

# **Emotion-Aware AI Communication Interfaces: Enhancing Expressive Capabilities for Deaf, Non-Verbal, and Speech-Impaired Users**

Oyanibi Oyindamola Esther MBA  
SC Johnson College of Business, Cornell University, Ithica, New York, USA

Ifeoma Eleweke  
College of Technology and Engineering, Westcliff University  
Florida, USA

Yusuff Bolaji Ajegbile  
Department of Health and Human Kinetic, University of Ibadan, Nigeria.

## **Abstract**

Communication is the structural base of human experience, and many people who have to meet difficult communication needs, such as deaf and non-verbal or speech-impaired people, face daunting structural challenges in reaching expressive parity. Although important, conventional Augmentative and Alternative Communication (AAC) systems are often limited due to the lack of conversational speed, contextual awareness, and, most importantly, a lack of the ability to communicate subtle emotional and affective conditions. The theoretical review fills this gap by integrating the emerging developments in the context of affective computing, assistive technology (AT), human-computer interaction (HCI) and accessibility research. In the overall analysis, it is evident that the ongoing Artificial Intelligence (AI) applications are characterized by serious developmental challenges, especially with regard to pervasive biases of data due to homogeneous or non-performed emotional datasets, and the inability to effectively analyze non-homogeneous, non-standard expressive modalities, including non-verbal speech and dysarthric communication. Most importantly, the current models are deficient in a prescriptive framework of ethically implementing automated emotional responsiveness. As a result, this article presents an original conceptual framework titled Emotion-Aware Multimodal AAC Interaction Cycle (EA-MAAC), which is based on individual multimodal input, ethically responsible affective perception, contextual thinking, and adapting emotional response. With an exceedingly well-organized sequence of attack, the EA-MAAC model offers engineering next-generation interfaces that focus on user autonomy, access policy requirements, and universal principles of design, and thus improves the comprehensive expressive abilities and social assimilation of underserved populations.

## **INTRODUCTION**

The ability to explain thoughts, intentions, and, most significantly, emotional states becomes one of the key pillars of social participation and cognitive integrity (Light, 2019; LeDoux, 1994). Persons with the requirement to make use of complex communication needs (CCN), including deaf or hard of hearing (DHH), the minimal verbal (mv) and speech impaired (SI) due to a condition, such as stroke, cerebral palsy, or neurodevelopmental disorders, can often require the mobilization of Augmentative

and Alternative Communication (AAC) technologies in order to gain access to this universal right (Light et al., 2019a; Evangeline and Anitha, 2022). AAC systems, which may include the simplest (e.g., pictographic boards) to the most advanced (e.g., speech-generating devices (SGDs)) technology, can be used to support or substitute natural speech (Elsahar et al., 2019; Evangeline & Anitha, 2022; American Speech Language Hearing Association, 1991).

Nevertheless, after several decades of innovation, the effectiveness of the AAC is still severely undermined by systemic limitations. The most notable of these weaknesses is the lack of support of subtle emotional expression and interpretation (Larradet et al., 2020; Na et al., 2016). The nature of communication is not restricted to the transfer of semantic information, in fact, it is inherently affected by the semantics of the affect, body language, and paralinguage (Liu et al., 2023; Ma, 2022). The current AAC modalities are frequently associated with low conversational rates and the absence of contextual flexibility needed to support the natural conversation, thus reducing the user impact and preventing the true interpersonal interaction (Konadl et al., 2023; Arnott and Alm, 2013). This is a major limitation of propositional or instrumental communications and greatly hinders the capacity of users to express identity, will, and the depth of emotions (Kane et al., 2017).

Incorporating the methods of Artificial Intelligence (AI), and Machine Learning (ML), the possible revolution in breaking these barriers of expression is likely to occur (Evangeline & Anitha, 2022; Barua et al., 2022). The AAC solutions based on AI promise possible solutions to automated personalization, flexibility of context, and improved data management efficiency (Evangeline and Anitha, 2022; Li et al., 2021). Most importantly, the new area of affective computing offers the technological framework that needs to be implemented to allow the incorporation of emotions in these interfaces (Picard, 2000; Ma, 2022). This entails the identification, decoding, and acting upon human emotional manifestations in various sensory modalities (multimodality) (Liu et al., 2023; Rios et al., 2023).

In this paper, a thorough theoretical overview will critically examine the state of Emotion-Awakening AI Communication Interfaces (EA-AICI) in a person to communicate to the deaf, non-verbal, and speech-impaired. This review provides a theoretical background that is needed by drawing upon seminal and recent peer-reviewed literature of affective computing, assistive technology, HCI, and accessibility research. Then, it critically examines the limitations of current AI methods, especially in relation to data bias and the identification of non-standard modalities of expression (Guo et al., 2019; Narain et al., 2022). Last but not least, the review states a new theoretical framework, the EA-MAAC model, which combines technical potential, with socio-ethical demands, and suggests strict future research directions based on the development of emotionally rich, user-independent, and universally designed communication systems.

## Theoretical Framework

In order to go on with the conceptualization of EA-AICI, a synergistic theoretical underpinning based on affective computing, models of communication competence, and the socio-technical principles of assistive technology implementation is needed.

