

Helping or Homogenizing? GenAI as a Design Partner to Pre-Service SLPs for Just-in-Time Programming of AAC

Cynthia Zastudil

Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
cynthia.zastudil@temple.edu

Christine Holyfield

Communication Sciences and
Disorders
University of Arkansas
Fayetteville, Arkansas, USA
ceholyfi@uark.edu

Christine Kapp

Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
christine.kapp@temple.edu

Kate Hamilton

Psychology
Temple University
Philadelphia, Pennsylvania, USA
kate.hamilton@temple.edu

Kriti Baru

Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
kriti.baru@temple.edu

Liam Newsam

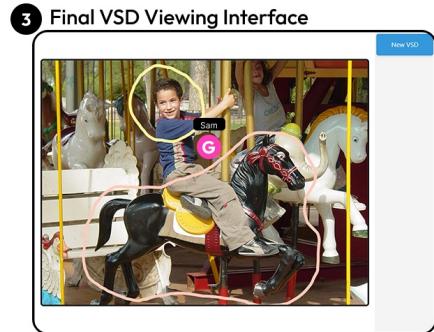
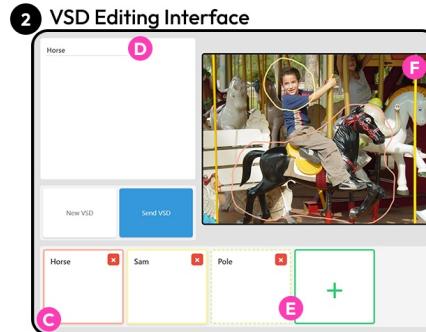
University of California, Berkeley
Berkeley, California, USA
liamcnewsam@berkeley.edu

June A. Smith

Computer Science
Berea College
Berea, Kentucky, USA
smithj7@berea.edu

Stephen MacNeil

Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania, USA
stephen.macneil@temple.edu



A VSD programmers can **upload** an image of a past activity or social interaction

B VSD programmers can **capture** an image of an actively occurring situation

C **Generative AI-suggested** hotspots are populated at the bottom of the screen

D VSD programmers can **edit hotspots** to modify the text content as necessary

E VSD programmers can **delete generated hotspots** or **add new hotspots** as necessary

F VSD programmers can **draw the hotspots** on the image on the image canvas

G When a hotspots is pressed, hotspot content is **displayed** and **spoken aloud**

Figure 1: The three screens that comprise the user interface of our prototype. (1) VSD programmers choose to upload or capture an image to use in a VSD. (2) The application automatically generates a set of potential hotspots for use in the VSD and the programmer can choose to edit, use, or delete these hotspots. They can also manually add their own hotspots. Once all of the hotspots are created, they can use the canvas to draw the hotspots on the image on the image canvas. (3) Once they have finished configuring the VSD, users can see a preview of what the VSD looks like and interact with the created hotspots.

Abstract

Augmentative and alternative communication (AAC) devices are used by many people around the world who experience difficulties in communicating verbally. One form of AAC device which is especially useful for minimally verbal autistic children in developing language and communication skills is the visual scene display (VSD). VSDs use images with interactive hotspots embedded in them to directly connect language to real-world contexts which are meaningful to the AAC user. While VSDs can effectively support emergent communicators (i.e., those who are beginning to learn how to use symbolic communication), their widespread adoption is impacted by how difficult these devices are to configure. We developed a prototype that uses generative AI to automatically suggest initial hotspots on an image to help non-experts efficiently create visual scene displays (VSDs). We conducted a within-subjects user study to understand how effective our prototype is in supporting non-expert users, specifically pre-service speech-language pathologists (SLPs) (N=16) who are not familiar with VSDs as an AAC intervention. Pre-service SLPs are actively studying to become clinically certified SLPs and have domain-specific knowledge about language and communication skill development. We evaluated the effectiveness of our prototype based on creation time, quality, and user confidence. We also analyzed the relevance and developmental appropriateness of the automatically generated hotspots and how often users interacted with (e.g., editing or deleting) the generated hotspots. Our results were mixed with SLPs becoming more efficient and confident. However, there were multiple negative impacts as well, including over-reliance and homogenization of communication options. The implications of these findings reach beyond the domain of AAC, especially as generative AI becomes more prevalent across domains, including assistive technology. Future work is needed to further identify and address these risks associated with integrating generative AI into assistive technology.

CCS Concepts

- Human-centered computing → Accessibility technologies; Accessibility systems and tools.

Keywords

augmentative and alternative communication, AAC, autism, visual screen displays, VSDs, generative AI, just-in-time programming

ACM Reference Format:

Cynthia Zastudil, Christine Holyfield, Christine Kapp, Kate Hamilton, Kriti Baru, Liam Newsam, June A. Smith, and Stephen MacNeil. 2025. Helping or Homogenizing? GenAI as a Design Partner to Pre-Service SLPs for Just-in-Time Programming of AAC. In *The 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25), October 26–29, 2025, Denver, CO, USA*. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3663547.3746384>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASSETS '25, Denver, CO, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0676-9/25/10
<https://doi.org/10.1145/3663547.3746384>

1 Introduction

Many individuals with complex communication needs, such as those with autism, apraxia, or other speech and language disorders, rely on augmentative and alternative communication (AAC) to supplement or replace verbal communication in their daily lives. AAC often takes the form of technologies such as grid displays [62] or visual scene displays [14], which present communication options that are spoken when selected. Visual scene displays (VSDs) use images with embedded words or phrases in the form of selectable hotspots [14]. VSDs are often used by emergent communicators (i.e., communicators who are working on learning and using symbolic language [27]) because they help to contextualize the communication options within a scene [59, 60]. This is especially helpful for young, minimally verbal autistic children¹ who are still developing symbolic language skills and benefit from concrete, context-based representations of meaning [59, 60]. Similarly, VSDs group concepts together in an image taken from the real world, which preserves the relationships between people and objects [59]. As a result, VSDs can reduce the cognitive demands of using AAC by aligning with the natural visual processing of scenes [62]. These benefits distinguish VSDs from other AAC layouts including the more common grid displays that display symbols, which are often line drawings that do not as clearly reflect referents (i.e., the object, concept, or activity to which the symbol refers), displaying them in isolation without maintaining the relationship between words and the real world [70].

Given how effective VSDs are with emergent communicators, they offer a promising solution for use outside of clinical contexts when SLPs are not present, such as in the home. Although they can be easier to use by communicators, VSDs are challenging to configure and often require experts to 'program' them by choosing the hotspots for each image. There is not a universally agreed upon set of guidelines, so expertise is needed to effectively configure them. It can be incredibly time consuming to configure each image with relevant communication options, so VSDs are often configured ahead of time for a specific communication setting. This manual configuration requires frequent updates to remain relevant for their users' communication needs [28]. Another way of configuring VSDs is just-in-time (JIT) programming [80] (i.e., capturing an image of an activity, social interaction, or object and configuring it in real-time for the VSD user). JIT programming supports communication partners in creating VSDs in the moment to take advantage of spontaneous opportunities for communication [80]. This approach still requires significant knowledge about how and when to program VSDs [44]. Parents and other communication partners outside of clinical settings often lack this expertise, preventing widespread adoption of VSDs [11, 41].

Prior work has investigated the use of AI to generate high-quality, contextually relevant topic-specific communication boards [34] and text-based communication suggestions [89]. We are interested in whether generative AI (genAI) can be used in AAC devices to reduce the effort required to program high-quality VSDs by scaffolding the VSD configuration process. While genAI may speed up the

¹In this paper, we primarily use identity first language (e.g., "autistic children"), as recent research has found that many individuals prefer identity first language [81]; however, this is not a universal preference and some people also prefer person-first language [84] (e.g., "children on the autism spectrum").

process, there are still some open questions. For example, it is vital that VSDs be programmed in a personalized way for each end user, taking into account their communication stage and personal interests and experiences [44]. Prior work has shown that genAI does not do a good job at incorporating personalized information about end users without very detailed prompting [97], which is not feasible for the just-in-time programming of VSDs. Rather than focusing on developing hyper-personalized user models which would be environmentally and financially expensive to build and would become outdated quickly, requiring frequent updating to stay up to date with users' communication needs, we are interested in ways to leverage communication partners' unique knowledge about VSD users to address the need for personalization.

In this paper, we investigate the impacts of integrating genAI into VSDs by designing and evaluating a prototype that uses genAI to scaffold the configuration of hotspots for VSDs. We evaluated this prototype with pre-service speech-language pathologists (SLPs) to understand the ways that it improves VSD configuration and also the potential negative consequences of integrating AI in this context. We conducted a within-subjects user study with our prototype and an existing VSD application as our control condition to investigate the impact our prototype has on VSD configuration done by pre-service SLPs.

Given these methods and research goals, we investigate the following research questions:

- RQ1:** How effective, in terms of creation time, quality, and user confidence, is a genAI-assisted VSD application in supporting untrained communication partners?
- RQ2:** Does a genAI-assisted VSD application provide relevant and developmentally appropriate hotspots?
- RQ3:** What are the impacts of genAI suggestions on the content of VSD hotspots?

Based on an analysis of the configured VSDs and participants' confidence and perceptions of the automatically generated hotspots, our key findings are:

- **VSDs were configured faster with increased confidence.** Participants were faster and more confident using our prototype with genAI suggestions.
- **Mixed results for VSD quality.** There were mixed results regarding the quality of the VSDs. While the VSDs configured with our prototype tended to use more developmentally appropriate hotspots, they also tended to contain too many hotspots, hotspots that were not directly related to the scene, and provided less hotspots for desired communication functions (e.g., socialization).
- **Increased homogeneity.** When using the VSD with generative AI suggestions, the resulting communication options were more semantically similar to the average VSD configured by participants.
- **Over-reliance on genAI suggestions.** Participants did not commonly delete or modify genAI suggestions, commonly accepting the majority of the suggestions provided to them with few edits or additions.

Our work makes the following contributions:

- A prototype which leverages genAI to provide a starting point for VSD creation, which could be potentially useful

in training settings for pre-service SLPs or other untrained communication partners.

