**Voice Bridge: An Assistive Technology for Nonverbal Communication**

**Submitted by:**

**Maxine Evelyn Hickey**

**Supervised by:**

Dr. Mohammed Mehdi Owrang

Dr. Arthur Gilman Shapiro

Presented to

Department of Computer Science

In Partial Fulfillment of Requirements

**CSC-493 Computer Science Capstone Project**

**American University**

**10 December 2023**

Department of Computer Science

**American University**

[Abstract 3](#_Toc182925961)

[Introduction 4](#_Toc182925962)

[Review 5](#_Toc182925963)

[Design of the Project 6](#_Toc182925964)

[Feasibility Discussion 10](#_Toc182925965)

[Final Implementation 12](#_Toc182925966)

[Results 14](#_Toc182925967)

[Conclusion 17](#_Toc182925968)

[References 19](#_Toc182925969)

# **Abstract**

This paper presents the development of Voice Bridge, an AI-powered assistive technology aimed at providing a nonverbal communication alternative for individuals with Autism Spectrum Disorder (ASD) and other conditions that limit verbal communication. The project, conducted in Visual Studio Code, addresses the significant communication barriers faced by nonverbal individuals, striving to mitigate isolation, frustration, and misunderstandings. The main objective of the project is to develop a user-friendly website powered by AI allowing nonverbal individuals to communicate and express their thoughts, needs, and emotions, while also promoting independence, facilitating inclusion, and improving from existing Augmentative and Alternative Communication (AAC) devices.

The foundational elements of this project include a website featuring various input methods, a Natural Language Processing (NLP) engine, sensory-friendly design, feedback mechanism, and technical support. Python, ML/AI libraries, and Visual Studio Code are essential tools for project completion. The proposed technology has the potential to significantly impact the lives of nonverbal individuals, fostering more meaningful communication and socialization. The project timeline spans from September to December and will require access to literature, a Python environment with ML/AI libraries, Visual Studio Code, and Pachyderm to complete the project.

This report documents the development of Voice Bridge, emphasizing its potential to address communication barriers for nonverbal individuals. The aim is to furnish a user-friendly, AI-powered solution that enables effective communication, fosters independence, and promotes inclusion. The report will conclude with a discussion of the project's final implementation, results, and recommended improvements.

# **Introduction**

In a world that relies heavily on verbal communication, the challenges faced by nonverbal individuals, particularly those with autism spectrum disorder (ASD) and related conditions, are profound. The barriers to effective communication not only led to isolation and frustration but also hinder the formation of meaningful connections. Addressing these challenges, the Voice Bridge project emerges as a pioneering effort to provide a transformative solution—an AI-powered assistive technology designed to offer a communication bridge for nonverbal individuals.

The primary objective of the development of Voice Bridge is to create a user-friendly website interface driven by artificial intelligence, facilitating communication and expression for nonverbal individuals. The project aspires to not only alleviate the challenges associated with existing AAC devices but also to promote independence, foster inclusion, and enhance overall communication outcomes for its users.

The primary focus is to document the journey from ideation to implementation, emphasizing the significance of creating a user-friendly website powered by artificial intelligence. It unfolds by exploring the organizational and developmental intricacies of the project, elucidating the methodological approach undertaken to address the identified communication problem. The subsequent sections will delve into the literature review, project design considerations, feasibility discussions, the actual implementation process, and the results obtained. The report will culminate in a discussion of lessons learned, recommendations for improvement, and the potential impact of Voice Bridge on the lives of nonverbal individuals. The reader will be guided through the project's timeline, objectives, and deliverables, ensuring a thorough comprehension of the initiative's scope and significance.