### Affective Computing and Emotional Modalities

The first issue of interest to affective computing (AC) involves the technology that can identify, analyse, and (in some cases) replicate human emotions (Picard, 2000; Ma, 2022). A complex state that leads to physiological and psychological changes that affect behaviour is emotion, which is conventionally conceptualised based on two main theoretical frameworks discrete and dimensional models (Ma, 2022; Plutchik, 2001).

Discrete emotion theories postulate few and universally identifiable, primitive emotions, which are joy, anger, fear and sadness (Ekman, 1992; Plutchik, 1980). Intelligibility to the user is a common choice-related feature in HCI applications, which leads to the popularity of these models (Ma, 2022). In contrast, dimensional models, including a popular Pleasure-Arousal-Dominance (PAD) (Mehrabian, 1996; Bradley and Lang, 1994) place emotions in a continuous space (Ma, 2022). Although dimensional models offer a theoretical and methodological framework of quantitative emotional computing, the choice of dimensions is still an issue (Ma, 2022). More importantly, AI systems will have to go beyond categorical or dimensional detection of simple emotions and focus on non-basic emotions, such as perplexity, boredom, or fatigue, which are essential in a conversation (Ma, 2022).

The data of emotions takes different forms in each of the human modalities:

1. **Vocal/Acoustic:** Speech emotion recognition (SER) applies the speech signal processing (SSP) method to identify emotion through such features as prosody and acoustic features (Ma, 2022; Schuller et al., 2003). In the case of people with serious speech impairments (SSI), it goes as far as acknowledging non-verbal vocal expressions (NVVEs) that convey information about affective and communicative messages (Narain et al., 2022).
2. **Visual/Kinesic:** Body language (BL) refers to non-verbal communication in the form of posture, gesture, and facial expression and in many cases it complements or dominates verbal information (Liu et al., 2023). One of them is facial gesture recognition, especially the ability to recognize subtle emotional expressions (Lyu et al., 2022; Li and Deng, 2020). In the case of DHH communities, it is obligatory to comprehend non-manual signs (facial expressions and head movement) that are inherent in sign languages in order to be able to fully grasp the linguistic context (Liu et al., 2023; Narain et al., 2022).
3. **Textual:** Sentiment analysis and natural language understanding (NLU) systems are text-based sentiment mining systems, which infer emotion and meaning based on written text and semantic context (Ma, 2022; Hsu and Ku, 2018).

To ensure affected states are accurately estimated, strong EA-AICI requires multimodal fusion, i.e. the combination of data of these different modalities to mitigate the

shortcomings of unimodal recognition, i.e. video, audio and text must be put together (Rios et al., 2023; Lyu et al., 2022; Zeng et al., 2008).

### **Assistive Technology and Communication Competence Models.**

Implementation of AAC technology should be placed within the set standards of human interaction and use of technology. The conceptual anchor that will be used in this paper is the Human Activity Assistive Technology (HAAT) model (Cook and Polgar 2015), which contends that the AT solutions must be considered in terms of the Human (physical, cognitive, and emotional), the Activity performed, and the Context (physical, cultural, and institutional) the device is operated in (Elsahar et al., 2019; Cook and Polgar, 2015).

An in-depth perspective on the subject of communication skill, especially among AAC users, is spread over four related areas, which are linguistic, operational, interpersonal and political (Light et al., 2019a). The existing AAC systems, which in some cases pay much attention to the linguistic and operational one (e.g., the speed of word recognition, access to vocabulary), do not assist in interpersonal competence- the ability to build social contacts, take turns, and be able to regulate emotional expression (Light et al., 2019a). This gap is specifically filled in EA-AICI which increases the interpersonal and political (autonomy/advocacy) dimensions using affective responsiveness (Light et al., 2019a).

The very design process should correspond to the principles of Universal Design (UD) and not be confined to the specifics of accessibility but should be built on the ideas of allowing people with a variety of capabilities to use a system without having to adjust to it (Fuglerud, 2014; Lidwell et al., 2010). UD requires the consideration of diversity, capabilities, preferences, and culture (Fuglerud, 2014; Smith and Yetim, 2004).

### **Literature Review**

The integration of AI, Affective computation and AAC research is a very interesting prospect, but when critically assessed, there exist very strong limitations and research gaps that still hinder the achievement of really good EA-AICI.

### **Artificial Intelligence in Speech and Non-Verbal Communication Enhancement.**

In the case of people who have severe speech impairments (SSI) or dysarthria, AI is becoming a more popular tool to increase the pace and fluency level of communication (Evangeline and Anitha, 2022). Voice Input Voice Output Communication Aids (VIVOCA) technologies have been used to generate speaker-specific recognition models of dysarthric speech by using statistical Automatic Speech Recognition (ASR) methods that include Hidden Markov Models (HMMs) (Hawley et al., 2012; Evangeline and Anitha, 2022). Nevertheless, the relationship between the level of impairment and accuracy tends to be negative (Larradet et al., 2020; Narain, 2021). Moreover, conventional ASR is difficult to use since speech recognition models had been constructed to accommodate standard articulation (Javed et al., 2022).

The AI application in the AAC systems is mostly concerned with efficiency in the conversation via predictive algorithms, including suggesting context-dependent

utterances or dynamically enlarging vocabulary in response to user location or selection history (Klauer et al., 2021; Neamtu et al., 2019; Li et al., 2021). However, research continuously demonstrates that AI applications in the AAC have gaps, which leads to a low rate of communication, frequently due to a perception of non-authentic and situationally inappropriateness of suggestions that are created (Konadl et al., 2023).