- Empirical contributions measuring the efficiency, confidence, quality, and homogeneity of VSDs configured with genAI.
- Identification of implications for integrating genAI into AAC devices, including over-reliance on AI and increased homogenization of AAC displays.

Our prototype examines the integration of genAI into an existing state-of-the-art AAC device, to better understand what kinds of guardrails may be needed to address some of the negative impacts we identified of AI integration into AAC devices. In our discussion we talk about ways in which different HCI approaches may be able to address these issues. Across domains, researchers are advocating for designing tools with guardrails which incorporate best practices. For example, this is common in education [25, 57], creativity research [82], and data analysis research [26].

2 Related Work

2.1 Visual Scene Displays

For people with complex communication needs, AAC intervention can aid in the development of language, literacy, communication, and cognitive skills [23]. As an alternative to the traditional grid layout, visual scene displays (VSDs) are images of meaningful events or shared activities in which vocabulary is embedded within the image [14]. For example, a photograph from a school dance could be used as a VSD with the faces of friends in the picture programmed to output their names or a picture of people dancing could be programmed to output "dance" [45]. This way, language is grounded within a familiar context for the AAC user due to its personally relevant meaning [45]. VSDs are frequently used by emergent communicators (i.e., those who do not reliably communicate with symbolic language [27]), such as minimally verbal autistic children [36, 37]. VSDs have also been found to be useful for people with aphasia [13, 16].

VSDs can offset some of the cognitive load required to communicate messages using AAC devices for individuals with cognitive or linguistic challenges [62] because they maintain the relationship between people and objects as they experience them in real life [59], represent social interactions through which language is learned [60], and reduce working memory demands because all people, objects, and activities are presented together [60]. VSDs also have the potential to provide communication partners ways to get more involved in the communication process [14] by working with AAC users to configure effective and engaging VSDs.

2.2 Just-in-Time Programming

To make AAC more contextually relevant in each moment, communication partners will often configure VSDs with language options right when they are needed. This technique is referred to as just-in-time (JIT) programming [80]. JIT programming allows communication partners to program new vocabulary in the moment based on events that unfold and interests demonstrated within them [44].

JIT configuration allows clinicians to configure communication options for VSD users in real-time associated with an image or naturally occurring scene and it supports taking advantage of teachable

moments as they occur [80]. However, the JIT programming approach raises questions regarding the risk of responding to false indicators of engagement or missing language learning opportunities [44]. Similarly, despite the flexibility JIT configuration provides, there are still concerns as to whether clinicians can adjust accordingly when faced with environment changes or unforeseen situations [80]. It also requires clinicians to be present and continuously reprogramming the model to capture engaging scenarios [44]. Automated JIT support may offer a simpler and more efficient approach to creating communication options that are relevant to an immediate context.

Automated JIT scaffolding has been explored in other research domains including data visualization literacy [66], explaining complex concepts in scientific research papers [7, 43], and in creativity research [65]. In an AAC context, prior work has investigated the use of JIT scaffolding through automatically generating topic-specific displays [34], providing text-based suggestions for starter phrases in conversations [89], using communication partner speech as input to program a topic display JIT [47], and the impact of JIT programming on social participation of autistic children [48]. There is also a significant amount of research in AAC dedicated to automation in AAC devices, including text prediction, abbreviation expansion, and keystroke automation [22].

2.3 AI-Enabled AAC Devices

Numerous studies have investigated whether and how AI can be integrated into AAC devices, including using computer vision algorithms to generate contextually relevant communication options based on a photograph [33, 71, 74], gaze-based AAC [31, 99], and automatic recognition of communication context (e.g., location or communication partner identity) [51]. Highly performant genAI provides new possibilities, and researchers have been exploring the ways in which it can be used to further enhance assistive technology. In the context of AAC devices, this includes automatically generating topic-specific displays [34], generating text to support conversations in text-based AAC devices [88, 89], and tailoring AAC systems to reflect individual user narratives for more personalized and authentic communication in text-based AAC devices [75].

Many researchers, ourselves included, as well as users of assistive technology have adopted a critical lens of using AI in assistive technology [2, 38, 39, 64]. Many have argued that genAI promotes ableism and that outputs contain ableist biases and stereotypes [2, 38, 39, 64] and risks of misinformation and hallucinations in genAI responses [39]. GenAI should not be viewed as a panacea which can solve all issues with assistive technology. Rather, it is simply a tool which should be used with an abundance of caution. In our work, we do not intend for genAI to replace the role of speech-language pathologists or communication partners.

3 A Prototype for AI-Assisted VSD Configuration

The goal of this research is to understand how non-expert users interact with genAI-provided hotspot suggestions in the VSD configuration process and how their use of AI suggestions impacts their confidence and effectiveness of configuring VSDs as well as any potential quality impacts of AI suggestions. To investigate this,

we needed to design and develop a prototype. We developed a web application in ReactJS which enables users to upload an image, automatically generate AI suggested hotspots, and then review and modify them if needed (see Figure 1). The prototype uses the computer vision capabilities of a multimodal large language model (LLM) to automatically generate hotspots which aid in scaffolding the VSD configuration process. The web application runs on a tablet and mimics an interface that is otherwise very similar to other VSD applications that VSD programmers and AAC users typically see. At this stage, our prototype supports the configuration of VSDs and the ability to preview the final VSD which an AAC user would use. This prototype is not meant for use with VSD end users. In the following section, we outline the design goals which inspired the design of our prototype. Then, we describe the interfaces and key features of our prototype.

3.1 Design Goals

The design of our prototype was informed by best practices in VSD design. For example, VSDs should use hotspots which are engaging and personalized for the user [44, 62], developmentally appropriate [44], and provide enough communication options to support language and communication skill development without overstimulating the end user [41, 87]. Our work is also informed by evidence that if AAC devices are easier to configure and use, they are less likely to be abandoned [11, 69]. Our work is guided by these design goals:

Design Goal 1: Reduce the barriers to configuring VSDs. VSDs are typically used by expert practitioners and researchers in a clinical context [85]; to use VSDs outside of this relatively narrow context, scaffolding the process of mapping relevant and developmentally appropriate vocabulary [44] needs to happen. Our goal, informed by prior research, is to include more communication partners who are programming VSDs and contexts of use [61]. In order to support pre-service SLPs in VSD creation, we decided to use genAI in creating an initial set of hotspots to scaffold users' experiences when configuring VSDs.

Design Goal 2: Make it easy to integrate VSD users' linguistic abilities and interests. While genAI has shown promise in AAC applications, there are drawbacks to incorporating genAI systems, such as the prevalence of harmful biases and stereotypes [5, 39, 64, 93, 94] and hallucinations [39, 93]. With these drawbacks in mind, communication partners should be included as moderators in the hotspot generation process to ensure that hotspots would not be harmful for VSD end users and that they would be developmentally appropriate and sufficiently personalized to be the most effective [97]. Prior research on VSD effectiveness shows that hotspots should be engaging for the end user [44, 61, 62], and this is primarily achieved by including personally relevant hotspots about the people and activities being depicted in the image [44, 61, 62].

Design Goal 3: Minimize changes to the VSD users' interface. The VSDs used by emerging communicators should match the format of existing VSD applications as much as possible. This is important because learning to use AAC devices takes a lot of time and effort [6, 69, 78]. Two popular VSD applications, Tobii Dynavox's Snap Scene² and Attainment Company's GoVisual³, both

²<https://us.tobiidynavox.com/products/snap-scene>

³<https://www.attainmentcompany.com/govisual>

present VSDs with drawn hotspots encircling a person or object of interest, which when pressed, display a label with the hotspot name. We preserved this interaction pattern in our system.

3.2 Features

There are several key features of our prototype: automatic generation of hotspots, manually creating hotspots, modification of existing hotspots, and drawing hotspots on the selected image.

3.2.1 Create Hotspots Automatically. In the past couple of years, there have been huge advancements in the capabilities of LLMs and within the last year multimodal LLMs which can accept text, audio, and image input have become available with impressive vision capabilities. We automatically generate hotspots using OpenAI's GPT-4o model⁴. We chose this model because at the time of our prototype's development, it was the most cost-effective and available pre-trained model. When an image is uploaded, our system sends the unannotated image to GPT-4o via its public API to automate the generation of hotspots from static visual media. The multimodal LLM processes the visual input and text-based prompt to return a ranked list of regions of "hotspots"—areas of visual or contextual significance. We used the following prompt instructions for the language model:

"This photo was taken to be used in a visual scene display for a child. The child using the visual scene display uses mostly single words. Please provide the contextually relevant hotspots for the image if you are focused on building engagement in interactions and the emergence of words. Please focus primarily on the objects and activities being done in the scene. Include nouns, verbs, and descriptors when possible in words children would recognize."

Prior work by Zastudil et al. used a prompt which used information specific to the image provided to the language model and gave explicit instructions about the AAC user's goals when using the VSD [97]. Our prompt was informed by theirs, and we modified their prompt to improve generalizability so that any image could be used to configure a VSD. The modifications also better emphasize the importance of using appropriate vocabulary and producing contextually relevant hotspots which would be useful in a VSD.

We chose to use a prompt without any personalized information (e.g., linguistic skills, names, special interests) about the VSD end users because it requires a lot of information about the user, which has privacy considerations and is not easy to collect. Therefore, we leveraged current capabilities of the multimodal LLM to create contextualized suggestions about the people, objects, and activities being depicted in an image [97]. Personalization should come from the users configuring the VSDs. This reflects the unique knowledge SLPs have about their clients that they use when configuring AAC devices. Additionally, in the prompt, we did not give a limit to the number of hotspots, as our prototype is designed to scaffold VSD configuration for the users, providing a set of suggestions rather than a final set of hotspots.

3.2.2 Create Hotspots Manually. Users can **add new hotspots** which were not generated by our prototype. These hotspots can be

⁴<https://openai.com/index/gpt-4o-system-card>

whatever the user wants them to be. Supporting manual additions to the VSD provides users the flexibility to add hotspots beyond the automatically generated hotspots. The automatically generated hotspots may not include all of the hotspots related to the objects, activities, or people in the scene which the user may want to use.

