

Assistive Technology Outcomes and Benefits
Volume 18, Spring 2024, pp. 100-123
Copyright ATIA 2024 ISSN 1938-7261
Available online: atia.org/atob

Voices from Industry

Revolutionizing Augmentative and Alternative Communication with Generative Artificial Intelligence

Kenneth R. Hackbarth, M.S., M.S., M.E.

President and Founder, Volksswitch.org

Corresponding Author

Kenneth R. Hackbarth
President and Founder
Volksswitch.org
11027 Balsam Street
Westminster, OH 80021
Phone: (303) 902-2603
Email: ken@volksswitch.org

Abstract

Generative AI, a type of Artificial Intelligence (AI), is set to revolutionize Augmentative and Alternative Communication (AAC). New AAC devices powered by this technology will enable non-speaking individuals to engage in real-time conversational speech, bridging the gap between AAC users and non-AAC users in daily life. The breakthrough lies in the ability of this AI to rapidly generate human-like text using contextual prompts. Generative AI, integrated with existing technologies like text-to-speech, facial recognition, and eye tracking, will unleash new possibilities for time-relevant, effortless, intuitive, and personalized communication for individuals with complex communication needs. The technology needed is currently available or in development, and companies are rapidly moving to bring these products to the market. We are on the verge of a disruptive revolution that will reshape most aspects of life and open doors for AAC users that have been largely closed until now.

Keywords: augmentative and alternative communication, generative AI, artificial intelligence, communication rates

Revolutionizing Augmentative and Alternative Communication with Generative Artificial Intelligence

A particular type of Artificial Intelligence (AI), called Generative AI, will revolutionize the Augmentative and Alternative Communication (AAC) domain. Soon, a new breed of AAC devices will help to break through the barrier that has separated non-speaking individuals from general society for too long. These devices will place the long-sought goal of real-time conversational speech within the reach of individuals with complex communication needs. This revolution will take advantage of a recent breakthrough in the field of AI: the ability to easily and quickly generate unique, yet relevant, human-like text using a few contextual prompts. This technology is called “generative artificial intelligence.”

While Generative AI is the linchpin of the architecture described in this article, a few additional technologies are needed to fulfill this vision. Fortunately, these technologies already exist. They simply need to be assembled and tuned. Many of these technologies, like text-to-speech and eye tracking, have been around for years, but recent advances have rendered them almost magical. Companies that have built up expertise in traditional AAC devices and services over decades will need to retool and retrain quickly to accommodate these changes in much the same way that camera and film companies had to reinvent themselves upon the arrival of digital imaging.

Waller (2019) predicted that unlocking the inherent value and promise of individuals with complex communication needs would require a paradigm shift in the design of AAC. Sennott et al. (2019) recognized that artificial intelligence might very well be the basis for that shift. Because these authors published in 2019, they could not have known about the qualitative change in AI that would burst onto the scene a few years later. Rather than opening up new possibilities for intelligent word prediction and encodings based on contextual clues and specialized situational logic, the application of prodigious amounts of computer hardware and training data has made it possible to generate entire thoughts, both relevant and personalized, in seconds (Hwang & Chen, 2023).

The technology described in this article is either readily available or currently in development. Companies are moving at a dizzying speed to bring these products to market. As such, the statements and recommendations included here must be prefaced with the words “for now” or “currently.” We have suddenly reached the inflection point in the advance of information technology predicted by Ray Kurzweil (2005, p. 11) almost two decades ago.

...we won't experience one hundred years of technological advance in the twenty-first century; we will witness on the order of twenty thousand years of progress (again, when measured by today's rate of progress), or about one thousand times greater than what was achieved in the twentieth century.

Target Audience and Relevance

This article is relevant to several audiences — AAC users, AAC vendors, AAC service providers, and AAC researchers. AAC users will learn that the capabilities of generative AI, that are currently benefiting individuals throughout society, can and will support them in demonstrating their inherent value and talents, along with putting them in settings where they can influence the thinking of others. AAC vendors and service providers will get a glimpse into the future of the industry, along with the skills they will need to develop and the technologies they will need to incorporate. AAC researchers will be inspired to explore entirely new approaches and, like the AAC vendors, create new alliances and collaborative relationships.

Disclaimer

In the interest of full transparency, I am neither a developer nor researcher of artificial intelligence, nor an expert in psychology, neurobiology, linguistics, conversational analysis, or augmentative and alternative communication. Instead, I have a three-decade history in designing and developing software and hardware products as a systems engineer at AT&T Bell Laboratories and its subsequent divestitures. That work included consulting with telecommunications companies and vendors to understand their current capabilities, describe their embedded operational systems, document requirements for new functionality, and direct the development of complex products involving multiple existing and custom components.

Communication Types

Table 1 represents attempts to organize human communication into three categories as a function of the time delay between information exchanges. This approach uses the categories and terminology associated with data processing (system-to-system data exchange) as a helpful model for categorizing types of human communication.

Table 1: Human Communication Types

Type	Delay Between Exchanges	Examples
Real-Time	instantaneous to a few seconds	voice/video calling, face-to-face spoken or signed interactions
Near-Real-Time	tens of seconds to a few hours	chatting, messaging
Batched	a few hours to days	email, letters, voice/video recordings

Communication forms like books, college lectures, graffiti, and bumper stickers are almost entirely one-directional and, as such, have an indeterminate exchange delay. On the other hand, body language can communicate emotional states in fractions of a second, but the construction and interpretation of the exchanges are mainly unconscious. Note that in the case of real-time communication, the delays between exchanges can even be negative, such as the interruption of, or speaking over, the questions and responses of others.

Each of these communication types can be used to establish, reinforce, and maintain personal relationships or to persuade others to adopt a particular point of view. However, the shorter the exchange and the more in-person a communication type, the more clarifying, reinforcing, and persuasive the interaction can be. More

delayed or remote exchanges can be useful when first establishing relationships and later when maintaining them.

Real-time exchanges represent the vast majority of interactions that take place in society. Anyone who cannot participate in these interactions will often be relegated to the periphery of community and relationships. Developing influence can begin through batched communication like letters to the editor, business proposals, or white papers. However, eventually, one will need to restate the core of the argument in person and in real-time to the people with power. Without this ability, individuals are often relegated to positions of little influence. Unfortunately, individuals who use AAC devices rarely progress beyond transactional to interactional communication (Waller, 2006). Waller defines transactional communication as “expressing concrete needs/wants and transferring information” (p. 221). Interactional communication is defined as “telling jokes, sharing experiences, discussing philosophy, etc.” These definitions were originally proposed by Cheepen in 1988. Waller emphasizes that interactional communication is critical to develop and maintain relationships but is significantly more complex than transactional communication, with AAC systems primarily supporting the latter. Waller indicates that little has changed since 2006 by restating this position in her 2019 article.

Near-real-time exchanges are odd because they fail to fulfill either goal of establishing or maintaining relationships. While popular today, they are a recent invention, too terse for establishing relationships and too distant for maintaining them. Their information payload is too small to make a convincing point and too remote to communicate the non-verbal cues that humans have evolved to detect and internalize.