In the case of minimally verbal people (mv), it is rather about the identification of non-verbal communication indicators. A study has already proven it possible to classify personalized, real-world nonverbal vocalizations (NVVEs) based on the analysis of audio only (Narain et al., 2022). These affective and communicative expressions are aural, and their frequency, duration, and tone differ; however, only close communication partners can comprehend them without much effort (Narain et al., 2022). The data collection in this specialized field must introduce new protocols about individualized and momentary labeling in real-life conditions to emphasize spontaneity and contextual sensitivity (Narain et al., 2022; Narain and Maes, 2020).

In the case of DHH community, the focus on AI progress is Sign Language Recognition (SLR) and translation (Papatsimouli et al., 2022; Aly and Aly, 2020). To learn linguistic structure and affect, SLR requires deep learning models that will combine the recognition of both manual (hand shapes and movements) and non-manual markers (NMMs) (facial gestures, head movements and mouth morphemes). Acoustic sensing, frequently based on smart glasses and earphones, has become an inexpensive, omnipresent means of identifying both manual and non-manual cues simultaneously, eliminating the use of privacy-invading cameras (Jin et al., 2021; Lyu et al., 2022).

### **The Reliability Gap of Recognition of Emotion.**

The ongoing technical and epistemological issue of EA-AICI is the fact that generalized emotion recognition systems are not reliable when used on a diverse or disabled population.

First, it can lead to a drop in performance when facial expressions do not follow the canonically established patterns because of such conditions as Parkinson disease, Bell Palsy, or autism (Guo et al., 2019). The algorithms used to process emotions might fail to interpret the facial expressions of an autistic person who might not express themselves in a typical way (Guo et al., 2019). Moreover, it is not possible to conclude on the basis of the movements of the human face only that a person is in a certain emotional state (Barrett et al., 2019).

Second, the issue of data bias is a crucial problem threatening the justice and credibility of AI systems to people with disabilities (PwD) (Guo et al., 2019; Trewin, 2018). Current AI applications tend to be trained on homogeneous and non-inclusive data that constitute historical biases and can cause discrimination or failure of the system to work with PwD (Guo et al., 2019; Gadiraju et al., 2023). The use of acted or simulated databases of emotional speech is a significant weakness of SER, as it is hard to determine whether the emotions a system detects are true and spontaneous in the conditions of a real environment (Ma, 2022; Akcay and Oguz, 2020; Poria et al., 2019). Even the development of deep learning does not help to detect real emotions more accurately (Ma, 2022).

Third, current studies show considerable differences in addressing the entire range of conversational settings and stages (Konadl et al., 2023). The middle stage of conversation is what AI tools are usually concerned with, whereas initial phases of conversation (e.g., initiating conversation on the basis of context recognition) and closing ones are largely neglected (Konadl et al., 2023). Emotional interaction needs to be powerful and it is essential to be aware of the entire communicative cycle.

### **Ethical and social technical imperatives.**

The emergence of emotion-sensitive chatbots brings about ethical and socio-technical issues on a deeper level (the autonomy of users and the fairness of algorithms, in particular) (Mensio et al., 2018; Ma, 2022). Researchers indicate that AI fairness in PwD needs to be approached with specific care because prejudices can be expressed in different ways: either in stereotyping more (and thus, redefining stereotypes) or in disabled disappearance (Gadiraju et al., 2023; Trewin, 2018).

The EA-AICI should be developed based on the principles of trustworthy AI, which are Transparency, Justice, Non-Maleficence, Accountability, and Privacy (Smuha, 2019). The issue of privacy is a pressing one, particularly with the systems based on the capturing of extremely personal information, including physiological indicators or facial data or other non-verbal sounds (Ma, 2022; Narain, 2021). Emotion-sensing technology is also a threat to the privacy of the user and his or her communication partners (Smuha, 2019). More importantly, living experience should be prioritized in the development, where PwD should be treated as specialists who should inform the data annotation and the design of the system (Gadiraju et al., 2023; Park et al., 2021). It is not the mere technological assistance, but the encouragement of user agency, where AAC systems provide the freedom to express, personality, and identity, instead of limiting it to a uniform, disemotional synthetic voice (Kane et al., 2017; Larradet et al., 2020).

### **The Development of the Conceptual Model.**

A basic need, which is identified in the critical review, is the transition of the current AAC tools into dynamic, emotionally evocative, and context-specific communicative partnerships. Such a change requires a conceptual systematism, integrating personalized affective input and ethically reasoned, adaptive output. The Emotion-Aware Multimodal AAC Interaction Cycle (EA-MAAC) is our proposal to address the limitations of current AAC studies, namely the lack of homogeneity, the lack of expressivity, and the lack of context awareness (Konadl et al., 2023; Larradet et al., 2020; Narain et al., 2022).

According to the EA-MAAC model, communication between the user and the communication partner occurs via continuous interaction over four mutually dependent phases between the user and the AI interface:

### Phase 1: Multimodal Affective personal input (P-MAI).

The stage is about user-generated communication acquisition in a variety of preferred modalities and no longer relies on the traditional text entry that is touch-based or standard symbols (Li et al., 2021; Patricio et al., 2020).