3.2.3 Modify Existing Hotspots. Once the hotspots are generated, users can **edit hotspots** directly in the web interface by clicking on the hotspot or **delete hotspots** which they deem irrelevant or simply do not want to use. We expected that users would edit hotspots to change the content to be more developmentally appropriate, personalized, or any other edits they would like to make.

3.2.4 Drawing Hotspots. Once the user has decided which hotspots they want to use for the VSD, they are able to draw the hotspots on the selected image using their finger, a mouse, or a stylus.

3.3 Interfaces

There are three primary interfaces in our prototype. The image capturing and upload interface, VSD editing interface, and a VSD preview interface. Figure 1 shows what these interfaces look like. Our prototype was designed to be used on a tablet computer, such as an iPad or similar device.

3.3.1 Image Capturing and Upload Interface. When opening the application to configure a new VSD, the VSD programmer must first upload a new image. Users can upload an existing image or take a new image using the onboard camera. Currently, the uploaded image is not stored on the device or in the cloud.

3.3.2 VSD Editing Interface. Once the image is uploaded, the page for editing the VSD's hotspots is automatically opened. The editing interface consists of three windows: the list of generated hotspots, an editing window for hotspots, and the image for users to draw hotspots on. There are also options for starting over and creating a new VSD and viewing the finalized VSD an end user would see. Each generated hotspot can be deleted if the user does not wish to use it. When a user selects a hotspot from the generated hotspots window, the editing interface is populated with the hotspot, where the user can then edit the text content of the hotspot. Additionally, users can choose to add hotspots by pressing the add hotspot button. Lastly, when the user selects a hotspot, they can draw the hotspot on the image populated in the image window.

3.3.3 VSD Preview. Once the user has decided to finalize the VSD, they are shown a preview of what an end user of the VSD would see. This includes the image with the hotspots they drew on the previous screen. When the hotspots are pressed, a label containing the hotspot text is temporarily overlaid over the image and the hotspot content is played aloud. In the current stage of this prototype, we do not send this VSD anywhere, it is solely for display purposes.

4 User Study

To evaluate our prototype and investigate our research questions, we conducted a within-subjects study consisting of two conditions for configuring VSDs. In the control condition, participants configured two VSDs using Tobii Dynavox's Snap Scene. In the experimental condition, participants used our prototype to complete the same task. The conditions were counterbalanced to address

ordering effects, and four images were used to ensure reliability across stimuli.

Snap Scene was selected as a control condition because it is a widely available commercial VSD application. Snap Scene is available on more platforms than iOS, unlike similar VSD applications (e.g., GoVisual, Scene Speak⁵), and it only costs \$49.99. In addition to being widely used and available, it was also designed based on best practices and empirical research within the field which found it to be effective at facilitating social interactions, sharing information and expressing needs, integrating new words and concepts, and combining words for more complex ideas [45].

4.1 Participants

VSDs have a lot of potential benefits for AAC users, especially for beginning communicators [59, 60]. However, the lack of training and knowledge of effective configuration of VSDs hinders their more widespread adoption [53]. While experts in VSDs have the knowledge for configuring them, untrained communication partners may need additional scaffolding in configuring effective VSDs. We recruited seventeen pre-service speech-language pathologists (SLPs) from a large R1 university in the United States. We recruited participants using convenience sampling via an email sent to students studying communication sciences and disorders. Our inclusion criteria is that they must be studying to become a clinically certified SLP (CCC-SLP) and that they have some clinical experience. Based on responses to demographic survey questions, only two participants reported frequently using or observing VSDs being used in a clinical setting. Therefore, participants could be considered mostly unfamiliar with VSDs. Additionally, none of our participants had used the control software before our study. See Table 1 for full participant information.

4.2 Procedure

The second author of this work, a clinically certified SLP and expert on VSDs, ran each user study in-person during September 2024. It consisted of three phases (1) information about VSDs, (2) VSD configuration with Snap Scene and our prototype, and (3) an online post-questionnaire. We received ethics approval to run this study by our institutional review board (IRB) before running the study. Participants signed a consent form before participating in our study. Participation was voluntary. After completion of the study, participants received information related to programming AAC for emerging symbolic communicators that could be useful to them in their future careers.

4.2.1 Information about VSDs. Before beginning the configuration portion of the user study, since our participants were largely unfamiliar with VSDs, we provided participants with a broad overview of VSDs, including how they are typically used, how images are selected for VSDs, and what functions hotspots perform. This included the second author providing verbal information along with some accompanying images of three example VSDs to demonstrate how they are used. After we provided them with this information, they were given the opportunity to ask any questions they had about VSDs to the researcher running the user study.

⁵<https://www.goodkarmaapplications.com/scene-speak1.html>

4.2.2 VSD Configuration. Once the participants had information about VSDs, we had participants configure VSDs with the control application (Snap Scene) and our prototype. The order of which application was used first and which images were used was counterbalanced across the entire experiment to mitigate any order effect. For each application, we showed participants a brief video tutorial to illustrate how each application worked. Participants configured VSDs for two contexts: playing and retelling a past activity. These contexts were selected because they are common use cases for VSDs [19, 55]. For each application, we provided the participant with two images (one for each context), each with a case study to inform their VSD creation. Case studies are commonly used in AAC research to describe AAC users' language capabilities, needs, and goals [45]. In order to ensure that VSD configuration was as naturalistic as possible to reflect how SLPs currently use VSDs in educational or therapeutic contexts, and how that behavior might change if they are assisted by AI, we did not inform participants about any assessment criteria (see Section 4.4) we used in our analysis to avoid biasing their use of either system. An example image and case study provided to participants is shown in Figure 2. All images and case studies we used are provided in Appendix A.

4.2.3 Post-Questionnaire. Once participants finished configuring their VSDs, we had them complete a post-questionnaire. Participants were asked in the questionnaire to rate their confidence in the VSDs they made with both Snap Scene and our prototype using a Likert-scale from "Not at all Confident" to "Extremely Confident". We then asked participants questions about the relevance and appropriateness of the generated vocabulary based off of questionnaires used by de Vargas et al. [33, 34]. The full post-questionnaire is provided in Appendix B.

4.3 Data Collection

We collected screen recordings of the VSD configuration process. We used these recordings to obtain information about how long VSD creation took and which hotspots were created. Two members of the research team reviewed each of the videos carefully capturing timestamps for the starting and ending times for each VSD and what hotspots were created. For the VSDs configured during the prototype condition, we also recorded which generated hotspots were used, edited, or deleted. Each video was reviewed by two members of the research team to ensure there were no mistakes in data collection. Additionally, we used the responses to the post-questionnaire to gain insights on participants' confidence when making their VSDs and their perceptions on the quality of the generated hotspots.

4.4 Data Analysis

To analyze the results of our user study we conducted multiple levels of analysis, both at the hotspot and VSD level. Our hotspot-level analysis consisted of analyzing the content of the hotspots using part-of-speech analysis [97] and deductive coding using Light's functions of communication framework [58, 97]. We also analyzed what automatically generated hotspots participants used, modified, deleted, and added when using our prototype. At the VSD-level we analyzed the quality of the configured VSDs based off best practices

	Program and Year	Clinical Experience	Client Ages	Emerging Communicators	VSD Experience
P1	2nd, GR	School	Young children, school-age children, transition-age youth, adults	Occasionally	Never
P2	2nd, GR	School, private therapy	Young children, school-age children, transition-age youth	Frequently	Occasionally
P3	2nd, GR	Hospital	Adults	Very rarely	Never
P4	2nd, GR	Hospital, private therapy	School-age children, transition-age youth, adults	Very rarely	Rarely
P5	2nd, GR	School, private therapy	Young children, school-age children	Occasionally	Occasionally
P6	2nd, GR	School, private therapy	Young children, school-age children	Very frequently	Frequently
P7	2nd, GR	School, private therapy	Young children, school-age children, transition-age youth	Frequently	Occasionally
P8	4th, UG	School, private therapy	Young children, school-age children	Occasionally	Occasionally
P9	3rd, UG	School	Young children	Never	Very rarely
P10	3rd, UG	School	Young children	Occasionally	Never
P11	4th, UG	School, private therapy	Young children, school-age children, transition-age youth, adults	Frequently	Frequently
P12	3rd, UG	Summer camp	Young children	Occasionally	Never
P13	1st, GR	Private therapy	Young children, school-age children	Occasionally	Occasionally
P14	1st, GR	Hospital	Adults	Very rarely	Never
P15	1st, GR	Hospital	Adults	Never	Never
P16	2nd, GR	School	School-age children	Frequently	Occasionally
P17	2nd, GR	School, private therapy	Young children, school-age children, transition-age youth, adults	Frequently	Rarely

Table 1: An overview of the demographic information for the 17 pre-service SLPs that participated in our study. We collected data about the number of years they had been in their program, the settings of their clinical experience, the age ranges of their clients, whether they work with emerging communicators, and whether they have either used or observed the use of a VSD. For the program year, we report which year they are in the program and whether they are undergraduate students (UG) or graduate students (GR).

for VSD creation informed by prior research. We also conducted a homogenization analysis of the configured VSDs.

4.4.1 Hotspot Analysis. We analyzed the content of the hotspots across both conditions to understand how AI suggestions might have impacted the VSDs participants configured. This included part-of-speech analysis and deductive coding to determine the primary functions of the hotspots. The part-of-speech analysis and deductive coding allowed us to compare the hotspots to understand if and how hotspot content differed by condition. Additionally, we analyzed the interaction logs for the prototype to understand which hotspots were added, edited, and deleted from the VSDs. We did not conduct analysis of the generated hotspots as this has been done in prior work [97]. Our analysis was focused on understanding how

participants used the generated suggestions in combination with the knowledge they had about the hypothetical VSD end users they were configuring based on the case studies provided.

For the part-of-speech analysis we followed the procedure outlined by Zastudil et al. [97]. Each hotspot was split into single words to ensure that every part-of-speech was accounted for. The first author then classified each word by its part-of-speech. We conducted this analysis to determine how the content and structure of the hotspots aligned across the Snap Scene and prototype VSDs.

We conducted the same deductive coding process outlined by Zastudil et al. [97] for the hotspots made with Snap Scene and with our prototype. These deductive codes are from Light's Functions of Communication framework [58]:



Figure 2: An example VSD image and case study prompt participants used during VSD creation. We provided information related to the hypothetical VSD user's linguistic abilities, communication goals, and personal details (e.g., special interests).