Communication systems for individuals with disabilities have, for decades, focused on giving people a voice. Initially, the concern is to support the individual in saying “something, anything,” that can communicate a need or a preference. Next, the goal moves to expanding their access to and use of a larger and larger vocabulary so that their needs and preferences can be expressed in richer and more personal ways (Waller, 2019). Unfortunately, support for more extensive vocabularies in AAC devices is typically accompanied by a requirement to possess or develop complex cognitive and operational skills.

The technologies available for these augmentative and alternative communication (AAC) systems can only communicate in near-real-time or batch modes. An AAC user, if given enough time, could surely prepare a speech in advance of an event and then play that speech on cue; however, as soon as the nature of the communication changes to a real-time format—say, a question/answer session—the quality of communication tends to fall dramatically due to the sizable delays introduced by the AAC system and the individual’s motor abilities.

At times like these, non-AAC user communication partners with insufficient or no AAC knowledge inevitably engage in several negative behaviors. Non-AAC users will interrupt, attempt to complete the AAC user's thoughts, preempt communication from the AAC user, and even discuss what the AAC user intends without involving the AAC user. This behavior often leads to misunderstanding, simplification of ideas, and disengagement by both sides of the discussion. Most importantly, this inequitable dynamic marginalizes, silences, and restricts the self-determination of the AAC user.

Speech Rate

Speech rate is calculated as the average number of words spoken per minute. Alternatively, it may be more accurately calculated as the average number of syllables per minute, since speech often involves words that vary significantly in their number of syllables. Since pre-literate users of AAC systems normally select words rather than syllables, we will focus here on words per minute (wpm). Note that AAC users can only convey the words or concepts that are available to them via their AAC device.

Typical rates of speech vary with the kind of communication taking place, but here are some commonly quoted numbers:

- Presentations: between 100–150 wpm
- Conversations: between 120–150 wpm
- Audiobooks: between 150–160 wpm
- Radio hosts and podcasters: between 150–160 wpm

With a lower bound of 100 wpm for presentations and 120 wpm for conversation, it is probably safe to assume that a speech rate of 80 wpm is close to the lower bound for comfortable conversation. Note that AAC users with significantly limited mobility, especially those using eye gaze as a selection method, normally average between 10 and 20 words per minute (Morris, 2021), more than five times slower than what might be considered functional in a typical conversation.

If we assume that an AAC device requires about half a second to speak a single word, then an AAC user has $60 - 80 \times \frac{1}{2} = 20$ seconds to select those 80 words. That is just $80/20 = 1/4^{\text{th}}$ of a second, on average, to choose each word to maintain an 80-word-per-minute rate. This is admittedly a rough calculation with rough assumptions, but it can be helpful as a ballpark figure.

If a slower speech rate is workable, then there will be additional time for word selection, but there is a limit if one wishes to be understood. If it regularly takes more than one or two seconds to select and voice each word in a sentence, it becomes difficult for the listener to retain and remember the string of words as a whole and to understand what is being said. When a collection of words is strung together in connected speech, the average time to speak each word goes down significantly. However, one must then consider the time required to prepare the string of words in the first place.

Literate users of AAC systems may choose to construct their responses by selecting letters from a physical or virtual keyboard. This gives them access to the entire language but often at a reduced rate of word construction due to a limited range of motion or poor hand function (von Tetzchner, 2018). In a review of relevant research, Koester and Arthanat (2018) found an average speed of 1.7 wpm for individuals employing onscreen keyboard scanning and selection using a switch. Letter selection using eye-gaze can improve text production to 8 to 10 wpm (Waller, 2019). Semantic Word prediction can increase communication rate but incurs a visual scanning cost (Trnka et al., 2007). Compaction (Baker, 1982; 1987) and other encoding strategies like Context-Aware Abbreviation Expansion (Cai et al., 2022) can increase an individual's word production but are accompanied by significant cognitive requirements. Note that a 50% to 100% improvement in speed, given a base rate of 10 wpm, produces only 15 to 20 words per minute and is still insufficient for real-time conversation.

Silence is NOT Golden

However, I believe focusing on words per minute is a red herring. In reality, the mismatch between an AAC user and a non-AAC user in conversation is fundamentally the result of a deeply rooted human aversion to “awkward” silence.

It is uncommon for English-speaking individuals to experience more than a moment of silence between exchanges during a conversation, with typical pauses in turn-taking lasting between a quarter and half a second. If the silence persists for longer than a few seconds, it tends to make people uneasy or disrupt the natural flow of the discussion. Research conducted by Koudenburg et al. (2011) revealed that, in the United States, we become uncomfortable in conversation after just four seconds of silence, and sometimes as little as one second of silence (Jefferson, 1989). Repeatedly encountering multi-second delays in a conversation can cause reactive and anticipatory anxiety in the listener, leading to the kind of conversational sabotage (however unintended) described earlier. Police interrogators regularly take advantage of this aversion to silence and can get suspects to reveal more than they otherwise would simply by sitting quietly and looking at the suspect for extended periods.

Imagine for a Moment

Now, imagine the hypothetical scenario presented in Table 2. Tom and Mark are old friends, and both like coffee. They encounter each other at their local Starbucks. Mark is a non-speaking AAC user, and Tom initiates a conversation. What follows is a relatively formal description of the beginning of their conversation. Suspend your disbelief for a moment.

In the table, Tom’s statements are in the column labeled “Tom,” and Mark’s statements are in the column labeled “Mark.” The sentences in the column labeled “AAC Device Display” are shown on Mark’s AAC device after each of Tom’s statements. The timestamps mark the end of Tom’s statements and the start of Mark’s. The timestamp marks the time at which the AAC’s display has finished updating. The bold statements represent the response option selected by Mark. Both Tom’s and Mark’s statements are verbalized. In Mark’s case, the verbalization is produced by his AAC device.

Table 2: An Example of Facilitated Conversation

Time	Tom	AAC Device Display	Mark
0:00	Hi Mark, how are you doing?		
0:05		<ol style="list-style-type: none"> 1. I'm having a slow morning. I need coffee. 2. About as well as the Yankees. They've lost their last three games. 3. I'm fine. How about you? 	
0:08			I'm having a slow morning. I need coffee.
0:10	You and me both!		
0:15		<ol style="list-style-type: none"> 1. How is your son doing in basketball? 2. You haven't told me about your trip to Disney. 3. Are you OK? 	

Time	Tom	AAC Device Display	Mark
0:18			Are you OK?
0:21	Yeah, I'm fine, just sleepy. I was up late last night watching Netflix.		
0:26		<ol style="list-style-type: none"> 1. What show were you watching? 2. By yourself or with your son? 3. I don't have Netflix; is it worth the price? 	
0:31			By yourself or with your son?