1.Heterogeneous Signal Acquisition: The interface will have to simultaneously receive signals relating to linguistic content (text/pictograms/SL manual markers) and affective signals (Narain et al., 2022; Liu et al., 2023). In the case of speech-impaired patients, it involves the use of robust SSP techniques that are customized according to the acoustic characteristics of a specific user to process non-standard voice input or dysarthric voice input (Hawley et al., 2012; Ma, 2022). In the case of mv\* people, they record nonverbal vocalizations (NVVEs), which use the cheap acoustic sensor (Narain et al., 2022; Jin et al., 2021). In the case of DHH users, non-manual cues (facial expressions/head movements) that enable the determination of signed linguistic and emotional context are captured in visual or acoustic sensory perception (Liu et al., 2023).

2.User-Defined Emotional Profiling: This stage will be required because it is difficult to implement generalized emotion modeling (Guo et al., 2019); thus, it requires highly personalized data labeling and training. The interface should facilitate the definition and labelling of unique expressions of the affective states by users or trusted communication partners (Narain et al., 2022; Gadiraju et al., 2023). This will allow the AI to acquire the user idiosyncratic emotional lexicon, as well as the individual expressions (Narain et al., 2022).

### Phase 2: Multimodal Fusion (AI-MF) and Affective Interpretation.

The input signals are read and combined to derive the desired emotional condition and communicative purpose of the user.

1. Multimodal Feature Extraction The deep learning algorithms are used to extract salient features within each modality (Rios et al., 2023; Lyu et al., 2022). One can use Mel-Frequency Cepstral Coefficients (MFCCs) on speech, deep convolutional networks on the extraction of visual features, and Natural Language Processing (NLP) on a textual context (Evangeline and Anitha, 2022; Ma, 2022).

2.Context-Sensitive Fusion: Late/Early fusion paradigms are also used so that features can be combined to maximise prediction accuracy (Rios et al., 2023). Most importantly, the fusion process should be context-sensitive, i.e. it should consider non-affective elements such as the activity, location of the user (Neamtu et al., 2019) and the stage of the conversation (Konadl et al., 2023). The model should be focused on the real-world spontaneity through the application of the transfer learning methodology that is adjusted with the help of the personal profiles created in Phase 1 by the user (Ma, 2021; Narain et al., 2022).

3.Risk Reduction and Uncertainty Management: The model should include the elements that can differentiate between the genuine emotional expression and possible imitation or false signals (Ma, 2021). It must display uncertain signs, which will encourage openness and leave the user with the choice of verifying the meaning (Smuha, 2019).

### Phase 3: Ethical and Contextual Reasoning (E-CR).

Before producing an output, the AI system will have to compare the obtained emotional or linguistic data with a set of socio-ethical and contextual regulations. This stage plays a vital role in overcoming the level of technological responsiveness to the advanced, human-like interactive process (Ma, 2022; Ren and Bao, 2020).

1.Contextual State Assessment: The system evaluates the social context (e.g., formal and informal context, conversation beginning and versus middle/closing) (Konadl et al., 2023). This defines the suitability of the level of emotional expression and language of response (Ma, 2022).

2.Ethical Constraint Application: Layers of reasoning implement non-maleficence and transparency constraints (Smuha, 2019). As an example, the output of the system should focus more on neutral, calming, or directive reactions than simply replicate the negative feeling in case the identified emotion is severe distress (Ma, 2022; Ma, 2022). This includes the human emotional reaction strategies (Ma, 2022). The AI should also evaluate the input in terms of the possible representational harms or magnification of stereotypes (Gadiraju et al., 2023).

3.Autonomy and Preference Check: The model checks whether the suggested output contravenes any of the pre-existing user preferences in terms of voice gender, personality modulation, or the level of emotional expressivity (Ma, 2022; Kane et al., 2017). The so-called Sceptics users may want to have less emotionality, whereas Enthusiasts may want to have a more expressive system (Ma, 2022).

### Phase 4: Adaptive Emotional Output Production (A-EOG)

The last step is the production of a multimodal output that is comprehensible and conveys both the linguistic and the emotional meaning, which should be produced in a way that is personalized and socially acceptable.

1.Emotionally Modulated Speech Synthesis: Linguistic text is translated into artificial speech through the methods that adjust the prosodic characteristics (pitch, rate, intensity) to achieve the desired emotion (Rizos et al., 2020; Larradet et al., 2020). To users of AAC, this has to involve the functionality that will add expressive individuality to the monotony of standard synthetic voices (Mills et al., 2014).

2.Visual Affective Representation: The system produces additional visual output, e.g. emotional 3D avatars, dynamic pictograms, or custom interface changes (Larradet et al., 2020; Catania and Garzotto, 2022). In the case of DHH users, it is a process of making the content of affect content contextually pertinent non-manual marks on a sign language avatar (Liu et al., 2023).

3.Feedback and Learning Integration: The output is captured and user/communication partner response is registered and fed back into Phase 1 to optimize the emotional profiling and enhance further predictions, completing the loop of the iterative learning (Ma, 2021).

## Discussion

The strong embodiment of Emotion-Aware AI Communication Interfaces (EA-AICI) has a devastating implication on various fields, which require keen concern on the policy of accessibility, ethical implementation, and the principles of design.