# **Review**

The development of a chatbot tailored to assist nonverbal communicators, particularly individuals with autism, is a result of a critical examination of past approaches and current challenges in communication interventions. Traditionally, augmentative and alternative communication (AAC) devices have been utilized to address the limitations of nonverbal individuals with autism. However, the limitations of these methods, coupled with recent advancements in natural language processing and artificial intelligence, call for a new solution. Current research highlights the growing interest in leveraging technology to improve communication outcomes for this population, necessitating an exploration of comparable projects and gaps in the existing research. Researchers suggest that "by using AI, AAC devices can adapt to the individual user’s needs, preferences, and communication style, making it easier for them to express themselves in a natural way" (Sarathy), including improved tone, creativity, and overall self-expression. This chatbot project aims to bridge these gaps by synthesizing insights from the historical context, technological landscape, and ethical considerations, contributing to the field of assistive technologies, and enabling better communication and social interactions for nonverbal communicators, especially those on the autism spectrum.

# **Design of the Project**

The Voice Bridge project aims to develop an assistive technology tool to address the major communication barrier faced by nonverbal individuals with conditions that limit verbal communication. The project considers potential users, user-friendliness, performance, compatibility with other technologies, functionality, acceptance, convenience, capacity, and availability. Additional objectives include promoting independence, facilitating inclusion, and improving upon current AAC devices. The project plans to respond to this need by creating a user-friendly interface with a commitment to accessibility and inclusivity, using the power of AI and natural language processing to create an adaptable communication system.

The UK National Autistic Society, in collaboration with Hassell Inclusion, has developed web accessibility guidelines to enhance the user experience for individuals with autism. The elements of accessible web pages are detailed below in order of importance.

Page Layout: Pages should be clutter-free with a clear design, simplicity, and inclusion of only elements relevant to the current user task. The length of pages should be appropriate for users with minimal scrolling. Sections should be clearly separated to help users visually discern elements within or outside the same section. All important information should be placed at the top of the page and above the fold. Design elements should be aligned and laid out symmetrically.

Navigation: The design of the web app must feature consistent navigation across the site to enable users to understand how to navigate to different pages. Links must be clearly visible and distinct from other text and non-clickable elements, achieved using color and underlining.

Colors: The color palette of the website should consist of simple, soft colors and color combinations. Autistic individuals prefer low-contrast color schemes, and users should be able to choose background and text colors that suit their needs.

Fonts: Font size on pages should be no smaller than 12 pt. to ensure text readability. The typeface on the website must be consistent, and a sans-serif font is recommended for better readability. Users should have the option to choose font size based on their needs.

Text: Ensure sufficient spacing between paragraphs and lines of text. Organize text into short paragraphs and use short lines for improved readability. Avoid splitting text with elements that may interrupt the flow of reading and eliminate scroll-stoppers. Limit the use of bold text to highlight keywords or important phrases in the text.

Non-textual formats: Consider providing information in video format with captions and transcripts. Use visuals to facilitate understanding of textual information and help users locate required information on the page. Meaningful and memorable icons should be used to assist users in finding information.

Images: Use simple images that can be understood briefly without the need for deep visual analysis. Ensure images add value and clarity to the text information on the page. Use background images sparingly, especially if overlaid with text. Decorative graphs should not distract users from reading text. Non-patterned background images on content pages facilitate uninterrupted reading. Only photograph people relevant to the story and allow users to turn off decorative graphics and all images on the website to focus on textual information.

Video and Audio: Avoid auto-playing content on pages, even with sound off, as users prefer control over when and what content plays. If auto-play is implemented, provide an option to turn it off. Include captions for all videos and provide transcripts for both video and audio, ensuring accessibility for text-first users. Allow users to toggle captions according to their preferences. Additionally, offer text-based introductory information for videos, enabling users to read content before playing. Ensure that video content enhances the clarity of textual information on the site.

Movement: Refrain from using pop-ups, moving, or animated design elements. If movement is unavoidable, implement a pause mechanism to return control to users. Small, animated movements may be utilized to draw attention to information or prompt users to act, especially when other visual emphasis methods have failed.

Help Pages: Provide effective support for all autistic users encountering issues on the site. Supply help in various formats and channels to ensure accessibility.