In a little over 30 seconds, Tom and Mark had three exchanges and further cemented their relationship. There is a 3- to 10-second delay between the end of each of Tom's statements and the start of Mark's reply. During that time, Mark is presented, via his AAC device, with three options to respond to Tom. Mark selects his preferred response, which causes that sentence to be vocalized by his AAC device. Later, Mark will provide feedback to his AAC device indicating why he preferred the responses that he did. In particular, he will indicate that he is currently interested in learning more about the relationship between Tom and his son. The next time Tom and Mark meet, the device will provide more response or conversation initiation options to lead the conversation in that direction. It will offer those suggestions earlier or first in the ordered list. Mark's AAC device serves as an interface to a local or cloud-based service that understands a great deal about Mark and his relationships. Its goal is to help Mark participate in real-time, turn-taking vocal conversations in arbitrary settings with acquaintances and strangers. It strives to give Mark a voice that best represents his worldview, kindness, sense of humor, and idiosyncrasies.

What is Going On?

Is this science fiction? Not at all. This scenario leverages technology that, for the most part, is available today. Of all of those technologies, the key one is generative AI. Without it, there would be no response options for Mark to choose from. However, before looking at the available technologies, let us explore the details of one of these exchanges.

Table 3: The First Exchange

Time	Tom	AAC Device Display	Mark
0:00	Hi, Mark, how are you doing?		
0:05		<ol style="list-style-type: none"> 1. I'm having a slow morning. I need coffee. 2. About as well as the Yankees. They've lost their last three games. 3. I'm fine. How about you? 	
0:08			I'm having a slow morning. I need coffee.

At time 0:00 in Table 3, Tom initiates the conversation with the statement, "Hi Mark, how are you doing?" This action is much more complex than it appears at first glance. It is valuable to examine the two phrases in his sentence separately. The first phrase, "Hi Mark," embodies multiple activities:

1. Tom must first recognize Mark’s presence in this setting and then confirm that this individual is a person with whom he is in a relationship.
2. Tom then calls out, “Hi, Mark,” either in an attempt to get Mark’s attention or to acknowledge that he has Mark’s attention.
3. Once he has Mark’s attention—generally in the form of shared eye contact—he generates and voices a greeting that is appropriate for the setting and the amount of time that has passed since he last interacted with Mark. Because Tom is a speaking individual, he can unconsciously generate human-interpretable speech.

This may seem like an overly detailed description of Tom’s greeting, but Mark and his device will also have to accomplish these things if he would like to initiate a similar conversation in a similar setting.

In response, by time 0:05, Mark’s AAC system must:

1. single out Tom’s statement from the surrounding conversations and noise,
2. recognize Tom as the speaker and understand Tom’s relationship to Mark,
3. convert Tom’s utterance from voice to text and parse the words into a meaningful form using natural language processing,
4. generate three appropriate responses for Mark to choose from, with what would likely be Mark’s preferred selection appearing first, and
5. display the responses on Mark’s AAC device (or use text-to-speech to play the choices to Mark in order of likelihood of selection).

An appropriate response, in addition to being representative of Mark’s relationship with Tom, could take into consideration the following factors and thereby improve the relevance of the provided responses:

- the location,
- time of day,
- Mark’s conversational style and preferences,
- Mark’s state of mind and a prediction of Tom’s state of mind,
- any goals Mark may have as part of his relationship with Tom, and
- recent conversations between Mark and Tom.

By time 0:08, Mark has made his selection among the options provided to him, and the AAC system must voice that response at a volume level appropriate for the ambient noise level of the setting. The system must include the proper inflection in its vocalization to ensure the maximum amount of information and meaning is communicated.

This process then repeats until one or both individuals indicate a desire to terminate the conversation, at which point, the exchange goes into a wrap-up phase. Note that people often use a verbal blocking technique to hold the conversation while they prepare their response. Words and phrases like “just a second,” “you know,” “um,” “I think,” and “that’s a good question” are an indication that the individual is deciding on a response, and it would be inappropriate to interrupt them. Such utterances are not to be interpreted as a response in and of themselves. Having the system automatically generate these placeholders may be helpful while the user is choosing among the offered responses or preparing a response from scratch.

Providing Feedback to Improve Performance

Later that day, when Mark has spare time, he reviews his recent conversations and shares his feedback with his AAC device. He critiques the suggested responses regarding both their form and their presentation order. He explains what the device failed to consider in generating responses. This information is then used to modify the “tendencies” of the device and improve its performance in future conversations with Tom and others.

Functional Components of this Scenario

Even to a non-linguist, an overwhelming number of complex tasks must be successfully executed to participate in what looks, at first glance, like a trivial conversation. Most of them appear to take place subconsciously but would have to be explicitly constructed if carried out by a machine. Here are the tasks that stand out.

Preparing for Conversation

1. **Location recognition.** The conversation's physical location often sets the conversation's foundational context, especially when interacting with strangers.
2. **Partner(s) identification/recognition/selection.** At a fundamental level, the communication partner or partners must be identified. That information can then be used to determine the kind of relationship that exists between the participants.
3. **Relationship recognition.** The speakers may be strangers, close friends, or family members.
4. **Partner relationship history.** By knowing the nature of past interactions, it is possible to make predictions about the nature of the current exchange.
5. **Communication goal identification.** Every conversation is entered into intentionally, and each participant has a goal, for example, to reinforce a friendship or persuade a customer to purchase a product.
6. **Goal achievement tactic selection.** For any but the most mundane conversations, some planning must take place regarding how the exchanges within that conversation will be directed.

Initiating the Conversation

1. **Attention capture.** In this step, the conversation's initiator gets the communication partner's attention by saying or doing something. A spoken greeting is an example of capturing the attention of a communication partner.

Turn-Taking

1. **Speech recognition.** There are two fundamental elements to speech recognition. The first involves separating the sound waves of interest from the ambient noise, and if multiple individuals are involved in the conversation, one must attend to all speakers individually. The second element is to parse the sound waves into words, phrases, and sentences.
2. **Conversation continuation.** One turn in a conversation can build on previous turns because the content of those turns is remembered, at least for the duration of the conversation.
3. **Interpretation of facial expressions and body language.** Proper interpretation of the conversation partner's facial expressions, micro-expressions, and body language can be critical to deciding whether a conversation should be stopped or started and whether, at some point, it should be redirected. This information can give the participants information about each other's emotional and intellectual states moment by moment as the conversation progresses.

4. **Response generation.** While this is listed as just a single element of turn-taking, it is where all the heavy-duty intellectual work occurs. It requires integrating and synthesizing all current and historical information relevant to the conversation. Consideration may also be made for cultural influences, emotional dynamics, and social context.
5. **Speech generation with inflection.** The ultimate goal of the AAC system is to render the user's thoughts in the form of audible speech. In addition, the meaning of a set of words can vary greatly depending on which words are emphasized or how the pitch of the speaker's voice changes along the way.
6. **Attention maintenance.** Small utterances of agreement/surprise, head nods, and facial expressions can help maintain the connection by demonstrating solid relationships between the communication partners.

Terminating the Conversation

1. **Termination choices.** Partners must be able to communicate that each has met their conversational goals and that their conversation is now ending. Sometimes, the individuals commit to further discussion and what that conversation may entail.

Relevant Technology

It must be possible to accomplish all of these functions through the application of technology. Many of the technologies listed below have been around for years, but all have significantly improved in accuracy, miniaturization, capacity, and speed.