- (1) **Expressing Wants or Needs** - communication intended to make requests
- (2) **Information Transfer** - communication meant to share information with others
- (3) **Social Closeness** - communication meant to develop or maintain relationships
- (4) **Social Etiquette** - communication meant to convey polite terms (e.g., "thank you", "please", "hello")

We also included an "Other" category to handle communication options which did not clearly align with these four functions. Two researchers performed the coding and inter-rater reliability was computed in accordance with best practices for qualitative research [68]. We computed the inter-rater reliability using Cohen's Kappa since we had two raters and categorical codes [20]. The inter-rater reliability score was 0.66 indicating substantial agreement [54].

4.4.2 VSD Analysis. Evaluating the quality of hotspots within a VSD is challenging because it is a highly subjective task. There is not an agreed upon set of guidelines for creating VSDs; however, there are recommendations for VSD configuration informed by prior research. We analyzed the quality of the VSDs configured by participants according to the guidelines described below:

- (1) Use 2–4 hotspots to ensure that there isn't too much visual stimulation or overlapping hotspots which may overwhelm users [41, 87] or result in stimulus over-selectivity [29, 95] (i.e., hyper-attentiveness to some stimuli and limited to no attention paid to other relevant stimuli [15]).
- (2) Focus hotspots on the people, activities, and social interactions (when applicable) in the scene [44, 96].
- (3) Use hotspots which align with the users' communication stage and learning goals [44, 61]. This includes using single words in hotspots over phrases which can be combined together to create more complex phrases [44, 61].

Additionally, we analyzed the VSDs configured to see if the genAI-assistance resulted in more homogeneous VSDs by using the homogenization analysis procedure developed by Anderson et al. [4]. We computed the sentence embeddings [79] for all of

the sets of hotspots made across conditions using the Python SentenceTransformers library [79] with the model all-MiniLM-L6-v2. We then calculated the average embedding for all hotspots across conditions. Then, we compared each set of hotspots to the average embedding using cosine similarity to understand how distinct each set of hotspots was from the group.

5 Results

We collected a total of 64 VSDs (32 from Snap Scene and 32 from our prototype) with 247 hotspots (94 from Snap Scene and 153 from our prototype). A total of 178 hotspots were generated by our prototype, and 118 of those were used either unmodified or they were edited. One participant's data (P5) was removed from the study due to a application error with the control software. All of the VSDs participants configured for both conditions are provided in our supplemental material.

5.1 RQ1: Effectiveness of Generative AI-Assisted VSD Creation

We found that participants configured VSDs more efficiently in terms of time and that they felt more confident about the VSDs they configured with our prototype than the VSDs they configured with Snap Scene. Overall, we saw that participants used more hotspots when creating VSDs with our prototype than with Snap Scene. The results of our quality analysis were mixed, with the VSDs configured with Snap Scene adhering to two out of three best practices, whereas our the VSDs configured with our prototype only adhered to one out of three.

5.1.1 Time-to-Create VSDs Across Conditions. Participants' median time to configure a VSD (i.e., time-to-create) VSDs when using Snap Scene and our prototype was 77 seconds and 63.5 seconds respectively. A Wilcoxon Signed-rank test shows that there is a statistically significant effect on the prototype on participants' time-to-create when creating VSDs ($W = 438.5, Z = 3.2635, p < 0.05$) with a medium effect size ($r = 0.41$). See Figure 3 for a visual comparison of participants' time-to-create across conditions.

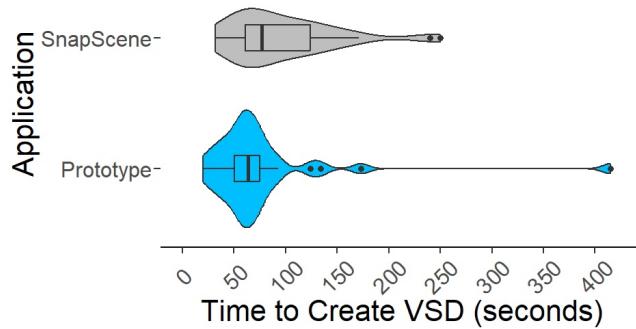


Figure 3: On average, participants configured VSDs faster when using our prototype compared to when they used Snap Scene.

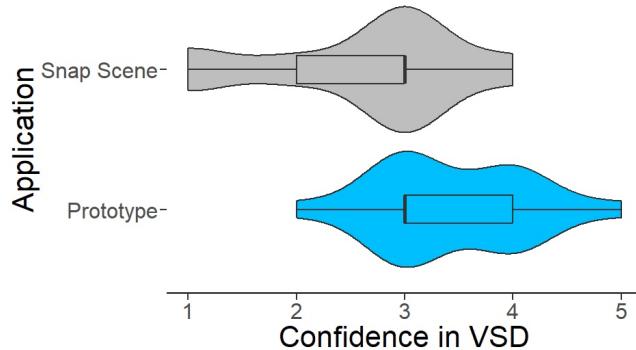


Figure 4: Participants' reported their confidence levels (on a 5-point Likert scale). We found that their confidence levels were higher for the VSDs they configured using our prototype compared to their reported confidence when configuring VSDs with Snap Scene.

5.1.2 Confidence in Configured VSDs. Participants ranked their confidence in the VSDs they configured along a 5-point Likert scale. A Wilcoxon Signed-rank test shows that there is a significant effect of the prototype on users' confidence when creating visual scene displays ($W = 5$, $Z = -2.8043$, $p < 0.05$) with a medium effect size ($r = 0.496$). Figure 4 provides a visual comparison of participants' confidence across conditions.

5.1.3 Number of Hotspots Made Across Conditions. Participants used, on average, 2 more hotspots per VSD when using our prototype ($n=153$) versus using Snap Scene ($n=94$). The median number of hotspots when using Snap Scene and our prototype were 3 and 5, respectively. A Wilcoxon Signed-rank test shows that there is a significant effect of the prototype on users' number of hotspots used when creating visual scene displays ($W = 5.5$, $Z = -4.5567$, $p < 0.05$) with a large effect size ($r = 0.57$). See Figure 5 for a visual comparison of participants' number of hotspots used per VSD.

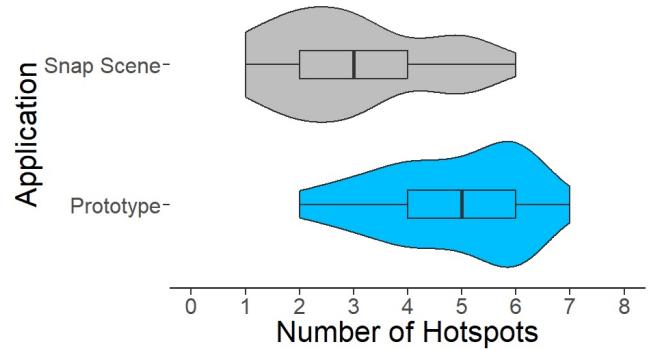


Figure 5: Participants tended to use more hotspots when programming VSDs with our prototype than when using Snap Scene.

5.1.4 Quality of VSDs Configured. Overall, we see that findings for quality were mixed. Our quality measures are outlined in Section 4.4.2. For quality measures 1 and 2, Snap Scene outperformed our prototype, but for quality measure 3 the prototype outperformed Snap Scene. Quality measure 1 is focused on the number of hotspots used in VSDs. In the Snap Scene condition, 60.50% of VSDs used between 2 and 4 hotspots. In the prototype condition, we saw that 40.63% of VSDs used between 2 and 4 hotspots. Quality measure 2 is focused on the relevance of the hotspots used to the scene. In the Snap Scene condition, we saw that 95.74% of hotspots were focused on the people, activities, or social interactions, whereas in the prototype condition we saw that 84.97% of hotspots were focused on the people, activities, or social interactions. Quality measure 3 is focused on the developmental appropriateness of the hotspots used. In the Snap Scene condition participants used single-word hotspots 78.70% of the time, whereas in the prototype condition, participants used single-word hotspots 91.50% of the time.

5.2 RQ2: Relevance and Appropriateness of Generated Hotspots

Participants felt that the generated hotspots were relevant and appropriate for the targeted goals outlined for their VSD creation in the case studies we provided them. But there were trends in the modifications made, including removing irrelevant hotspots and editing hotspots to be more personalized or developmentally appropriate.

5.2.1 Participants' Perceptions of Relevance and Appropriateness. Through our post-questionnaire (see Appendix B), we found that participants found that our prototype generated hotspots which they wanted to use, and did not frequently generate hotspots which they did not want to use. Additionally, our participants found that the generated hotspots were effective in supporting them in creating VSDs with hotspots they would use in educational or speech therapy and achieving the goals outlined in the case studies provided to them in the study (see Figure 6).

5.2.2 Modifications Required for Prototype VSDs. Our analysis of the generated hotspots aligned with our participants' responses.

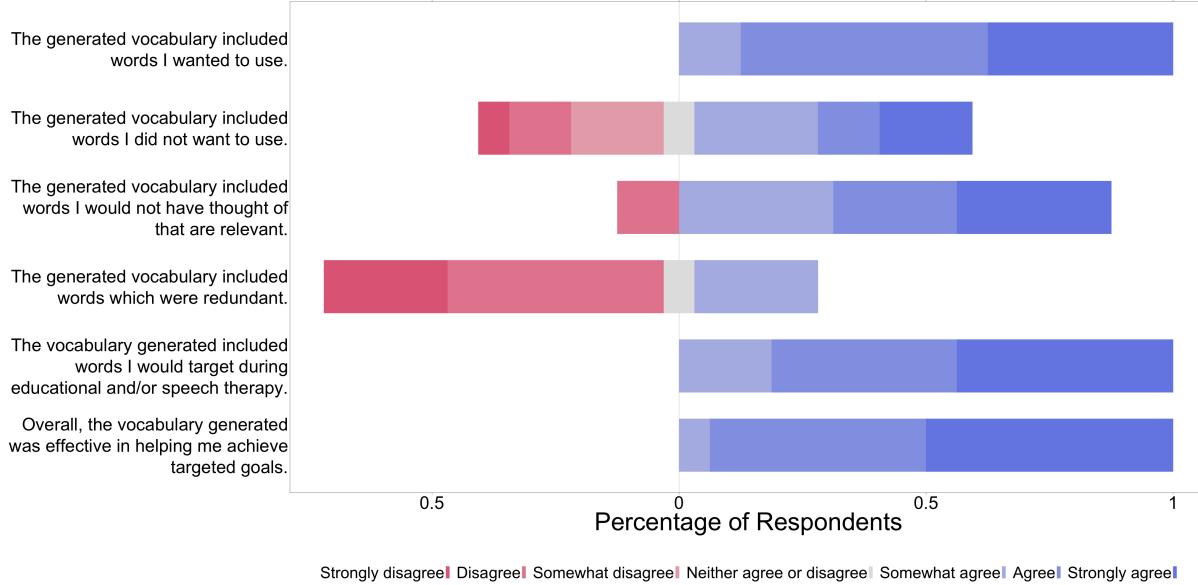


Figure 6: Participants reported that the automatically generated hotspots were relevant a majority of the time, and felt that they would be useful in achieving targeted goals in speech therapy.