Forms: When designing forms, use unique labels and clearly explain the information that needs to be entered. Offer immediate feedback to allow users to quickly identify and rectify errors.

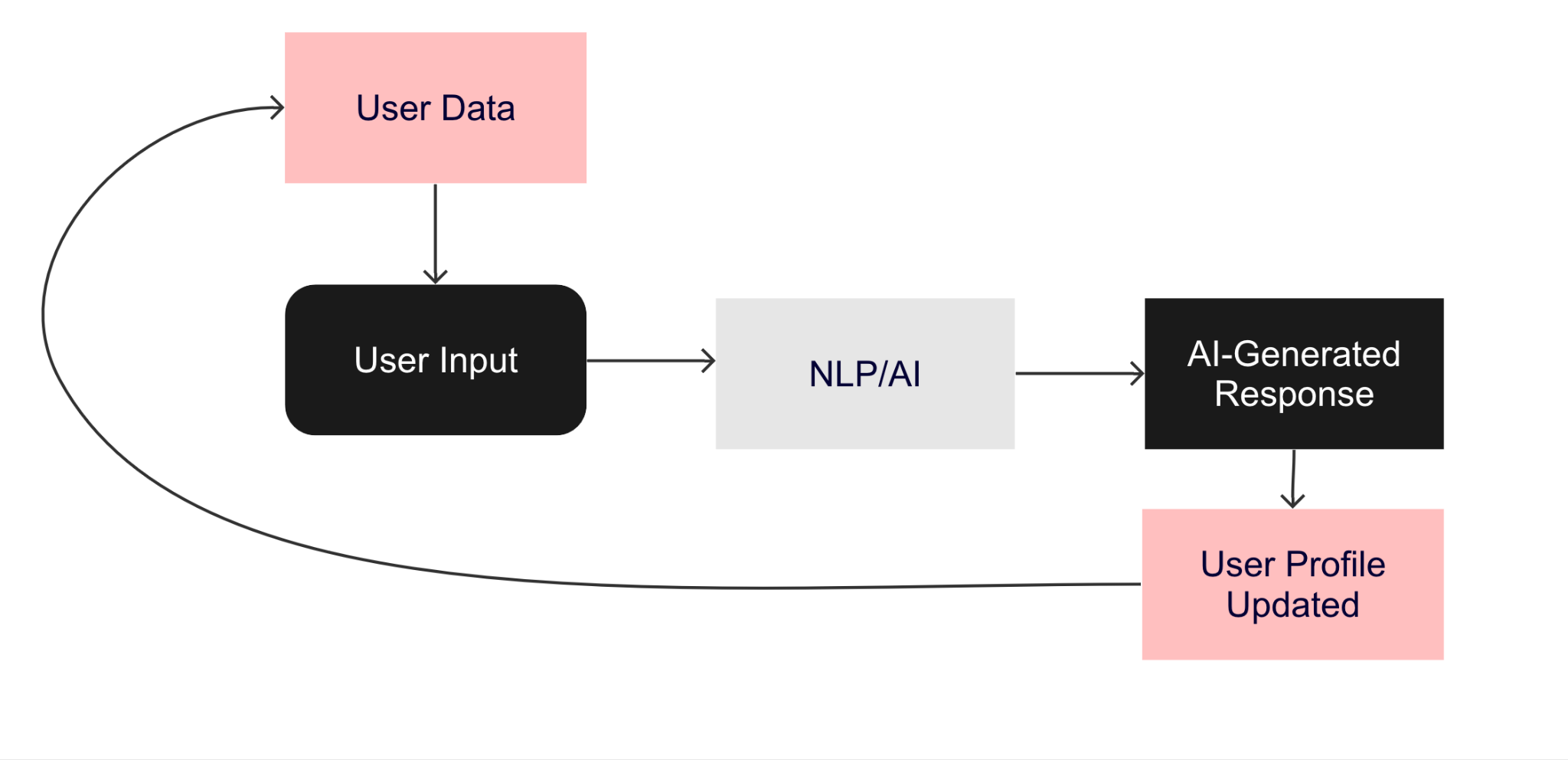
Use of Language: Utilize clear, precise language and avoid metaphors and ambiguous terms. Steer clear of specialist niche terminology or colloquial language on the site. Elect for clear and concise language familiar to users.