Situational Awareness

- **GPS.** The accuracy of public GPS receivers depends on weather and obstructions but is typically on the order of 5 to 10 feet. That is more than sufficient precision to determine the physical location in which one is currently situated.
- **Electronic calendars.** By recording your plans about whom you expect to meet and where you expect to meet them in your online calendar, this information can be made available to the system to set expectations for any conversation in which you may be involved.
- **Facial recognition.** Current facial recognition algorithms can predict an individual's identity with over 99% accuracy based on an image of their face. This accuracy is reflected in the fact that many people use facial recognition instead of passwords to log onto their computers and phones.
- **Voice recognition.** A machine's ability to recognize a particular individual based on characteristics of their voice is currently performed with over 98% accuracy and is regularly used during phone conversations with financial institutions to establish the account holder's identity.

Natural Language Processing (NLP)

- **Speech-to-text** is the act of interpreting audio signals produced by the speech of an individual and converting them into a text representation. Current algorithms are 90 to 95% accurate, depending on environmental conditions. Microsoft claims to have a "word error rate" (WER) of 5.1%, while Google boasts a WER of 4.9%. This is comparable to a professional human transcriptionist at 4%.
- **Text-to-speech**, sometimes called speech synthesis, is a technology that can take text as input and produce results indistinguishable from human speech – including sex and age differences, regional accents, and voice inflection.

- **Chatbots** are computer programs designed to simulate a human conversation. They have been used for decades and specialize in retrieving structured information that has been predetermined based on the industry in which the chatbot will be deployed. A simple chatbot will not support open-ended conversations.
- **Sentiment and emotions analysis/classification** uses computers to find and classify emotions in a body of text as positive, negative, or neutral based on the opinions expressed.

Statement/Response Generation

- **Large Language Models (LLM)** are deep-learning algorithms that use huge datasets to recognize, summarize, translate, and predict textual language.
- **Generative AI** is a type of artificial intelligence (AI) that can assemble text, images, or other media from scratch when given instructions or prompts. Generative AI systems normally have an LLM at their core.
- **Relationship modeling** is used in social psychology. Relationships exist as a social connection, link, or tie between two or more people. A relationship model is a formalization of these relationships as nodes (individual people) and connections between the nodes (the nature of their relationships).
- A **worldview** is a set of ideas and beliefs about the world, oneself, and life as a whole; a group of personally tailored theories about how the world works and answers to a wide range of questions. A **worldview model** is a formalization of these beliefs and ideas into a structure representing a particular individual's thoughts and preferences.
- **Response shaping** is the process of teaching or training an organism (or, in this case, a large language model) to generate a specific response by rewarding responses that are close to or match the response one wants it to provide.

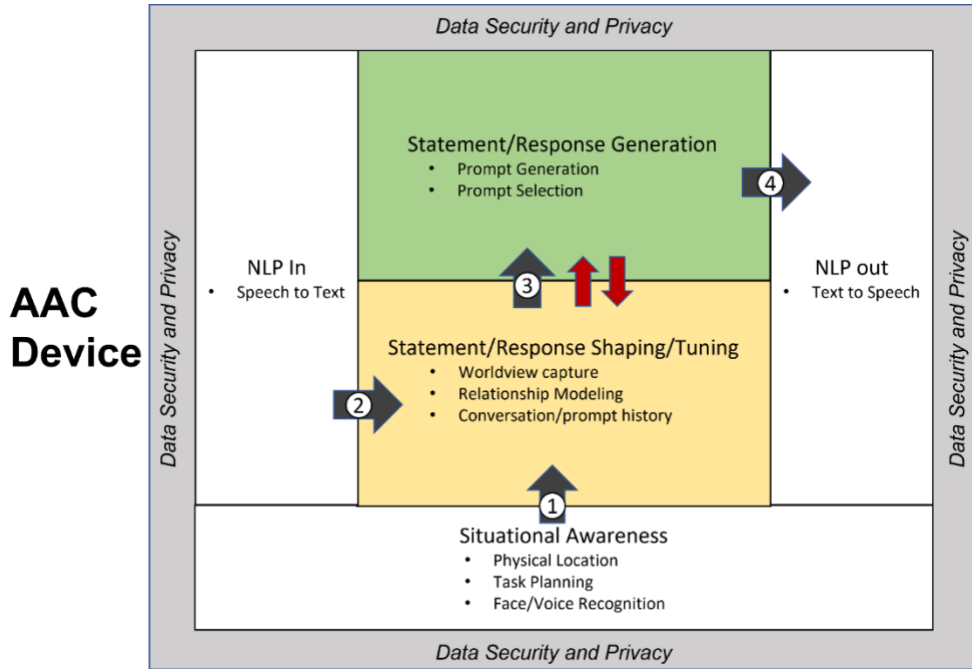
User Interface

- **Tablet/Laptop computers and smartphones** are the most common hardware platforms for use in AAC. They have several built-in capabilities that can make them an excellent starting point for AI-driven communication: microphones, cameras, visual displays, and speakers. Many include GPS receivers and switch access features as well.
- **Eye tracking** is a technology that attempts to determine what a person is looking at by identifying and following the movement and location of the person's pupils. This information can then be used to select among targets in the person's visible surroundings—in particular, potential communication partners.
- **Augmented Reality (AR)** overlays a computer-generated image on an image of the real world to literally "augment" reality. AR goggles or glasses could combine the user's view of their current (or potential) communication partners with AI-generated statement/response options. If the goggles or glasses can monitor the location and movement of the user's eyes, then selecting partners and selecting responses could be driven directly via eye movements.
- **Switch-based selection** is a collection of technologies that recognize an individual's intentional muscle movement by closing of an electric circuit. The movement is then tied to choosing an option in an associated user interface through direct selection or selection scanning methods.

High-Level Architecture

Figure 1 shows a 30,000-foot description of how these technologies might fit together to provide the functionality needed for facilitated real-time communication. Boxes that share a border exchange information. The arrows represent the flow of that information. The large black arrows represent the flow of information that takes place during conversations. The smaller red arrows represent information flow that takes place during system feedback, training, and tuning sessions.

Figure 1: High-Level Architecture – Entirely Local



All boxes are wrapped in a data security and privacy strategy. Because the system is intended to represent a repository of the user's thoughts, opinions, and goals, the data is inherently private and must be protected.

Since this is fundamentally a tool for verbal conversation, both the input and output NLP boxes (on the left and right) are responsible for taking speech as input and generating speech as output. This technology has been around for decades, though it has advanced tremendously. High-quality, speaker-independent speech recognition is now possible, as evidenced by the abundance of "smart speakers" with which we surround ourselves. Note that the recognition can occur in relatively noisy environments and can be associated with a particular individual in order to provide custom responses. Speech generation has, similarly, seen incredible advancement to the point where generated speech sounds completely lifelike and incorporates all the necessary voice inflections.

Functions in the "Situational Awareness" box gather information about the environment to set a context for the conversation. Conversation in a coffee shop is likely to involve brief exchanges with strangers. In contrast, discussions in the workplace are likely to be lengthier, with a need to be sensitive to particular protocols and power dynamics.