We found that, of the hotspots generated by our prototype ($N = 178$), 64.6% of them were directly related to the scene, either about the people in the scene or the activity being depicted. However, we observed a high frequency of participants modifying generated hotspots to further personalize them to incorporate details provided to them in the case study (e.g., changing “boy” to the name provided in the case study). The other trend we observed was participants modifying hotspots to make them more developmentally appropriate. Of the modifications made by participants ($n = 8$), 37.5% made the hotspots more personalized and 37.5% made them more developmentally appropriate. We also found that of the hotspots participants deleted ($n = 57$), 59.6% were not relevant to the scene. Lastly, of the hotspots participants’ added ($N = 35$), 79.41% of them added personalized hotspots to the VSD.

5.3 RQ3: Impact of Generated Hotspots

VSDs configured with our prototype tended to be more homogeneous and participants tended to use more unique hotspots when using Snap Scene compared to our prototype. Additionally, participants seemed to rely heavily on the suggestions generated for them. We also observed some differences in the content across conditions. The parts-of-speech widely aligned across conditions, however, we saw a decrease in the use of hotspots for social interactions.

5.3.1 Reliance on Generated Hotspots. Of all of the hotspots generated by our prototype ($N=178$), 33.71% were deleted, 4.49% were edits, and 61.80% were used unmodified. Generated hotspots composed 77.12% of all hotspots ($N = 153$) used by participants when creating VSDs using our prototype.

5.3.2 Homogeneity of VSDs Configured Across Conditions. When participants configured VSDs with our prototype, the hotspots (i.e.,

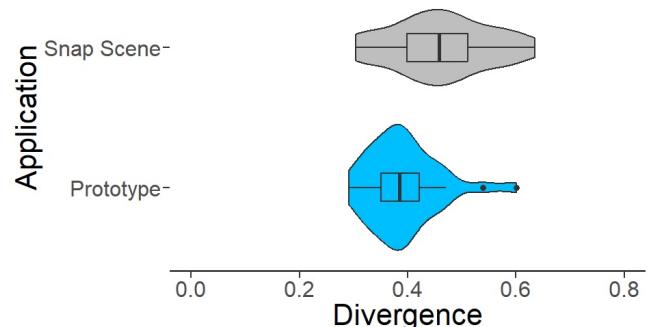


Figure 7: We found that VSDs that were configured with our prototype were more homogeneous than VSDs made with Snap Scene. The hotspots created with Snap Scene tended to diverge more from the average VSD based on semantic difference.

communication options) were more similar to the average embedding of all hotspots created across both conditions using cosine similarity to the average embedding ($\bar{x} = 0.39 \pm 0.09$) compared to VSDs configured using Snap Scene ($\bar{x} = 0.46 \pm 0.07$). We found a statistically significant difference for the application used on the homogeneity of the hotspots created ($t(31) = 3.4037, p < 0.05$), with a medium effect size ($d = 0.60$), 95% CI[0.26, 1.02], with higher homogeneity for VSDs configured using our prototype (see Figure 7).

5.3.3 Unique Hotspots Used Across Conditions. Across both conditions, 84 out of the total 247 hotspots were unique (34.01%). In the

Snap Scene condition 51 of 94 hotspots were unique (54.26%). In the prototype condition 56 of 153 hotspots were unique (36.60%).

5.3.4 Hotspot Content Across Conditions. Through our part-of-speech analysis we found that the content of the hotspots was largely similar across conditions. The hotspots were primarily composed of nouns and verbs in similar proportions in the Snap Scene and prototype conditions. The primary difference between the Snap Scene and prototype conditions was the inclusion of more advanced parts of speech (e.g., adjectives, pronouns, prepositions, adverbs, and articles) in the Snap Scene hotspots. See Table 2 for full description of the parts of speech used across conditions.

Through our analysis of the function of the hotspots created across conditions, we found that, in both conditions, the hotspots participants used heavily focused on information transfer. In the Snap Scene condition, however, participants used more hotspots focused on social closeness than in the prototype condition. See Table 3 for a full description of the functions of the hotspots used across conditions.

6 Discussion

We conducted a within-subjects user study with pre-service SLPs to better understand how our genAI-enabled VSD application prototype, which automatically generates hotspots using a pre-trained multimodal LLM, affected users' efficiency and confidence in configuring VSDs. We also conducted analysis about how our prototype affected three quality metrics, relating to hotspot content, number of hotspots, and developmental appropriateness of hotspots. Our analysis also included a comparison of the content of hotspots across conditions and how much users relied on the automatically generated hotspots. In the following section, we discuss these results and the broader implications for our work.

6.1 Support for Pre-Service SLPs Users in VSD Configuration

SLPs possess a wealth of knowledge about linguistic and communication skill development, incorporating AAC interventions to aid in developing these skills, and personal information about their clients to use these interventions effectively. There are a lot of different kinds of interventions, and developing knowledge about their effective use and application may face barriers such as lack of access to experts [24], training opportunities [9, 32], and educational opportunities [40, 86]. We found that when participants, who were largely unfamiliar with VSDs prior to participating in our study, configured VSDs using our prototype they were faster and more confident in their VSDs. However, our findings regarding quality of the VSDs were mixed. The prototype performed better in one quality measure, specifically developmental appropriateness, but worse along others, specifically the number of hotspots used and the relevance of the hotspots to the scene. This indicates that while the use of a VSD creation application with automatically suggested hotspots is useful in increases users' confidence and efficiency, it is not a perfect solution.

6.1.1 Pre-service SLPs configure VSDs more quickly. One of the main advantages of our prototype was its speed. Compared to Snap

Scene, a current state-of-the-art tool, our prototype allowed participants to configure VSDs more quickly. This builds on prior work by Caron, Light, and Drager, who found that JIT programming of VSDs was more effective when there were fewer steps required [18]. They demonstrated this with a prototype, EasyVSD, which is what Snap Scene was based on. Our prototype further reduces steps by automatically producing hotspot suggestions. With our prototype, SLPs can quickly evaluate these suggestions and add to them, modify, or delete them from the scene. By speeding up the process, SLPs and communication partners can quickly configure VSDs tailored to the immediate context which can improve rate of interaction, AAC user adoption, and increase the vocabulary with which AAC users interact [18, 61]. Building upon work from Fontana de Vargas et al. [34], participants reported that the generated hotspots were largely relevant to the scene and useful in achieving the goals outlined in the case studies we provided them. Participants incorporated some or all of the automatically generated hotspots in most of the VSDs they configured (31/32 of VSDs configured with our prototype). Beyond our participants' perceptions; however, we observed a high incidence rate of the automatically generated hotspots being about objects in the background of the images or not relevant to the people or activities being depicted in the images.

6.1.2 Pre-service SLPs are more confident in the VSDs they configure. Participants also reported feeling significantly more confident about the VSDs they configured with our prototype compared to Snap Scene. This is important because when a user feels more confident in their ability to use an application, they are more likely to continue using that application. This phenomenon has been documented consistently across application domains, such as assistive technology [8, 100], educational technology [63, 83], and older adults' technology adoption [3, 72]. A lack of confidence has been linked to higher rates of abandonment of AAC devices with AAC users and their communication partners, often parents [69]. Our findings indicate that a system such as ours, may be beneficial in helping communication partners feel more confident using VSDs. Providing initial suggestions of hotspots in our prototype to scaffold VSD configuration may have impacted participants' perceptions of the difficulty of the task and, in turn, made them feel more confident in configuring the VSDs. This has been observed in other contexts where JIT scaffolding has increased users' confidence [7, 43, 65]. Participants did not just feel more confident in their VSDs, they felt that they would be able to achieve the targeted goals of the case studies we presented to them. This feeling of self-efficacy in VSD use is promising, because, while the general feeling of self-efficacy and competence in AAC use of SLPs has increased in recent years [53], SLPs confidence in choosing the appropriate AAC devices for clients (i.e., feature matching) is not as high [12]. Increasing SLPs confidence and general knowledge in VSDs may be a step towards increasing self-efficacy in feature matching.

6.1.3 Pre-service SLPs used developmentally appropriate hotspots. Participants created more single-word hotspots when using the prototype (91.5%) than when using Snap Scene (78.7%), which aligns with recommendations for VSD creation [44, 61]. We also saw that when participants modified automatically generated hotspots ($n = 8$), they often changed hotspots to be more developmentally appropriate (37.5% of the time). By using more single-word hotspots

Part of Speech	Snap Scene	Prototype
Nouns	64.0%	74.1%
Verbs	15.2%	15.7%
Adjectives	8.0%	5.4%
Pronouns	6.4%	4.2%
Prepositions	2.4%	0.0%
Adverbs	0.8%	0.0%
Articles	1.6%	0.0%
Other	1.6%	0.6%

Table 2: Part of speech frequency for the Snap Scene and prototype conditions. The use of nouns, verbs, adjectives, and pronouns was similar across conditions. However, Snap Scene VSDs had more prepositions, adverbs, and articles.

in VSDs, users are able to more flexibly combine a larger variety of words to communicate increasingly complex ideas, which is supportive of language development and acquisition [76].