### **Policy on Accessibility and Development of Assistive Technology.**

The existing AT growth currently encounters a problem of accessibility because it is expensive, complicated infrastructure, and it is less portable (Elsahar et al., 2019; Evangeline and Anitha, 2022). The policy should require the creation of AI-powered AAC systems, which use the cheap, ubiquitous devices, including smartphones and tablets, to make them more accessible and usable in the real world (Evangeline and Anitha, 2022; Light et al., 2019a). An example can be the systematic review on AI in AAC, which provides information about the opportunities of creating mobile application that combines AI to personalize and recommend based on the circumstances, especially in developing countries where resources are limited (Farzana et al., 2020; Farzana et al., 2021).

Moreover, the policy should consider the need of multilingual and culturally consistent design (Farzana et al., 2021). The weakness of generic tools in the region-specific languages or culturally specific pictogram representations limits the conceptual flexibility and applicability of AAC (Evangeline & Anitha, 2022). This gap should be resolved by implementing AI systems that should have regional and contextual communicating data (Evangeline & Anitha, 2022). The frameworks of the policy must encourage the development of standardized data sets that facilitate inclusion and imitation to vulnerable groups, requiring the complex ethical concerns to be managed during their development (Guo et al., 2019; Park et al., 2021).

### Artificial Intelligence in Ethical Implementation and User Control.

Strict ethical considerations should also be applied in the implementation of the AI that can sense and create emotion to ensure user autonomy and avoid harm. These are some of the core principles of trustful AI: Transparency, Justice, Non-Maleficence, and Privacy (Smuha, 2019).

1.Privacy and Safety: The fact that the process of gathering and analyzing very sensitive emotional and biometric data (e.g., nonverbal vocalizations, facial expressions) involves an abundance of sensitive data necessitates the use of effective data protection tools (Ma, 2022; Narain, 2021). Users should also be allowed complete access to information about what data is gathered, the purpose of its use, and access by whom, which will eliminate the risk of surveillance (Smuha, 2019; Narain, 2021).

2.Algorithms Fairness and Bias: Algorithms should be explicitly outlined regarding their bias on PwD and that should be evaluated in the form of identifying and countering failure modes as a result of training data not being inclusive (Guo et al., 2019). It demands the prioritization of lived experience by being knowledgeable about data annotation pipelines to describe harms that are subtle (Gadiraju et al., 2023). Existing

system failures, e.g. loss of disability or reinforcement of stereotypes (Gadiraju et al., 2023), need to be systematically mitigated.

3. User Autonomy and Control: The basic socio-technical aspect is to increase user control over the emotional behavior of the AI. The EA-MAAC model helps to make sure that users can tailor the expression of emotion and the extent of emotion detection by the AI to avoid the imposition of unwanted emotions states or a limitation of self-expression by the AI (Ma, 2022; Kane et al., 2017). This will respond to the urgent necessity of PwD to own their digital image and identity of communication (Larradet et al., 2020).

### Universal Design Principles.

The continued development of EA-AICI supports the need to apply the universal design (UD) and similar principles of ability-based design carefully (Wobbrock et al., 2018). UD is focused on the flexibility of interaction and independence of dialogue and does not bind the software logic and user interface (Burzagli and Emiliani, 2020). This independence of dialogue enables interfaces to change dynamically to the preferred input modality of the user (e.g., eye gaze, switch access, silent speech, vocalization) (Elsahar et al., 2019; Larradet et al., 2020).

In addition, cognitive differences should be considered in the design. Indicatively, certain research notes that, visual content, e.g., the use of icons or pictograms with text, is advantageous to the dyslexic user and makes it easier to access (Muftah and Altaboli, 2020). Intellectually disabled people can benefit from AI-based systems such as real-time captioning, virtual reality (VR) communication, and personalized communication support that ensure their inclusivity and independence (Intellectual Disability and AI, 2023; Kawas et al., 2016). Expressive visual features, including emotionally expressive avatars or a well-considered use of emojis, can also facilitate the conceptualization and communication of emotions to the population, including people with neurodevelopmental disorders (Larradet et al., 2020; Catania et al., 2022).

### Future Research Directions

In order to achieve the potential of the EA-MAAC model and address the identified limitations of this review, a number of research directions which are interdependent should be thoroughly investigated.

1. **Modeling Real-World, Spontaneous Affect:** The concept of systems being trained on the basis of acted-emotion databases should be abandoned in favor of highly sophisticated models that will reliably identify spontaneous emotion and non-basic affective states (Ma, 2022; Poria et al., 2019). This requires the development of new, longitudinal data collection procedures that can render highly individualized and contextualized affective data of mv+ and speech-deafened people in real-life circumstances (Narain et al., 2022; Ma, 2021). The future research ought to be aimed at the means of quantifying the restorative outcomes and exhaustion produced by the speech cues (Ma et al., 2021), so that VIs can react in advance.
2. **Contextual and Conversational Fluency:** The gaps in the coverage of the conversational phase should be solved in the future research (Konadl et al., 2023). This involves creating AI methods to detect the start of a conversation (e.g. greeting

signals or social signals) and the end of a conversation (e.g. farewell signals) and, as a consequence, offer contextually relevant emotional reactions and linguistic proposals (Konadl et al., 2023). Improved NLP models should be researched on so that conversational cohesion and semantic accuracy of complex dialogue is high (Vertanen, 2017).