The two boxes in the center are where the "sausage" of conversation gets made, and they must work closely together. The "Statement/Response Generation" box is responsible for creating a finite set of reasonable and grammatically correct response options for the user. This box would be implemented using generative artificial

intelligence via a large language model. Creating reasonable and grammatically correct text is where large language models already shine. This box is the key to revolutionizing AAC and propelling AAC users into the center of society. The second box, "Statement/Response Shaping/Tuning," is critical to generating suggested responses that are appropriate for the setting, topic, and communication partner. Most of all, this box ensures that the suggested responses are representative of the sort of things that this AAC user would want to say.

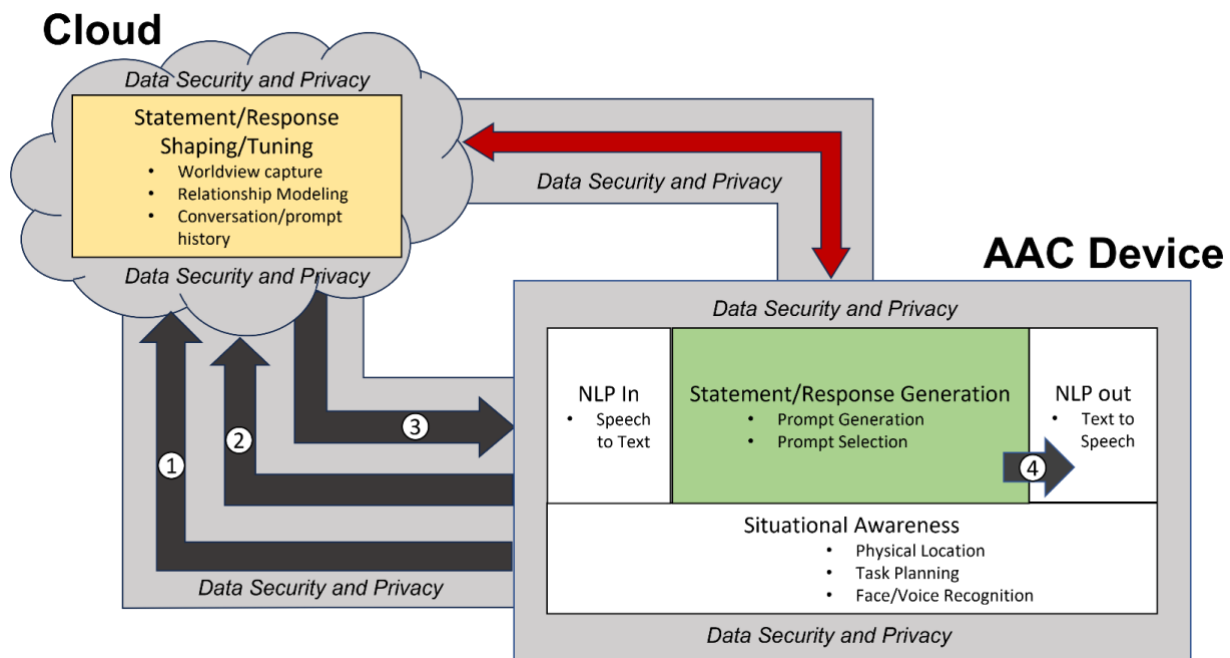
The typical information flow between the architectural components starts with:

1. the context for the conversation coming from the Situational Awareness component,
2. that is then combined with a text representation of what the communication partner has said as processed by the NLP In component,
3. the AAC user's worldview processes both of these pieces of information, and a request for a set of goal-oriented response options is made of the Statement/Response Generation component, which is then presented to the AAC user,
4. the selected response is passed to the NLP Out component for conversion into quality human speech.

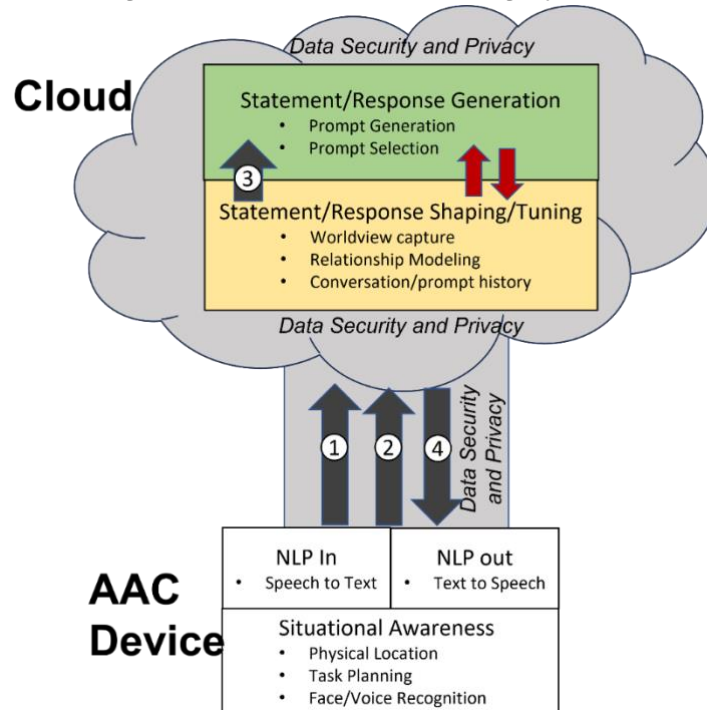
This flow repeats with each exchange in the conversation.

The architecture in Figure 1 puts all functionality directly on the AAC device. This limits protected data flow into and out of the device—thereby reducing security concerns—but places significant hardware and software requirements on the AAC device itself. Figure 2 shows the same components but with the Statement Response Shaping/Tuning component moved to the cloud, possibly for greater functionality or speed.

Figure 2: High-Level Architecture – Local and Cloud



This architecture has a greater need for data security and privacy, both surrounding the individual components and the communication between them. Also, the greater processing speed available in the cloud could be nullified by the communication delays introduced between the cloud and the AAC device. Finally, Figure 3 shows the two most sophisticated and processing-hungry components moved to the cloud.

Figure 3: High-Level Architecture – Largely Cloud-based

This architecture has several advantages over the other two. The cloud can provide the two components with the greatest computational requirements access to the most powerful computer hardware available. By locating these logically sophisticated components in the cloud, one can ensure that they are updated regularly with the latest programming, regularly backed up, and can take advantage of the latest data security/privacy software. By moving these processing-heavy components to the cloud along with the data requiring the most protection, the AAC device itself can be much simpler. Note that mapping functionality to physical hardware is a challenging exercise in practice and even more of a moving target because both the software and candidate hardware are rapidly becoming smaller, more capable, and less power-hungry.

The latter two architectures take advantage of processing in the cloud but, at the same time, place a requirement on the system to have access to the internet. Mobile broadband protocols like 5G, which are needed to support high-bandwidth applications like autonomous vehicles, will probably provide more than sufficient capacity for these designs. There may be an issue when using such a device in a school setting where internet access is more controlled.

Issues Around Privacy for Communication Partners

I specifically identify a need for data protection and privacy for the AAC device user in these architectures. To what degree must information associated with the communication partner or partners also be protected?