6.2 Negative Impacts of AI-Suggestions

In the previous section, we outlined some of the positive effects of the automatic hotspot suggestions: faster configuration of VSDs, increased confidence, and using developmentally appropriate hotspots. However, these benefits come at a cost.

6.2.1 Configured VSDs were less aligned with best practices. Despite these positive impacts, we also observed some more negative effects of genAI suggestions. In terms of quality impacts, we saw that participants tended to use significantly more hotspots when using our prototype and use hotspots which were not directly related to the scene. This conflicts with best practices from AAC research by potentially overwhelming the visual-cognitive processing capabilities of AAC users. Having too many hotspots can negatively impact linguistic development due to the likelihood for stimulus over-selectivity and overwhelming end users [29, 41, 95]. By using hotspots which are not clearly referring to the people or activities within the scene, it circumvents one of the unique attributes of VSDs which makes them effective for language acquisition, incorporating a familiar context to anchor vocabulary [45]. These results reflect the complexity of measuring quality in this context, but also raises concerns about just replacing VSD programming software with AI-enabled software, as it may lead to overwhelming visual cognitive processing. Lastly, we saw that when participants configured VSDs using our prototype, in almost every case (98.7% of the time) the hotspots were meant to convey information and rarely develop social closeness (0.7% of the time). When using Snap Scene, participants used proportionally more hotspots intended to develop social closeness (5.3% of the time). VSDs have been shown to be effective in supporting social engagement between the VSD users and their communication partners, specifically their peers [55, 56], so it is an important function to retain in VSDs. While this might be addressed by further refining the prompts for the genAI models, there currently exist misalignments between model behavior and best practices which need to be mitigated before focusing on aspects like prompt engineering. There may be interaction methods informed by prior HCI research which can be implemented to apply

Function	Snap Scene	Prototype
Information Transfer	93.6%	98.7%
Social Closeness	5.3%	0.7%
Expressing Wants & Needs	0.0%	0.0%
Social Etiquette	0.0%	0.0%
Other	1.1%	0.7%

Table 3: The functions, as described by Light [58], of the hotspots used across conditions. We found that, across both conditions, participants' hotspots tended to focus on information transfer, however, there was substantially less focus on social closeness in the hotspots used when configuring VSDs with our prototype.

checks to ensure appropriateness and quality when configuring VSDs. One example of such an intervention could be to include prompting the users to reflect [42] upon their created hotspots based on best practices and whether or not they have included personalized hotspots and hotspots for social closeness.

6.2.2 Pre-service SLPs relied heavily on AI suggestions. Through our analysis of how many of the generated hotspots were edited or deleted or when new hotspots were manually added, we saw that participants exhibited signs of over-reliance on the generated suggestions. Participants often accepted the hotspots suggested without any modifications (61.80% of generated hotspots). Generated hotspots, included those which were edited made up 77.12% of all of the hotspots created when using our prototype. Only 22.88% of the total hotspots made using our prototype were manually added by our participants. This could be related to participants having a false sense of confidence in the genAI to generate high-quality suggestions [52]. Although our intention was for participants' to use the generated hotspots as a starting point for their VSD creation, it is not unexpected that they over-relied on the automated hotspots. This commonly occurs in genAI-powered systems used by non-experts for complex tasks [17, 77]. Over-reliance on AI is a well-studied phenomenon, and there are multiple approaches which have been validated to combat over-reliance. A potential solution could be to implement a validation step which checks a user's configured VSD against best practices and explains [91] what changes should be made.

6.2.3 Communication options were more homogenous. The last and potentially most impactful negative impact we saw observed was the increased homogeneity and reduction of unique hotspots of VSDs configured with our prototype. Increased homogeneity when using genAI-powered systems has been observed in the context of using genAI as a creativity support tool [4]. This homogenizing effect has been observed in other contexts as well, including writing [1, 10], crowdworker output [92], and people's opinions and beliefs [30, 50, 93]. As previously mentioned, the reason why VSDs are engaging and effective for users is that they are personally relevant, and just-in-time programming has enabled communication partners to specifically tailor the VSDs for users [19, 28, 45], sometimes with input from the VSD users themselves [46]. Evidence of homogenization of VSDs is potentially harmful, as homogenized

VSDs represent a narrower range of options that can be communicated. It is important to offer a diverse range of options to support broader expression via VSDs. The goal is to expand communication, and it is problematic if integrating AI into VSDs inadvertently limits it. The inadvertent homogenization effect of genAI has led to many researchers to call for solutions to mitigate this homogenization [1, 4], and more research needs to be done to understand how to apply these potential solutions to AI-enabled AAC devices.

6.3 Addressing the Risks of AI in AAC Devices

In the previous sections we discussed some of the positive and negative findings of using genAI in our VSD configuration prototype. While there were some positive impacts of using our prototype - increased efficiency, confidence, and the use of developmentally appropriate hotspots - there are important negative impacts, such as over-reliance and homogenization which are known in design and education contexts, but have been overlooked in previous studies of AI-enabled AAC devices [34, 89].

Because of these newly discovered issues, this prototype could be potentially useful in an educational context for scaffolding VSD creation rather than requiring pre-service SLPs to start from scratch. Training is a really important part of self-efficacy and the effective use of AAC devices [21, 32]. Introducing this as a training tool could make it easier for pre-service SLPs to develop confidence and familiarity in using VSDs while helping professionals and experts to train and correct pre-service SLPs with some of the best practices we have discussed. But the same problems remain for in many AI systems: over-reliance [35, 49, 98], homogenization [1, 10, 30, 50, 93], and a lack of critical engagement [73, 77, 90].

Researchers have previously advocated for the integration of AI into AAC by presenting positive empirical outcomes [34, 89]. However, our findings offer a more nuanced perspective that reiterates some of the positive aspects while also identifying potential pitfalls. We identified a potential risk of over-reliance on AI suggestions where participants passively accepted suggestions without modifying them. We also identified tradeoffs between efficiency and quality. Participants were more confident when using AI suggestions in our study despite mixed findings regarding the quality of VSDs generated with AI support. Trusting that the suggestions are good when they are not, in the case of AAC, can significantly harm users' linguistic and social development. By providing communication options that are not related to scene or by overwhelming users with too many communication options, it reduces the usefulness of AAC devices [11], which can also risk the abandonment of AAC devices entirely [11].

Finally, we observed that the hotspots created with AI assistance tended to be more homogeneous. Given that AAC devices are most effective when carefully personalized to the user, this homogenization is problematic because it essentially represents sensible defaults. The homogenizing effect of genAI that we observed presents a disturbing trend - the creation of VSDs and potentially other AAC displays which are becoming more and more similar. The benefit of JIT programming of AAC displays is that they can be incredibly personalized to each user [44]. Using JIT programming to configure less unique AAC displays results in a display that is more similar to a default display, which are not as useful to anyone.

7 Limitations and Future Work

While the core functionality (i.e., taking or uploading an image, manual creation and editing of hotspots) of the two application used in our study was the same, there are minor differences (e.g., icons, fonts) which may impact the usability of our prototype as compared to Snap Scene. Additionally, our user study only included one group of communication partners, pre-service SLPs. Future work could also include other communication partners such as parents, caregivers, and friends. We hypothesize that the trends of over-reliance and homogenization would be more apparent with these communication partners, as pre-service SLPs contain specific knowledge about how language and communication skills are developed.

Our prototype represents a lower bound for how AI might be integrated into AAC, specifically VSDs. With more carefully crafted guardrails, some of the issues we identified in this study (e.g., over-reliance, homogeneity) might be mitigated. For example, through more advanced prompting techniques (e.g., self-critique [67]), homogeneity could be reduced; however, given the propensity for over-reliance, this approach gives more agency to the AI. Future work is necessary to refine our prototype and add in interactive features that would better support the configuration of high-quality VSDs.

8 Conclusion

VSDs have been shown to be effective for many communicators who use AAC devices; however, there are many barriers which prevent their widespread adoption outside of clinical contexts. The primary barrier is that they are difficult to configure for communication partners who are not experts in VSDs and JIT configuration. We developed a prototype VSD which uses genAI to scaffold the configuration process. We evaluated this prototype with 16 pre-service SLPs with mixed results. We found that while our prototype increases users' confidence and self-efficacy, we see substantial negative impacts related to over-reliance on AI suggestions, divergence from best practices in VSD creation, and homogenization of VSDs configured with the support of AI.

Acknowledgments

This work was partially funded by the Temple University Pervasive Computing REU Site and Convergence Accelerator Grant (National Science Foundation grant numbers CNS-2150152 and ITR-2236352).

References

- [1] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. Ai suggestions homogenize writing toward western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [2] Bedour Alshaigy and Virginia Grande. 2024. Forgotten Again: Addressing Accessibility Challenges of Generative AI Tools for People with Disabilities. In *Adjunct Proceedings of the 2024 Nordic Conference on Human-Computer Interaction*. 1–6.
- [3] Lan An, Diego Muñoz, Sonja Pedell, and Leon Sterling. 2022. Understanding confidence of older adults for embracing mobile technologies. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction*. 38–50.
- [4] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th conference on creativity & cognition*. 413–425.
- [5] Lena Armstrong, Abbey Liu, Stephen MacNeil, and Danaë Metaxa. 2024. The Silicon Ceiling: Auditing GPT's Race and Gender Biases in Hiring. In *Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and*

Optimization (San Luis Potosi, Mexico) (EAAMO '24). Association for Computing Machinery, New York, NY, USA, Article 2, 18 pages. <https://doi.org/10.1145/3689904.3694699>

[6] Allyson P. Arserio, Elizabeth E. Biggs, and Emily Holz. 2023. Experiences of Parents Supporting Their Young Children Who Use Speech-Generating Devices. *Journal of Early Intervention* 0, 0 (2023), 10538151231199898. <https://doi.org/10.1177/10538151231199898> arXiv:<https://doi.org/10.1177/10538151231199898>

[7] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A. Hearst, Andrew Head, and Kyle Lo. 2023. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 74 (Sept. 2023), 38 pages. <https://doi.org/10.1145/3589955>

[8] Maryam Bandukda, Catherine Holloway, Aneesha Singh, Giulia Barbareschi, and Nadia Berthouze. 2021. Opportunities for supporting self-efficacy through orientation & mobility training technologies for blind and partially sighted people. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[9] Brooke E Barman, Virginia L Dubasik, Tim Brackenbury, and Donna J Colcord. 2023. Graduate students' perceived preparedness and confidence to work with individuals who use augmentative and alternative communication. *American Journal of Speech-Language Pathology* 32, 3 (2023), 1165–1181.