3. **Optimizing Multimodal Fusion and Non-Standard Expression:** More work still needs to be done to optimize multi-modal fusion architectures on non-standard inputs (Rios et al., 2023). In particular, the areas of interest should include: (a) better integration of non-manual markers (NMMs) with real-time sign languages with low-cost acoustic or wearable sensors (Jin et al., 2021; Liu et al., 2023), and (b) the personalized diagnosis and optimization of the expressive capacity of individuals with Intellectual Disability (ID) using AI-based methods (Intellectual Disability and AI, 2023).
4. **Socio-Technical and User-Centric Evaluation:** Intense evaluation should not be limited to their conventional accuracy indicators (Ma, 2022). Future research should examine the attitudes and preferences of users toward an emotionally and personality-sensitive VAs in different cultural settings (Ma, 2022). There should be research aimed at effectively humanizing the ways of responding to negative emotions expressed through an AI interface (Ma, 2022). In addition, more research should be directed towards design-oriented studies to determine how different visual aesthetics and interface features (e.g., personalized emojis, appearance of an avatar) influence the cognitive and psychological reaction of people with a certain disabilities (e.g., autism) (fulltext, 2021).

## **Conclusion**

The next important trajectory of evolution in the Augmentative and Alternative Communication technology is Emotion-Aware AI Communication Interfaces. The traditional emphasis on the simple word-searching or word-client inquiry has been not enough to sustain the interpersonal variety of human communication (Light et al., 2019a). It was found in this review that the construction of sound EA-AICI must be a careful interdisciplinary synthesis between the technical skills of multimodal affective computing and the deep ethical and usability requirements of accessible technology and accessibility research.

Along with the natural limitations, such as algorithmic bias based on homogeneous data (Guo et al., 2019), the enduring problem of the inability to recognize non-standard and spontaneous expressions (Narain et al., 2022; Ma, 2022), and technical limitations in supporting the entire communicative process (Konadl et al., 2023), this requires a systematic theoretical response. The suggested EA-MAAC model provides this framework, which has the form of a cyclical model in which personalized input and ethically considered, adaptive emotional output are the basis. With a stepwise implementation of user autonomy, advanced consideration of privacy, and compulsory universality design principles to focus on diversity and cultural context, the limitations of assistive technology can be surpassed in future studies that are directed by the EA-MAAC model. It is not only a technical project; it is a moral one applicable to tear down the communication gap, enhance the full expressive potential, and have deaf, non-verbal, and speech-impaired people as fully included in the digital era.

## References

Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116, 56–76.

Aly, S., & Aly, W. (2020). DeepArSLR: A Novel Signer-Independent Deep Learning Framework for Isolated Arabic Sign Language Gestures Recognition. *IEEE Access*, 8, 83199–83212.

American Speech Language Hearing Association. (1991). Report: augmentative and alternative communication. *ASHA*, 33, 9–12.

Arnott, J. L., & Alm, N. (2013). Towards the improvement of Augmentative and Alternative Communication through the modelling of conversation. *Computer Speech & Language*, 27(6), 1194–1211.

Ascari, R. E., Pereira, R., & Silva, L. (2020). Computer vision-based methodology to improve interaction for people with motor and speech impairment. *ACM Transactions on Accessible Computing*, 13(4), 1–33.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68.

Barua, P. D., Vicnesh, J., Gururajan, R., Oh, S. L., Palmer, E., Azizan, M. M., Kadri, N. A., & Acharya, U. R. (2022). Artificial Intelligence Enabled Personalised Assistive Tools to Enhance Education of Children with Neurodevelopmental Disorders—A Review. *International Journal of Environmental Research and Public Health*, 19(3), 1192.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59.

Burzagli, L., & Emiliani, P. L. (2020). From Accessible Interfaces to Useful and Adapted Interactions. In C. Stephanidis & M. Antona (Eds.), *Universal Access in Human–Computer Interaction. Design Approaches and Methods* (pp. 28–36). Springer International Publishing.

Catania, F., & Garzotto, F. (2022). A conversational agent for emotion expression stimulation in persons with neurodevelopmental disorders. *Multimedia Tools and Applications*, 797–12828.

Catania, F., Di Nardo, N., Garzotto, F., & Occhiuto, D. (2019). Emoty: an emotionally sensitive conversational agent for people with neurodevelopmental disorders. In

*Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 2014–2023).

Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion recognition through multiple modalities: face, body gesture, speech. In R. Beale & C. Peter (Eds.), *Affect and Emotion in Human-Computer Interaction* (pp. 92–103). Springer.

Cook, A. M., & Polgar, J. M. (2015). *Assistive Technologies Principles and Practices* (4th ed.). Elsevier.

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99(3), 550–553.

Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., & Mansor, A. (2019). Augmentative and alternative communication (AAC) advances: A review of configurations for individuals with a speech disability. *Sensors*, 19(8), 1911.

Evangeline, B., & Anitha. (2022). A survey on Artificial Intelligent based solutions using Augmentative and Alternative Communication for Speech Disabled. *Research Article*.

Farzana, W., Sarker, F., Hossain, Q. D., Chau, T., & Mamun, K. A. (2020). An evaluation of augmentative and alternative communication research for ASD children in developing countries: Benefits and barriers. In *International Conference on Human-Computer Interaction* (pp. 51-62). Springer, Cham.