Every day, when we converse with others, our brains automatically and subconsciously perform “partner recognition,” both facial and voice. It would be absurd to ask these individuals for permission to remember what their faces look like or the sound of their voices. Similarly, one would not ask permission to remember what a partner said during a conversation and is often appreciated.

However, the approach described here could challenge these intuitions, primarily because the system would create and maintain a record of the specific facial and voice characteristics that made recognition possible, along with a highly accurate record of the conversation, in order to better inform future conversations with these and other individuals. Further complicating the matter is the fact that this information may need to be communicated to the cloud for storage and processing by the artificial intelligence components. These are issues that deserve further thought and, if necessary, legislation.

A Brief Introduction to Generative AI – e.g., ChatGPT

Generative AI tools like ChatGPT are based on an artificial neural network. This kind of AI is particularly good at understanding sequential data, such as sentences in a body of text, because it can pay attention to different parts of the data depending on the context (Vaswani et al., 2017). ChatGPT has been trained on a large amount of text data from the internet. This text data could include books, articles, websites, and more. It learns by predicting what comes next in a sentence and then being rewarded for correct predictions. For example, if you gave it the sentence "The quick brown fox jumps over the...", it would predict that the next word is "lazy" based on the famous English pangram it has seen frequently during its training.

Through this process of predicting the next word over and over again on a vast scale, it starts to learn not just about individual words but also grammar, sentence structure, context, and even some factual information. It also captures some of the biases in the data on which it was trained (Epstein & Hertzmann, 2023). While this process may seem reminiscent of word prediction, it is more like complete thought prediction.

When interacting with ChatGPT, you provide a request or prompt, and the model generates a continuation. It does not know or remember specific documents from its training but uses its learned understanding of language usage to generate contextually relevant and grammatically correct text. This incredible capability already contributes to increased productivity in several professions (Bowles & Kruger, 2023; Noy & Zhang, 2023). There is no reason why similar gains in productivity cannot be achieved by AAC users when this capability is accessible via their AAC systems.

The "generative" part of the name comes from the fact that it can produce new, original sentences and paragraphs based on what it has learned. It is not simply choosing a response from a set of predefined options but instead generating something unique each time, guided by the patterns it learned during training (Vaswani et al., 2017).

Generative AI systems seem to have come from out of nowhere. In reality, they have been around for a decade but were largely ineffective. In the last few years, three enabling technologies came together to create a breakthrough in artificial intelligence: cheap and powerful computer hardware called graphical processing units (GPUs), easy access to terabytes of written text via the internet, and new computer programming algorithms. These resources have steadily increased in availability and capability over time. No one in the industry can explain why the current level of computing power, data, and programming has resulted in this breakthrough.

ChatGPT is only one of the several available generative AI systems. Google has a large language model (LLM) called PaLM, and Meta (the parent company of Facebook) has one called LLaMA. Links to these LLMs are included in the Resources section of this article. It is no coincidence that these systems are associated

with the largest technology companies on the planet. The amount of computing necessary to create them, and therefore the cost, is mind-boggling. ChatGPT was created by a company called OpenAI with several investors, including Microsoft and Elon Musk.

Creating a Worldview Model

A well-designed large language model like ChatGPT is more than capable of quickly generating a collection of grammatically, semantically, and syntactically correct response options all by itself. It is one thing to give the AAC user final control over the statements spoken by the AAC system, but that alone would remove the individuality of the AAC user from the conversation. The goal of this design is not just to speak “for” the AAC user but to speak “as” the AAC user. To achieve that end, we need to exert some control over the response generator so that the proposed responses align closely with the opinions and attitudes of this specific AAC user. To do that, there needs to be a mechanism that can capture this person's personality, knowledge, and worldview. Other terms like “mental model,” “knowledge model,” “mind map,” “digital identity,” “borrowable identity,” and “personal AI clone” are sometimes used to represent the same organized collection of information about an individual.

This worldview information can then be used in at least two ways. First, it can be used to train the large language model so that it naturally prefers concept formation that aligns with the beliefs of the target user. Second, specific, detailed prompting can be provided to the model in preparation for a conversation so that its output is bent/shaped in the direction the user would prefer. The example in the next section takes the second approach.

One can imagine several ways to collect information about an individual’s worldview:

- social media contributions/selections: posts, photos, likes, watched videos, etc.,
- a review of previous writings or spoken statements,
- answers to survey questions,
- holding focused interviews to collect facts about the individual,
- holding mock conversations, and
- interviews with family and friends.

Fortunately, several researchers have already attempted to create a digital, immortal avatar for individuals. In the process, they had to first capture and organize as much information about the individual as possible. Rahnema et al. (2021) proposed an approach that combines symbolic reasoning and data learning as part of a double-loop learning system to create a:

Knowledge structure that represents someone’s intuitive perception of her environment, the relationships between different entities in the environment, and also her way of thinking or reasoning upon the perceived world.

In the Resources section below, I have included a link to a documentary called *Living Forever Through AI: Digital Immortality and the Future of Death* that highlights the work of Rahnema and his team at Ryerson University, RTA School of Media, in Toronto, Canada to create such an avatar for the documentary’s narrator.

The documentary shows one method of establishing a relationship model for the narrator by performing facial recognition on a collection of family photos.

Note that the concept of interest in this document is not the creation of an avatar for digital immortality but, instead, the collection and organization of an individual's worldview that can be used to direct the generation of personalized responses during a real-time conversation.


A Similar ChatGPT Session

Next, let us see how the Starbucks example from earlier can be simulated in a dialog with version 3.5 of ChatGPT. To hold a conversation with ChatGPT, one provides a question along with some context for that question and then asks for a response. In the first exchange shown in Figure 4, I have provided several pieces of context:

1. the location is a Starbucks coffee shop,
2. there are two individuals involved in this conversation – Tom and Mark,
3. Tom is a good friend of Mark,
4. Tom has initiated the conversation by saying, “Hi Mark, how are you doing?” and
5. ChatGPT should suggest three brief responses.

In the “future-AAC” scenario described earlier, the first two pieces of context would have been provided by the “Situational Awareness” component of the architecture, the third piece would come from the Relationship Model, the fourth piece would come from the “NLP In” component, and the last piece would appear in the AAC device's user interface.

Figure 4: Responding to a Request for Conversation Using ChatGPT 3.5

 Now imagine the following scenario. Tom and Mark are old friends and both like coffee. Mark and Tom run into each other at their local Starbucks. Tom initiates a conversation. Tom says, “Hi Mark, how are you doing?” Please suggest three possible brief responses for Mark to say.




ChatGPT can generate these three potential responses in less than a second. They would be displayed on the target AAC system screen or announced to the AAC user audibly.

In the following exchange with ChatGPT, and shown in Figure 5, I provide the response Mark has selected and Tom's corresponding statement. Additionally, in suggesting three new responses for Mark, I add that ChatGPT should consider Mark's goal in this conversation to learn more about Tom's basketball-playing son. This information about Mark's conversational goal when in dialogue with Tom would come from Mark's "Worldview" and is available to the system as soon as it understands that Mark will be involved in a conversation with Tom.