[10] Nicolas F Bauer. 2025. Does ChatGPT Increase Language Homogenization? *KI in Medien, Kommunikation und Marketing: Wirtschaftliche, gesellschaftliche und rechtliche Perspektiven* (2025), 11.

[11] Susan Baxter, Pam Enderby, Philippa Evans, and Simon Judge. 2012. Barriers and facilitators to the use of high-technology augmentative and alternative communication devices: a systematic review and qualitative synthesis. *International Journal of Language & Communication Disorders* 47, 2 (2012), 115–129.

[12] David Beukelman and Janice Light. 2020. Augmentative and alternative communication: Supporting children and adults with complex communication needs. (2020).

[13] David R Beukelman, Karen Hux, Aimee Dietz, Michelle McKelvey, and Kristy Weissling. 2015. Using visual scene displays as communication support options for people with chronic, severe aphasia: A summary of AAC research and future research directions. *Augmentative and Alternative Communication* 31, 3 (2015), 234–245.

[14] Sarah Blackstone, J Light, D Beukelman, and H Shane. 2004. Visual scene displays. *Augmentative Communication News* 16, 2 (2004), 1–16.

[15] Jessica Bradshaw. 2013. *Stimulus Overselectivity*. Springer New York, New York, NY, 3007–3008. https://doi.org/10.1007/978-1-4419-1698-3_137

[16] Kris Brock, Rajinder Koul, Melinda Corwin, and Ralf Schlosser. 2017. A comparison of visual scene and grid displays for people with chronic aphasia: A pilot study to improve communication using AAC. *Aphasiology* 31, 11 (2017), 1282–1306.

[17] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (April 2021), 21 pages. <https://doi.org/10.1145/3449287>

[18] Jessica Caron, Janice Light, and Kathryn Drager. 2016. Operational demands of AAC mobile technology applications on programming vocabulary and engagement during professional and child interactions. *Augmentative and Alternative Communication* 32, 1 (2016), 12–24.

[19] Shelley E. Chapin, David McNaughton, Janice Light, Ashley McCoy, Jessica Caron, and David L. Lee. 2022. The effects of AAC video visual scene display technology on the communicative turns of preschoolers with autism spectrum disorder. *Assistive Technology* 34, 5 (Sept. 2022), 577–587. <https://doi.org/10.1080/10400435.2021.1893235> Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/10400435.2021.1893235>

[20] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[21] Clancy Conlon, Barbra Zupan, and Robyn Preston. 2024. The confidence and competence of speech language pathologists in augmentative and alternative communication: a scoping review. *Augmentative and Alternative Communication* 40, 4 (2024), 292–305.

[22] Humphrey Curtiss, Timothy Neate, and Carlota Vazquez Gonzalez. 2022. State of the art in AAC: A systematic review and taxonomy. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–22.

[23] Beth E Davidoff. 2017. AAC With Energy—Earlier: Research shows that children with communication challenges do best when introduced to augmentative and alternative communication as early as 12 months. *The ASHA leader* 22, 1 (2017), 48–53.

[24] Katherine M DeJarnette and Jane R Wegner. 2020. Preservice education in augmentative and alternative communication: An update. *Perspectives of the ASHA Special Interest Groups* 5, 5 (2020), 1243–1261.

[25] Paul Denny, Stephen MacNeil, Jaromir Savelka, Leo Porter, and Andrew Luxton-Reilly. 2024. Desirable Characteristics for AI Teaching Assistants in Programming Education. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1* (Milan, Italy) (ITiCSE 2024). Association for Computing Machinery, New York, NY, USA, 408–414. <https://doi.org/10.1145/3649217.3653574>

[26] Zijian Ding, Michelle Brachman, Joel Chan, and Werner Geyer. 2025. "The Diagram is like Guardrails": Structuring GenAI-assisted Hypotheses Exploration with an Interactive Shared Representation. In *Proceedings of the 2025 Conference on Creativity and Cognition (C&C '25)*. Association for Computing Machinery, New York, NY, USA, 606–625. <https://doi.org/10.1145/3698061.3726935>

[27] P Dowden. 2004. Summary of the communicative independence model. *UW Aug-comm.* http://depts.washington.edu/augcomm/03_cimodel/commands5_summary.htm (2004).

[28] Kathryn DR Drager, Janice Light, Jessica Currall, Nimisha Muttiah, Vanessa Smith, Danielle Kreis, Alyssa Nilam-Hall, Daniel Parratt, Kaitlin Schuessler, Kaitlin Shermetna, et al. 2019. AAC technologies with visual scene displays and "just in time" programming and symbolic communication turns expressed by students with severe disability. *Journal of intellectual & developmental disability* 44, 3 (2019), 321–336.

[29] William V Dube and Krista M Wilkinson. 2014. The potential influence of stimulus overselectivity in AAC: Information from eye tracking and behavioral studies of attention with individuals with intellectual disabilities. *Augmentative and Alternative Communication* 30, 2 (2014), 172–185.

[30] Esin Durmus, Karina Nguyen, Thomas I Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, Zac Hatfield-Dodds, Danny Hernandez, Nicholas Joseph, et al. 2023. Towards measuring the representation of subjective global opinions in language models. *arXiv preprint arXiv:2306.16388* (2023).

[31] Alexander Fiananca, Ann Paradiso, Mira Shah, and Meredith Ringel Morris. 2017. AACrobot: Using Mobile Devices to Lower Communication Barriers and Provide Autonomy with Gaze-Based AAC. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (CSCW '17). Association for Computing Machinery, New York, NY, USA, 683–695. <https://doi.org/10.1145/2998181.2998215>

[32] Cathy Flores and Shakila Dada. 2025. The effect of AAC training programs on professionals' knowledge, skills and self-efficacy in AAC: a scoping review. *Augmentative and Alternative Communication* 41, 1 (2025), 65–77.

[33] Mauricio Fontana de Vargas, Jiamin Dai, and Karyn Moffatt. 2022. AAC with automated vocabulary from photographs: Insights from school and speech-language therapy settings. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.

[34] Mauricio Fontana De Vargas, Christina Yu, Howard C Shane, and Karyn Moffatt. 2024. Co-Designing QuickPic: Automated Topic-Specific Communication Boards from Photographs for AAC-Based Language Instruction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.

[35] Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. , 277–304 pages.

[36] Jennifer B Ganz. 2015. AAC interventions for individuals with autism spectrum disorders: State of the science and future research directions. *Augmentative and Alternative Communication* 31, 3 (2015), 203–214.

[37] Cindy Gevarter, Mark F O'Reilly, Laura Rojeski, Nicolette Sammarco, Jeff Sigafoos, Giulio E Lancioni, and Russell Lang. 2014. Comparing acquisition of AAC-based mands in three young children with autism spectrum disorder using iPad® applications with different display and design elements. *Journal of Autism and Developmental Disorders* 44 (2014), 2464–2474.

[38] Deepak Giri and Erin Brady. 2023. Exploring outlooks towards generative AI-based assistive technologies for people with Autism. *arXiv preprint arXiv:2305.09815* (2023).

[39] Kate S Glazko, Momona Yamagami, Aashaka Desai, Kelly Avery Mack, Venkatesh Potluri, Xuhai Xu, and Jennifer Mankoff. 2023. An autoethnographic case study of generative artificial intelligence's utility for accessibility. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.

[40] Meredith K Gohsman and Rachel K Johnson. 2023. Reported barriers to augmentative and alternative communication service delivery and learning preferences among speech-language pathologists. *American Journal of Speech-Language Pathology* 32, 4 (2023), 1595–1609.

[41] David J Hajjar and Kathleen Mulkerin. 2023. Visual scene displays for children and adults: Using case studies to bridge research and clinical practice. *Perspectives of the ASHA Special Interest Groups* 8, 2 (2023), 322–336.

[42] Pim Haselager, Hanna Schraffenberger, Serge Thill, Simon Fischer, Pablo Lanillos, Sebastiaan Van De Goeij, and Miranda Van Hooff. 2024. Reflection machines: Supporting effective human oversight over medical decision support systems. *Cambridge Quarterly of Healthcare Ethics* 33, 3 (2024), 380–389.

[43] Andrew Head, Kyle Lo, Dongyeeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 413, 18 pages. <https://doi.org/10.1145/3411764.3445648>

[44] Christine Holyfield, Jessica Caron, and Janice Light. 2019. Programming AAC just-in-time for beginning communicators: The process. *Augmentative and Alternative Communication* 35, 4 (2019), 309–318.

[45] Christine Holyfield, Jessica Gosnell Caron, Kathryn Drager, and Janice Light. 2019. Effect of mobile technology featuring visual scene displays and just-in-time programming on communication turns by preadolescent and adolescent beginning communicators. *International Journal of Speech-Language Pathology* 21, 2 (March 2019), 201–211. <https://doi.org/10.1080/17549507.2018.1441440> Publisher: Taylor & Francis .eprint: <https://doi.org/10.1080/17549507.2018.1441440>.

[46] Christine Holyfield, Kathryn Drager, Janice Light, and Jessica Gosnell Caron. 2017. Typical toddlers' participation in "just-in-time" programming of vocabulary for visual scene display augmentative and alternative communication apps on mobile technology: A descriptive study. *American Journal of Speech-Language Pathology* 26, 3 (2017), 737–749.

[47] Christine Holyfield, Stephen MacNeil, Nicolette Caldwell, Tara O'Neill Zimmerman, Elizabeth Lorah, Eduard Dragut, and Slobodan Vučetić. 2024. Leveraging communication partner speech to automate augmented input for children on the autism spectrum who are minimally verbal: Prototype development and preliminary efficacy investigation. Technical Report. American Speech-Language-Hearing Association.