Customization: Enable users to tailor their website experience to their needs and preferences. Essential customization tools should include options to change font type, size, text and background colors, the ability to turn off decorative graphics, and the ability to disable captions for video content. Make accessibility tools clearly visible and user-friendly.

Re-learnability: Announce planned changes to the website in advance, informing users about what changes are imminent and when they will be implemented.

The deliverables from the project include a comprehensive assistive technology solution with key deliverables including a Natural Language Processing (NLP) engine, a sensory-friendly design adhering to sensitivity guidelines, a robust feedback mechanism for iterative improvements, accessible technical support, and training materials/documentation. Developed within the Visual Studio Code environment, the project leverages ML/AI libraries and Hugging Face to enhance functionality. To illustrate the system's intricate workings, a Data-Flow Diagram (DFD) has been created, visually capturing the flow of information between the website and the NLP engine, fostering transparency and efficiency in user input, feedback mechanisms, and data processing.

Simplified Data Flow Diagram



# **Feasibility Discussion**

The literature search for the Voice Bridge project provided valuable insights into the feasibility of the project across various factors. The project aims to develop an assistive technology tool powered by AI to enable nonverbal individuals with conditions such as autism spectrum disorder (ASD) and speech disorders to communicate effectively. The literature review highlighted the importance of addressing the major communication barrier that nonverbal individuals face and the potential of AI to transform assistive technology for nonverbal communication, making it more accessible and customizable.

Economic Feasibility

The project's economic feasibility is supported by the literature, which emphasizes the need for more affordable and customizable assistive technology for nonverbal individuals. Typically, AAC devices cost between $1000-15,000 (Barnett, MCSD/CCC and Persels). The cost of the project is free because it should align with the need for accessibility and affordability.

Environmental and Sustainability Feasibility

While there is no direct environmental impact and sustainability of the project, the focus on creating a more inclusive life for nonverbal individuals aligns with sustainability goals by promoting equal access to communication and independence for users.

Manufacturability Feasibility

The project's focus on developing a website, powered by AI, indicates a reliance on off-the-shelf components such as AI and NLP libraries, making the project more feasible in terms of manufacturability.

Technical Feasibility

The use of reliable access to a Python environment and Visual Studio Code further supports the technical feasibility of the project. There is also technical feasibility of the project by highlighting the potential of AI to improve AAC devices and overall communication for nonverbal individuals.

Ethical and Political Feasibility

The project emphasizes the ethical need to address the communication barriers faced by nonverbal individuals and the potential of AI to create a more inclusive life for them. There are no direct references to political issues related to the project.

In conclusion, the literature search provided strong justification for the Voice Bridge project, highlighting the economic, technical, and ethical feasibility of developing an AI-powered assistive technology tool for nonverbal individuals. The project's focus on affordability, accessibility, and inclusivity aligns with the literature's emphasis on the need for customizable and more efficient technology for nonverbal communication.

# **Final Implementation**

A functional website has been developed featuring a rough majority of the planned interfaces, including a baseline text generation interface, which aims to take limited input from the user reflective of their and provide text output that re-articulates the input with more depth and expression, serving as a more comprehensive and expressive representation of the user’s thoughts and emotions. The language model currently implemented is Pythia-160M. This model is also leveraged to drive a next-word predictive text interface, in which a user is given an initial set of four suggested words to use in their prompt. If a user selects a word, it appends to the last place in the input text box, and all four suggested word options are automatically updated to reflect a new set of words predicted by the model, based on the previously selected word. Lastly, a working implementation of text-to-speech is completed at a foundational level, which consists of a text-to-speech Hugging Face model (currently Bark) that inherits the generated text output and performs inference to generate an audio clip (in .WAV format). This audio clip features an artificial spoken representation of the user’s previously articulated expression from the language model and becomes available once inference is completed.