Figure 5: Continuing and Directing the Conversation Using ChatGPT 3.5

CH Mark responds, "Hey there, Tom! All good, thanks. How's life treating you?" and Tom says, "I'm really beat, I had a long night last night." Tom's son plays basketball for his high school team. Mark would like to learn more about how Tom's son is doing. Please suggest three possible brief responses for Mark to say.

 Certainly! Here are three possible brief responses for Mark to inquire about Tom's son's basketball endeavors: 📄 👍 🗨️

1. "I understand, Tom. By the way, how's your son's basketball season going? Any exciting games or standout moments?"
2. "Hey, Tom. How is your son doing on the basketball team? I remember you mentioning his passion for the sport. Any recent achievements or upcoming matches?"
3. "Speaking of long nights, how's your son managing with his basketball commitments? I recall he plays for the high school team. Any interesting developments or memorable games?"

Send a message. ➤

As before, ChatGPT can generate these potential responses in less than one second. By putting Mark in the critical position of selecting among the responses, he maintains control of his half of the conversation. Mark can also provide feedback to the response generator after the fact to further shape, tune, and personalize the suggestions. For example, Mark can say that one does not play a basketball "match." One plays a basketball "game." Ultimately, the goal of both the system and Mark is for the first suggestion to be the most appropriate of the three and, therefore, be visually or audibly available to vocalize first.

There may be situations where more or fewer suggestions would be presented or can be immediately requested. The user interface must also include a "free form" response generation capability as an ultimate fallback.

The feedback process would reduce the need for these alternatives over time as the system becomes more and more effective at predicting Mark's preferred responses.

Personal Autonomy

It would be natural to question whether such an AAC system would put words in the user's mouth or limit their ability to advocate for themselves. The answer to those questions depends on the level of control that the user of the system has over the range of available vocabulary and possible concepts that the system can produce.

1. At a superficial but critical level, this proposal specifies that the user interface should include a fallback capability so that the individual can create a response from scratch.
2. Because such a system has access to the entire language—either those words available to the AI or those that user can construct — the AAC user is not limited to the words or phrases that happen to be programmed into the AAC device.
3. For the “AI-prepared” responses, the individual is the final arbiter of the response that is spoken. A good user interface design would support generating three (or more) new options if the first three did not quite fit the situation. The individual may still choose to go with a less-than-optimal response in the interest of time.
4. The fundamental goal of the system is to speak "as" someone, not "for" someone. That is where the training/tuning and worldview capture come in. This is why follow-up with the AI is critical to improving the suggestions over time.

Another question might be whether a user of the system could be tempted to respond with the first statement offered by the system in the interest of time and thereby lose a measure of autonomy. We encounter a similar decision many times a day when we cannot quite think of how we want to say something or what word to use. If we sense that the conversation will be derailed or terminated if we do not say the one thing that has come into consciousness—no matter how inadequate—we usually go with that option. Smartphones offer functionality that is very much like this when texting. The messaging app will suggest entire responses that can be sent simply by tapping them. They are relatively generic, but they are usually reasonable responses and are less time-consuming and error-prone than generating a response from scratch using the small, virtual keyboard or a smartphone.

Such generic but reasonable responses are a common and critical component of everyday conversation with strangers and acquaintances. “Speaking” them with speed and fidelity could be a highly desirable capability for AAC users. However, by fine-tuning the system through an ongoing review with the AI, the user can increase the probability that the initial response proposed is also the most appropriate.

Is This Approach Only Relevant for Literate Users?

The online Merriam-Webster dictionary defines “literate” simply as “able to read and write.” So, the natural interpretation of this question would be, “Must someone be able to read and write in order to use such a system?” The examples above all assume that the words spoken by the communication partner and the response options presented by the device would be textual, therefore requiring the user to be able to read. However, one can easily imagine a feature whereby the response options could be spoken in a manner similar to auditory scanning. In addition, current LLMs are very good at restating concepts at an alternative grade, age, or Lexile level to match the user's abilities. Lastly, one can imagine the LLM translating the communication partner's speech and the suggested responses into a pictographic representation using some standard or personal graphic image set, which would extend these capabilities to even preliterate individuals.

Roadmap: How Such a System Could Roll Out

As mentioned, this is primarily a story of bringing available technologies together and connecting all the internal wiring. I would expect that the initial components would be the Speech-to-Text => Statement/Response Generation => Text-to-Speech components, which could reside on a user's laptop or tablet. This design would support the real-time generation of reasonable-sounding spoken responses to spoken statements for an arbitrary communication partner. This functionality alone would be a breakthrough in societal integration for AAC users.

Next in the rollout, elements of the Situational Awareness component could be incorporated so that response recommendations would better fit the location associated with the conversation and the partner with whom the conversation is taking place. Standard responses might also be added to the mix, which are statically tied to a particular physical location (e.g., preferred Starbucks coffee order) or conversation fillers (e.g., "Hi, how are you," or "Just a second while I think about that.").

The component that needs the most work and is the most significant challenge to integrate with the others is the Statement/Response Shaping/Tuning component. This functionality will probably come last. In addition to agreeing on a computer-accessible representation of the user's worldview and relationships, gathering this information and giving the user a way to modify and tune it will be challenging. Most challenging task will be securing and ensuring the privacy of this information. That effort will be complicated because putting this processing into the information flow will probably require more computing power and necessitate moving some of the data and processing to the cloud.

Fortunately, at each stage of development, the system adds demonstrable value and is usable as is. Having all components in place and functioning at the highest levels is unnecessary for the system to provide compelling capabilities.

Outcomes and Benefits

So much of the daily life of AAC users is constrained to communication of basic needs and preferences and only with dedicated communication partners. Due to the limitations of their AAC devices, these individuals are separated from the most vibrant and fulfilling aspects of society: real-time interactions with friends, acquaintances, and strangers. We sit on the cusp of a revolution in all aspects of our lives due to the advent of powerful, generative AI. We should expect that work, lifestyles, relationships, and technology will change to accommodate this disruption. AAC is no exception. In this article, I have laid out a case for how and why AAC devices of the near future will incorporate the capabilities of generative AI to provide a real-time voice for people with complex communication needs. In 2013, Judge and Townend surveyed 43 users of AAC devices and 68 AAC professionals to understand the factors related to AAC abandonment. The results of the survey could be boiled down to the systems being:

- hard to use and non-intuitive,
- hard for others to understand because of poor-quality voices or lack of volume control,
- a slow communication rate.

Waller (2009, p. 163) added that,

One of the reasons for the lack of extended conversation is that the design of most augmentative communication systems focuses on the communication of needs and wants. The ability to engage in more complex types of communication, including the sharing of personal narrative seldom develops in people who have grown up using AAC; the operation construction of narrative discourse is prohibitively slow and physically exhausting, and without the experience and technological support to construct and use pragmatically, the ability and desire to extend conversation remains elusive.

An AAC system leveraging situational awareness, modern text-to-speech, and speech-to-text software, along with the capabilities of generative AI shaped by the worldview of the individual user, can facilitate a conversational communication rate that is effortless, personal, and intuitive. Effortlessness is taken to a new level by limiting the user's real-time involvement to simply choosing between three to five options rather than constructing individual words and phrases from scratch by hand, gaze, or switch-press. The user will later engage in more effortful interactions in "offline," low-pressure, and patient exchanges with the AI system to hone the system's real-time behavior.

