, R., Ozbaran, B., & Hatipo, A. (2026). A new approach in autism diagnosis : Evaluating natural interaction using point of view (POV) glasses. 116(December 2025). https://doi.org/10.1016/j.ajp.2025.104798

Kristina, D., Colasante, T., Roldan, E., Carlos, J., Martinez, R., Li, H., Mukherjee, S., & Malti, T. (2026). An experimental study exploring human – AI complementarity in early social-emotional learning. Computers and Education Open, 10(January), 100331. https://doi.org/10.1016/j.caeo.2026.100331

Levy, E. T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse processes*, *15*(3), 277-301.

Liu, J., Shi, Y., Gu, C., & Li, Q. (2025). Optimization design and application of mechanical characteristics in ergonomics of children ' s intelligent toys. Systems and Soft Computing, 7(December 2024), 200245. https://doi.org/10.1016/j.sasc.2025.200245

Liu, Z., Cheon, S., Stanbury, A., Jiao, X., Xing, W., & Kang, H. (2026). Computers and Education : Artificial Intelligence Towards contextual-based AI : A scoping review of artificial intelligence in X reality for personalized learning. Computers and Education: Artificial Intelligence, 10(December 2025), 100523. https://doi.org/10.1016/j.caeai.2025.100523

Ma, W., & Liu, Q. (2023). Language and health studies in the era of holistic health : Achievements and prospects. Language and Health, 1(2), 1–11. https://doi.org/10.1016/j.laheal.2023.11.001

Mcstay, A., & Bakir, V. (2025). Soft law for unintentional empathy : addressing the governance gap in emotion-recognition AI technologies. Journal of Responsible Technology, 23(July), 100126. https://doi.org/10.1016/j.jrt.2025.100126

Oruma, S. O. (2026). Computer Standards & Interfaces Enhancing security , privacy , and usability in social robots : A software development framework. 96(June 2025).

Peters, M. D., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., ... & Khalil, H. (2021). Updated methodological guidance for the conduct of scoping reviews. *JBI evidence implementation*, *19*(1), 3-10.

Quan, S., Tu-shea, X., Ding, Y., Du, Y., Zheng, Q., & Gerdich, L. E. (2026). Computers and Education : Artificial Intelligence Conversational AI in children ' s home literacy learning : effectiveness , advantages , challenges , and family perception. Computers and Education: Artificial Intelligence, 10(December 2025), 100549. https://doi.org/10.1016/j.caeai.2026.100549

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PmLR.

Seaborn, K., Altmeyer, M., Rikaku, G., Ku, B., Kobuki, S., & Urakami, J. (2025). International Journal of Human - Computer Studies The Voice Experience Inventory (VOXI): Validating a consensus-driven instrument for measuring user impressions of computer voice $. International Journal of Human - Computer Studies, 203(June), 103576. https://doi.org/10.1016/j.ijhcs.2025.103576

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Terkourafi, M., & Hansen, M. M. (2026). Hearer ' s Meaning 2 . 0 : A reply to Li and Xie (2025). Journal of Pragmatics, 261, 4–18. https://doi.org/10.1016/j.pragma.2026.04.007

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine, 169*(7), 467-473.

Turner, M., & Tour, E. (2025). Using a translanguaging multimodal approach to develop teachers ' language awareness in linguistically diverse classrooms in Australia. Teaching and

Teacher Education, 153(November 2024), 104835. https://doi.org/10.1016/j.tate.2024.104835

Wu, Q., Wang, W., Liu, Q., Zhang, R., Suen, Y., & Billinghurst, M. (2025). International Journal of Human - Computer Studies Augmented tabletop interaction as an assistive tool : Tidd ' s role in daily life skills training for autistic children. International Journal of Human - Computer Studies, 202(April), 103517. https://doi.org/10.1016/j.ijhcs.2025.103517

Xiao, F., Li, Z., Lin, J., Zou, X., Yang, D., & Zou, W. (2025). Computers and Education : Artificial Intelligence Leveraging an LLM-enhanced bilingual conversational agent for EFL children ' s dialogic reading : Insights from children , parents , and educators. Computers and Education: Artificial Intelligence, 9(December 2024), 100484. https://doi.org/10.1016/j.caeai.2025.100484

Yang, W. (2022). Computers and Education : Artificial Intelligence Artificial Intelligence education for young children : Why , what , and how in curriculum design and implementation ☆. Computers and Education: Artificial Intelligence, 3(February), 100061. https://doi.org/10.1016/j.caeai.2022.100061

Zhang, B., & Xie, M. (2026). Computers in Human Behavior Reports CALA-Net : Intelligent depression detection with multimodal fusion and Tree-of-Thoughts framework. Computers in Human Behavior Reports, 22(March), 101051. https://doi.org/10.1016/j.chbr.2026.101051

Zhu, Y., Gao, T., Fan, L., Huang, S., Edmonds, M., Liu, H., Gao, F., Zhang, C., Qi, S., Nian, Y., Tenenbaum, J. B., & Zhu, S. (2020). Dark , Beyond Deep : A Paradigm Shift to Cognitive AI with Humanlike Common Sense. Engineering, 6(3), 310–345. https://doi.org/10.1016/j.eng.2020.01.011

2026 by the Authors. This Article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/)