With unique personalization needs for each user in consideration, several options for personalization of the website’s appearance for each user have also been implemented to date, including the ability to adjust the text size input text, output text, suggested word text, and the input field. The user is also now able to cycle through a palette of sensory-friendly pastel colors by clicking repeatedly on a dedicated button (this was perceived as ideal over a dropdown list of the color options, which would have involved the placement of multiple colors in a similar region, potentially damaging the sensory-friendliness of the interface). Currently, all personalization settings are stored only for the duration of the server session and will not be retained after the termination of the session. Personalization functionality does not interfere with the Flask server, with all functions (increasing/decreasing text size, changing background color, and locally storing these values) written in JavaScript, albeit utilizing the jQuery library for consistency with planned future functions that will require access to the server via AJAX methods that are more easily written using the library.

Lastly, several frameworks have been set in place to provide foundation for further planned developments. A key envisioned component of Voice Bridge is the ability of the platform to remember information about the user, including submitted feedback, usage information and eventually other identifying traits as part of the mission for inclusivity. The purpose of this new intelligence is to allow the application to eventually become tailored to the specific needs, preferences, and personality of each user. While a working implementation of these learning mechanisms, foundational steps have been taken. SQLite, recognized for its convenient accessibility via Python, is used to manage a local SQL database containing each input of the user and respective output from the language model, and is automatically updated every time the inference pipeline completes (a message/TTS clip is generated upon user input). While databases could be leveraged to record additional information about the user in the future, this current implementation of database management is viewed in a hopeful way as aiding the learning process by readying data for the necessary future fine-tuning of the text model.

# **Results**

While most vital features of the program have been successfully implemented at a fundamental level (e.g. text can be generated based on input, speech clips can be generated based on generated output text, usage information can be stored), significant challenges were encountered regarding the performance of the AI architecture – particularly the language model. The objective is to leverage the capabilities of language models to deliver a new articulation of limited input from the user that more comprehensively expresses thoughts and emotions for them, with learning capabilities tailoring model output to the user over time. Though the model can successfully generate text based on user input for display on the webpage and conversion to a speech clip (if this option is enabled by the user), the resulting output often exhibits little relevance to the input, outrageously so on some occasions. Likewise, the model’s prediction of next words is not consistently relevant to the previous selected word. It is observed that, in direct contrast to the goals, performance currently seems proportional overall to the depth of the input prompt, in that the lack of output relevance or coherence increases the more limited the input is.

It is realized that even the most advanced language models, which are not used here, are not specially designed or equipped for the intended use case. When the model receives limited raw input, there naturally exists little guidance for the output. This is further complicated by language models not being attuned to the unique task of utilizing limited input and expanding upon it specifically. A certain framework must exist to allow a language model to understand this specific task. As a starting point, certain prompt-engineering practices were attempted in effort to mitigate illogical output. This entails designing a larger prompt in the backend (not visible to the user) that provides guidance for the model regarding the desired output before appending the user’s input at the end. It is this larger prompt that is presented to the model. The solution of few-shot in-prompt learning is currently implemented, where several hard-coded examples are provided to the model that demonstrate the unpacking of certain limited inputs into something more comprehensive and meaningful. This follows the Q&A format:

|  |
| --- |
| *“Q: (example of user expression, limited input). A: (example of unpacked input, as expected output). Q: [user’s actual prompt, retrieved from frontend]. A:”* |

The performance gain from this methodology is minimal. Upon examining the performance of the model after employing this method, it becomes evident that severe limitations exist when confined to small language processing models. These limitations, as understood throughout the development of this project, include the inability for contextual understanding, recognizing patterns (such as few-shot prompting), and recognizing sentiment. Moreover, it is tantamount to acknowledge that the language model currently deployed originates from a group of models that are heavily intended for research purposes, rather than for deployment on programs. Further, it is stipulated by EleutherAI that Pythia models have not been subject to fine-tuning practices akin to that of more advanced languages models as later iterations of OpenAI’s GPT architecture, limiting its ability to adhere properly to the instructions of the user and mold overtime in accordance with their feedback. It is believed that the quality of text generation would increase significantly under several conditions, including the use of a larger-in-scale model, training on myriads of different text content, incorporating sentiment analysis, and rigorous fine-tuning practices to adapt the model to the unique instructions associated with this program’s task.