The technology to accomplish this end is, for the most part, available right now. The remaining challenges are primarily associated with system integration and creating a model of the user that can be used to shape the system's recommendations. The field of generative AI is moving so quickly that many of these challenges may be solved or minimized without additional work from AAC vendors or researchers.

Acknowledgment

The following individuals provided insightful suggestions and thoughtful critiques of this article during its development:

- Bruce Alter, PT, AT Consultant
- Michael Hipple, AAC Communicator
- Christine Holyfield, Ph.D., CCC-SLP
- Terry Jennings, Software Architect
- Mike Marotta, Director, The Richard West Assistive Technology Advocacy Center at Disability Rights New Jersey
- Alanna Raffel, OTD, OTR/L, Assistive Technology Specialist
- Gemma White, M.S., CCC-SLP
- Heather Williams, MAT, AT/AAC Specialist
- Lyle Williams, Adaptive Technology Specialist

Resources

These links may be helpful if you would like to explore generative artificial intelligence, or the other technologies referenced in this article.

Generative AI Systems

- [ChatGPT](#) from OpenAI

- [PaLM](#) from Google
- [LLaMA](#) from Meta
- [Large Language Models](#) sit at the center of all generative AI systems
- [What is Generative AI, and how does it work?](#)
- [Large Language Models Explained](#)

Personal Worldview

- [Living Forever Through AI: Digital Immortality and the Future of Death | ENDEVR Documentary on YouTube](#) (The link begins at the start of the interview with Hossein Rahnama).
- [AI Foundation](#) is a company highlighted in the ENDEVR documentary that aims to develop digital humans with conversation generation.
- [The Primals Project](#) is a survey of primal world beliefs at the University of Pennsylvania.
- [The Big Five Personality Traits](#) try to capture the core dimensions of human personality.
- [Knowledge Graphs](#) provide a generic structure for representing an individual's opinions, experiences, and relationships.
- These links describe the process for downloading your personal data captured by [Google](#) and [Facebook](#).
- [Video on Using ChatGPT with Your Own Data](#) demonstrates an early example of how you can employ a few lines of Python code and your personal data to enhance and personalize your prompting of ChatGPT.

Spoken/Written Language and Facial Expression Processing Technologies

- [Sentiment Classification: A Beginner's Guide](#)
- [Recognizing Human Facial Expressions with Machine Learning](#)
- [Facial Micro-Expression Recognition through Machine Learning](#)

Situational Awareness Technologies

- [Google Glass](#) was first to market but doomed by the privacy concerns of others.
- [Apple Vision Pro](#) is Apple's specialized AR goggles with built-in, accurate eye tracking.
- [Apple Glasses](#) is Apple's "planned" attempt at wearable AR.

Declarations

This content is solely the responsibility of the author(s) and does not necessarily represent the official views of ATIA. No financial disclosures and no non-financial disclosures were reported by the author of this paper.

References

Baker, B. (1982). Minspeak: A semantic compaction system that makes self-expression easier for communicatively disabled individuals. *Byte*, 7(9), 186–202. <https://archive.org/details/byte-magazine-1982-09/page/n183/mode/2up>

- Baker, B. (1987). *Semantic compaction for sub-sentence vocabulary units compared with other encoding and prediction systems* [Conference paper]. 10th Annual Conference on Rehabilitation Engineering, San Jose, CA. RESNA.
- Bowles, D. C., & Kruger, J. S. (2023). Generating employable, intelligent graduates in a world with generative AI: Thoughts for educators. *Pedagogy in Health Promotion*, 9(2), 75–77. <https://doi.org/10.1177/23733799231175171>.
- Cai, S., Venugopalan, S., Tomanek, K., Narayanan, A., Morris, M., & Brenner, M. (2022). Context-aware abbreviation expansion using large language models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1261–1275). Association for Computational Linguistics.
- Cheepen, C. (1988). *The predictability of informal conversation*. Oxford: Printer Publishers Ltd.
- Epstein, Z., & Hertzmann, A. (2023). Art and the science of generative AI. *Science*, 380(6650), 1110–1111. <https://doi.org/10.1126/science.adh4451>.
- Hwang, G.-J., & Chen, N.-S. (2023). Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Journal of Educational Technology & Society*, 26(2), 1–18. [https://doi.org/10.30191/ETS.202304_26\(2\).0014](https://doi.org/10.30191/ETS.202304_26(2).0014).
- Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a ‘standard maximum’ silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166–196). Multilingual Matters.
- Judge, S., & Townend, G. (2013). Perceptions of the design of voice output communication aids. *International Journal of Language and Communication Disorders*, 48(4), 366–381. <https://doi.org/10.1111/1460-6984.12012>.
- Koester, H., & Arthanat, S. (2018). Text entry rate of access interfaces used by people with physical disabilities: A systematic review. *Assistive Technology*, 30(3), 151–163. <https://doi.org/10.1080/10400435.2017.1291544>.
- Koudenburg, N., Postmes, T., & Gordijn, E. (2011). Disrupting the flow: How brief silences in group conversations affect social needs. *Journal of Experimental Social Psychology*, 47(2), 512–515. <https://doi.org/10.1016/j.jesp.2010.12.006>.
- Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin Books.
- Morris, M. (2021). Accessibility as an opportunity and challenge for intelligent user interfaces. *26th International Conference on Intelligent User Interfaces*. <https://dblp.org/db/conf/iui/iui2021.html>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>.

- Rahnama, H., Alirezaie, M., & Pentland, A. (2021). A neural-symbolic approach for user mental modeling: A step towards building exchangeable identities. In *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*. <https://ceur-ws.org/Vol-2846/paper36.pdf>
- Sennott, S., Akagi, L., Lee, M., & Rhodes, A. (2019). AAC and artificial intelligence (AI). *Topics in Language Disorders*, 39(4), 389–403. <https://doi.org/10.1097/TLD.0000000000000197>.
- Trnka, K., Yarrington, D., McCaw, J., McCoy, K., & Pennington, C. (2007). The effects of word prediction on communication rate for AAC. In *Human language technologies 2007: The conference of the north American chapter of the association for computational linguistics; companion volume, short papers* (pp. 173–176). Association for Computational Linguistics.
- Vaswani, A., Shazeer, A., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998-6008). <https://papers.neurips.cc/paper/7181-attention-is-all-you-need.pdf>
- Von Tetzchner, S. (2018). Introduction to the special issue on aided language processes, development, and use: An international perspective. *Augmentative and Alternative Communication*, 34(1), 1–15. <https://doi.org/10.1080/07434618.2017.1422020>.
- Waller, A. (2006). Communication access to conversational narrative. *Topics in Language Disorders*, 26(3), 221–239. <https://doi.org/10.1097/00011363-200607000-00006>.
- Waller, A. (2019). Telling tales: Unlocking the potential of AAC technologies. *International Journal of Language & Communication Disorders*, 54(2), 159–169. <https://doi.org/10.1111/1460-6984.12449>