Farzana, W., Sarker, F., Chau, T., & Mamun, K. A. (2021). Technological Evolvement in AAC Modalities to Foster Communications of Verbally Challenged ASD Children: A Systematic Review. *IEEE Access*, 9, 10041–10061.

Fuglerud, K. S. (2014). Inclusive design of ICT: the challenge of diversity. (Unpublished doctoral dissertation). University of Oslo.

Gadiraju, V., Kane, S., Dev, S., Taylor, A., Wang, D., Denton, R., & Brewer, R. (2023). "I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)* (pp. 12 pages). ACM.

Gao, Y., Jin, Y., Li, J., Choi, S., & Jin, Z. (2020). Echowhisper: Exploring an acoustic-based silent speech interface for smartphone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3), 1–27.

Guo, A., Kamar, E., Vaughan, J. W., Wallach, H., & Morris, M. R. (2019). Toward Fairness in AI for People with Disabilities: A Research Roadmap. *arXiv preprint arXiv:1907.02227*.

Hawley, M. S., Cunningham, S. P., Green, P. D., Enderby, P., Palmer, R., Sehgal, S., & O'Neill, P. (2012). A voice-input voice-output communication aid for people with severe speech impairment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(1), 23–31.

Hsu, C. C., & Ku, L. W. (2018). Socialnlp 2018 emotionx challenge overview: Recognizing emotions in dialogues. In *Proceedings of the 6th international workshop on natural language processing for social media* (pp. 27–31). IEEE.

Intellectual Disability and AI. (2023). *Journal of Developmental Research*, 3(1), 30–38.

Javed, T., Doddapaneni, S., Raman, A., Bhogale, K. S., Ramesh, G., Kunchukuttan, A., Kumar, P., & Khapra, M. M. (2022). Towards building ASR systems for the next billion users. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Jin, Y., Gao, Y., Zhu, Y., Wang, W., Li, J., Choi, S., Li, Z., Chauhan, J., Dey, A. K., & Jin, Z. (2021). SonicASL: An acoustic-based sign language gesture recognizer using earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 5(2), 1–30.

Kane, S. K., Morris, M. R., Paradiso, A., & Campbell, J. (2017). “At Times Avuncular and Cantankerous, with the Reflexes of a Mongoose”: Understanding Self-Expression Through Augmentative and Alternative Communication Devices. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)* (pp. 1166–1179). ACM.

Kawas, S., Karalis, G., Wen, T., & Ladner, R. E. (2016). Improving real-time captioning experiences for deaf and hard of hearing students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 15–23). ACM.

Klauer, M., Konadl, F., Reinkemeier, A., & Matzner, M. (2021). AI-based recommendations of augmentative and alternative communication systems: Systematic literature review. *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia* (pp. 1–11). ACM.

Konadl, F., Klauer, M., Reinkemeier, A., & Matzner, M. (2023). Assessing AI in AAC Systems: A Systematic Review on AI Approaches and their Coverage of Conversation Phases and Contexts. In *Proceedings of the European Conference on Information Systems (ECIS)*.

Konijn, E. A., & Van Vugt, H. C. (2008). Emotions in mediated interpersonal communication: Toward modeling emotion in virtual humans. In *Mediated interpersonal communication* (pp. 114–144). Routledge.

Larradet, F., Raba, D., Garzotto, F., Cattai, A., & Dautenhahn, K. (2020). Affective Gaze-Based Speaking Tool for Locked-In Syndrome Patients with Personalized Emotional Avatar and Speech. In *Human-Computer Interaction. Design Approaches and Methods* (pp. 138–151). Springer.

LeDoux, J. E. (1994). Emotional processing, but not emotions, can occur unconsciously. In P. E. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 291–292). Oxford University Press.

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*.

Li, W., Qiu, X., Li, Y., Ji, J., Liu, X., & Li, S. (2021). Towards a novel machine learning approach to support augmentative and alternative communication (AAC). *International Journal of Speech Technology*, 25, 331–341.

Lidwell, W., Holden, K., & Butler, J. (2010). *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design*. Rockport Pub.

Light, J., McNaughton, D., Beukelman, D., Fager, S. K., Fried-Oken, M., Jakobs, T., & Jakobs, E. (2019a). Challenges and opportunities in augmentative and alternative communication: Research and technology development to enhance communication and participation for individuals with complex communication needs. *Augmentative and Alternative Communication*, 35(1), 1–12.

Liu, L., Gao, L., Lei, W., Ma, F., Lin, X., & Wang, J. (2023). A Survey on Deep Multi-modal Learning for Body Language Recognition and Generation. *arXiv preprint arXiv:2308.08849*.

Lotfian, R., Busso, C., & Schuller, B. W. (2016). Building Naturalistic Emotionally Balanced Speech Corpus by Retrieving Emotional Speech From Existing Podcast Recordings. *IEEE Transactions on Affective Computing*.

Lyu, Z., Yang, W., Yan, B., Du, W., Sun, X., & Yan, Y. (2022). Deep Learning Technology in Intelligent Human–Computer Interaction: A Survey of Recent Advancements. *Applied Sciences*, 12(22), 11457.

Ma, Y. (2021). Fake moods: Can users trick an emotion-aware voicebot? In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–4). ACM.