In the case of utilizing more complex models and executing the suggested fine-tuning training tasks, the process is severely limited at this time by hardware constraints due to complications in feasibly obtaining a cloud-based framework, through which the application would entail more streamlined mechanisms for fine-tuning on user data and feedback, as well as exhibit significantly faster model inferences (if cloud GPUs are utilized), which swiftly eliminates the current computational burden of the user’s local machine.

# **Conclusion**

The Voice Bridge project represents a stride in addressing the formidable communication challenges faced by nonverbal individuals, particularly those with autism spectrum disorder (ASD) and speech disorders. Grounded in thorough research, the project recognized the pivotal role in assistive technology and leverages the potential of AI to revolutionize communication accessibility, customization, and affordability for this specific population.

The completion of essential deliverables, including the user interface website, NLP engine, sensory-friendly design, feedback mechanism, technical support, and comprehensive training materials/documentation, reflects the tangible outcomes of the project. These components collectively form a promising assistive technology tool with the potential to empower nonverbal individuals in expressing their thoughts, needs, and emotions while engaging in meaningful social interactions.

While the project demonstrates considerable success, reflections on its performance suggest opportunities for improvement in scheduling and planning. A closer examination of these aspects could lead to enhanced efficiency and resource utilization, ensuring a more streamlined development process. Regular progress reviews and adaptive adjustments are vital components to address unforeseen challenges and ensure the project's continued success.

Considering the constraints of a semester time frame, the prospective impact of Voice Bridge on the lives of nonverbal individuals is considerable, presenting a promising avenue for a more accessible and customizable communication tool. To fully harness this potential within the limited timeframe, a steadfast dedication to continuous planning, thorough evaluation, and adaptability proves crucial. The insights gained underscore the dynamic nature inherent in the development of assistive technology, highlighting the necessity for flexibility and ongoing refinement to effectively address the evolving needs of the user community.

# **References**

“Accessibility.” *National Autistic Society*, https://www.autism.org.uk/what-we-

[do/website/accessibility](https://www.autism.org.uk/accessibility).

Barnett, MCSD/CCC, Laura, and Tina Persels. “Augmentative Communication (AAC).”

*Medical Home Portal*, February 2019, <https://www.medicalhomeportal.org/living-with-child/assistive-technology/augmentative-communication#d31278155e383>.

“How Web Accessibility Affects People with Autism.” *Bureau of Internet Accessibility*, 2 April

2018, <https://www.boia.org/blog/how-web-accessibility-affects-people-with-autism>.

Ledbetter, Jason. “Revolutionizing Autism Care: The Transformative Role of Artificial

Intelligence in Diagnosis, Communication, and Therapy.” *LinkedIn*, 3 August 2023, <https://www.linkedin.com/pulse/revolutionizing-autism-care-transformative-role-jason-ledbetter/>.

Macias, Melba. “AI Helps Non-Verbal Man to Communicate.” *Enabling Devices*, 5 August

2021, <https://enablingdevices.com/blog/artificial-intelligence-helps-non-verbal-man-to-communicate/>.

Raymaker, PhD, Dora M., et al. “Development of the AASPIRE Web Accessibility Guidelines

for Autistic Web Users.” *NCBI*, 13 April 2019, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485264/>.

Sarathy, Divya. “AI and AAC: How AI can Empower AAC Users.” *Avaz Inc.*, 23 February

2023, <https://avazapp.com/blog/ai-and-aac-how-ai-can-empower-aac-users/>.