Ma, Y. (2022). *Emotion-Aware Voice Interfaces*. (Unpublished doctoral dissertation). Ludwig-Maximilians-Universität München.

Ma, Y., Drewes, H., & Butz, A. (2022). How should voice assistants deal with users' emotions? *arXiv preprint arXiv:2204.02212*.

Ma, Y., Li, J., Drewes, H., & Butz, A. (2021). You sound relaxed now - Measuring restorative effects from speech signals. In *IFIP Conference on Human-Computer Interaction* (pp. 585–594). Springer.

Mehrabian, A. (1996). Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261–292.

Mensio, M., Rizzo, G., & Morisio, M. (2018). The rise of emotion-aware conversational agents: threats in digital emotions. In *Companion Proceedings of the The Web Conference 2018* (pp. 1541–1544).

Mills, C., Poulton, T., Arnott, J., & Alm, N. (2014). Tailoring synthetic speech to the individual. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 235–236). ACM.

Muftah, M., & Altaboli, A. (2020). Investigating the Effect of Adding Visual Content to Textual Search Interfaces on Accessibility of Dyslexic Users. In C. Stephanidis & M. Antona (Eds.), *Universal Access in Human-Computer Interaction. Design Approaches and Methods* (pp. 290–300). Springer International Publishing.

Myers, D. G. (2004). *Theories of emotion. Psychology: Seventh Edition*. Worth Publishers.

Na, J. Y., Wilkinson, K., Karny, M., Blackstone, S., & Stifter, C. (2016). A synthesis of relevant literature on the development of emotional competence: implications for design of augmentative and alternative communication systems. *American Journal of Speech-Language Pathology*, 25(3), 441–452.

Narain, J. (2021). *Modeling Real-World Affective and Communicative Nonverbal Vocalizations from Minimally Speaking Individuals*. (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Narain, J., Johnson, K. T., Picard, R. W., & Maes, P. (2022). Modeling Real-World Affective and Communicative Nonverbal Vocalizations From Minimally Speaking Individuals. *IEEE Transactions on Affective Computing*.

Narain, J., & Maes, P. (2020). Understanding AAC Usage and Needs through a Web Survey with Device Users and Families. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (pp. 3864–3868). IEEE.

Neamtu, R., Camara, A., Pereira, C., & Ferreira, R. (2019). Using artificial intelligence for augmentative alternative communication for children with disabilities. In D. Lamas et al. (Eds.), *INTERACT 2019* (pp. 234–243). Springer, Cham.

Park, J. S., Bragg, D., Kamar, E., & Morris, M. R. (2021). Designing an online infrastructure for collecting AI data from people with disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 52–63).

Papatsimouli, M., Sarigiannidis, P., & Fragulis, G. F. (2022). A Survey of Advancements in Real-Time Sign Language Translators: Integration with IoT Technology. *Technologies*, 11(1), 83.

Patrício, C., Lucas, T., Goncalves, C., Pinto, N., Freitas, D., & Figueiredo, L. (2020). MagicContact Web: A Multiplatform Application for Augmentative and Alternative Communication. In J. C. R. P. da Silva et al. (Eds.), *Advances in Usability, User Experience, Wearable and Assistive Technologies* (pp. 807–817). Springer International Publishing.

Picard, R. W. (2000). *Affective Computing*. MIT Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Theories of emotion* (pp. 3–33). Elsevier.

Plutchik, R. (2001). The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4), 344–350.

Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access*, 7, 100943–100953.

Ren, F., & Bao, Y. (2020). A review on human-computer interaction and intelligent robots. *International Journal of Information Technology and Decision Making*, 19(1), 5–47.

Rios, A., Reichel, U., Bhuvaneshwara, C., Filntisis, P., Maragos, P., Burkhardt, F., Eyben, F., Schuller, B., Nunnari, F., & Ebling, S. (2023). Multimodal Recognition of Valence, Arousal and Dominance via Late-Fusion of Text, Audio and Facial Expressions. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 571–576). i6doc.com.

Rizos, G., Baird, A., Elliott, M., & Schuller, B. (2020). Stargan for emotional speech conversion: Validated by data augmentation of end-to-end emotion recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3502–3506). IEEE.

Schuller, B., Rigoll, G., & Lang, M. (2003). Hidden markov model-based speech emotion recognition. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03)*, 2, II–1. Ieee.

Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International*, 20(4), 97–106.

Smith, A., & Yetim, F. (2004). *Global human–computer systems: cultural determinants of usability*. Oxford University Press.

Trewin, S. (2018). AI fairness for people with disabilities: Point of view. *arXiv preprint arXiv:1811.10670*.

Vertanen, K. (2017). Generating holistic utterances in Augmentative and Alternative Communication. *Proceedings of the 14th European Conference on Speech Communication and Technology (Interspeech)*.

Voon, H. R., Li, S., Wang, T., Zhang, S., Zhang, W., & Lim, W. M. (2019). Autisay, a Communication Aid for Autism Spectrum Disorder (ASD). In *International Conference on Human-Computer Interaction* (pp. 160–171). Springer, Cham.

Wobbrock, J. O., Gajos, K. Z., Kane, S. K., & Vanderheiden, G. C. (2018). Ability-based design. *Communications of the ACM*, 61(6), 62–71.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2008). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39–58.