[48] Christine Holyfield, Tara O'Neill Zimmerman, Stephen MacNeil, Nicolette Sammarco Caldwell, Parth Patel, Breanna Griffen, Elizabeth Lorah, Eduard Dragut, and Slobodan Vučetić. [n. d.]. Preliminary Investigation of Context-Aware Augmentative and Alternative Communication with Automated Just-in-Time Cloze Phrase Response Options for Social Participation from Children on the Autism Spectrum. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP)* [n. d.], 1–15.

[49] Irene Hou, Sophia Mettillé, Owen Man, Zhuo Li, Cynthia Zastudil, and Stephen MacNeil. 2024. The effects of generative ai on computing students' help-seeking preferences. In *Proceedings of the 26th australasian computing education conference*. 39–48.

[50] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. <https://doi.org/10.1145/3544548.3581196>

[51] Shaun K Kane, Barbara Linam-Church, Kyle Althoff, and Denise McCall. 2012. What we talk about: designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 49–56.

[52] Markelle Kelly, Aakriti Kumar, Padhraic Smyth, and Mark Steyvers. 2023. Capturing Humans' Mental Models of AI: An Item Response Theory Approach. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1723–1734. <https://doi.org/10.1145/3593013.3594111>

[53] Thomas Kovacs. 2021. A survey of American speech-language pathologists' perspectives on augmentative and alternative communication assessment and intervention across language domains. *American Journal of Speech-Language Pathology* 30, 3 (2021), 1038–1048.

[54] J. Richard Landis and Gary G. Koch. 1977. An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics* 33, 2 (1977), 363–374. <http://www.jstor.org/stable/2529786>

[55] Emily Laubscher, Janice Light, and David McNaughton. 2019. Effect of an application with video visual scene displays on communication during play: Pilot study of a child with autism spectrum disorder and a peer. *Augmentative and Alternative Communication* 35, 4 (2019), 299–308.

[56] Emily Laubscher, Tracy J Raulston, and Ciara Ousley. 2022. Supporting peer interactions in the inclusive preschool classroom using visual scene displays. *Journal of Special Education Technology* 37, 2 (2022), 318–326.

[57] Mark Liffiton, Brad E Sheese, Jaromír Savelka, and Paul Denny. 2024. Code-Help: Using Large Language Models with Guardrails for Scalable Support in Programming Classes. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research* (Koli, Finland) (Koli Calling '23). Association for Computing Machinery, New York, NY, USA, Article 8, 11 pages. <https://doi.org/10.1145/3631802.3631830>

[58] Janice Light. 1988. Interaction involving individuals using augmentative and alternative communication systems: State of the art and future directions. *Augmentative and alternative communication* 4, 2 (1988), 66–82.

[59] Janice Light, Kathryn Drager, John McCarthy, Suzanne Mellott, Diane Millar, Craig Parrish, Arielle Parsons, Stacy Rhoads, Maricka Ward, and Michelle Welliver. 2004. Performance of typically developing four-and five-year-old children with AAC systems using different language organization techniques. *Augmentative and Alternative Communication* 20, 2 (2004), 63–88.

[60] Janice Light and David McNaughton. 2012. Supporting the communication, language, and literacy development of children with complex communication needs: State of the science and future research priorities. *Assistive technology* 24, 1 (2012), 34–44.

[61] Janice Light, David McNaughton, and Jessica Caron. 2019. New and emerging AAC technology supports for children with complex communication needs and their communication partners: State of the science and future research directions. *Augmentative and Alternative Communication* 35, 1 (2019), 26–41.

[62] Janice Light, Krista M Wilkinson, Amber Thiessen, David R Beukelman, and Susan Koch Fager. 2019. Designing effective AAC displays for individuals with developmental or acquired disabilities: State of the science and future research directions. *Augmentative and Alternative Communication* 35, 1 (2019), 42–55.

[63] Nannan Liu. 2025. Exploring the factors influencing the adoption of artificial intelligence technology by university teachers: the mediating role of confidence and AI readiness. *BMC psychology* 13, 1 (2025), 311.

[64] Kelly Avery Mack, Rida Qadri, Remi Denton, Shaun K. Kane, and Cynthia L. Bennett. 2024. "They only care to show us the wheelchair": disability representation in text-to-image AI models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 288, 23 pages. <https://doi.org/10.1145/3613904.3642166>

[65] Stephen MacNeil, Zijian Ding, Kexin Quan, Thomas J Parashos, Yajie Sun, and Steven P. Dow. 2021. Framing Creative Work: Helping Novices Frame Better Problems through Interactive Scaffolding. In *Proceedings of the 13th Conference on Creativity and Cognition* (Virtual Event, Italy) (C&C '21). Association for Computing Machinery, New York, NY, USA, Article 30, 10 pages. <https://doi.org/10.1145/3450741.3465261>

[66] Stephen MacNeil, Parth Patel, and Benjamin E. Smolin. 2022. Expert Goggles: Detecting and Annotating Visualizations using a Machine Learning Classifier. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (UIST '22 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 58, 3 pages. <https://doi.org/10.1145/3526114.3558627>

[67] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. 2023. Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems* 36 (2023), 46534–46594.

[68] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. <https://doi.org/10.1145/3359174>

[69] David McNaughton, Tracy Rackensperger, Elizabeth Benedek-Wood, Carole Krezman, Michael B Williams, and Janice Light. 2008. "A child needs to be given a chance to succeed": Parents of individuals who use AAC describe the benefits and challenges of learning AAC technologies. *Augmentative and alternative communication* 24, 1 (2008), 43–55.

[70] Pat Mirenda and Peggy A. Locke. 1989. A Comparison of Symbol Transparency in Nonspeaking Persons with Intellectual Disabilities. *Journal of Speech and Hearing Disorders* 54, 2 (May 1989), 131–140. <https://doi.org/10.1044/jshd.5402.131> Publisher: American Speech-Language-Hearing Association.

[71] Aimee Mooney, Steven Bedrick, Glory Noethe, Scott Spaulding, and Melanie Fried-Oken. 2018. Mobile technology to support lexical retrieval during activity recall in primary progressive aphasia. *Aphasiology* 32, 6 (Mar 2018), 666–692. <https://doi.org/10.1080/02687038.2018.1447640>

[72] Jerad Moxley, Joseph Sharit, Sara J Czaja, et al. 2022. The factors influencing older adults' decisions surrounding adoption of technology: Quantitative experimental study. *JMIR aging* 5, 4 (2022), e39890.

[73] Matej Nakić, Goran Oblaković, Ivan David Dogan, and Mato Njavro. [n. d.]. ChatGPT & Metacognitive Confidence: The Role of Generative AI in Promoting False Certainty Over Critical Thought. In *Cambridge Generative AI in Education Conference 2024*.

[74] Mmachi God'sglory Obiorah, Anne Marie Marie Piper, and Michael Horn. 2021. Designing AACs for People with Aphasia Dining in Restaurants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 496, 14 pages. <https://doi.org/10.1145/3411764.3445280>

[75] Sayantan Pal, Souvik Das, Rohini K Srikari, Jeff Higginborham, and Jenna Bizovi. 2024. Empowering AAC Users: A Systematic Integration of Personal Narratives with Conversational AI. In *Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)*. 12–25.

[76] Rhea Paul. 1997. Facilitating transitions in language development for children using AAC. *Augmentative and Alternative Communication* 13, 3 (1997), 141–148.

[77] James Prather, Brent N Reeves, Juho Leinonen, Stephen MacNeil, Arisoa S Randidrasolo, Brett A Becker, Bailey Kimmel, Jared Wright, and Ben Briggs. 2024. The widening gap: The benefits and harms of generative ai for novice programmers. In *Proceedings of the 2024 ACM Conference on International Computing Education Research-Volume 1*. 469–486.

[78] Tracy Rackensperger, Carole Krezman, David Mcnaughton, Michael B Williams, and Karen D'silva. 2005. "When I first got it, I wanted to throw it off a cliff": The challenges and benefits of learning AAC technologies as described by adults who use AAC. *Augmentative and alternative communication* 21, 3 (2005), 165–186.

[79] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[80] Ralf W Schlosser, Howard C Shane, Anna A Allen, Jennifer Abramson, Emily Laubscher, and Katherine Dimery. 2016. Just-in-time supports in augmentative and alternative communication. *Journal of Developmental and Physical Disabilities* 28 (2016), 177–193.

[81] Ather Sharif, Aedan Liam McCall, and Kianna Roces Bolante. 2022. Should I Say “Disabled People” or “People with Disabilities”? Language Preferences of Disabled People Between Identity- and Person-First Language. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. Renee Shelby, Shalaleh Rismani, and Negar Rostamzadeh. 2024. Generative AI in Creative Practice: ML-Artist Folk Theories of T2I Use, Harm, and Harm-Reduction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 32, 17 pages. <https://doi.org/10.1145/3613904.3642461>

[83] Rachel V. Staddon. 2023. Exploring higher education students’ perspectives on factors affecting use, attitudes and confidence with learning technologies. *International Journal of Instruction* 16 (2023), 31–52. Issue 2. <https://durhamrepository.worktribe.com/output/1189600> EPrint Processing Status: Full text deposited in DRO.

[84] Amanda Taboas, Karla Doepke, and Corinne Zimmerman. 2023. Preferences for identity-first versus person-first language in a US sample of autism stakeholders. *Autism* 27, 2 (2023), 565–570.

[85] Jennifer Thistle and Amber Thiessen. 2021. Effectiveness and utility of visual scene displays as communication supports for adults and children with complex communication needs: Perspectives of clinicians and researchers. *American Journal of Speech-Language Pathology* 30, 3 (2021), 1181–1193.

[86] Jennifer J Thistle, Amber Thiessen, and Jeffery Hart. 2023. “I don’t think you can really understand it until you put it to action”: school-based SLPs’ reflections on graduate and continuing education in AAC. *Augmentative and Alternative Communication* 39, 4 (2023), 270–281.

[87] Jennifer J Thistle and Krista M Wilkinson. 2021. Speech-Language pathologists’ decisions when designing an aided AAC display for a compilation case study of a beginning communicator. *Disability and Rehabilitation: Assistive Technology* 16, 8 (2021), 871–879.

[88] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shana K. Kane. 2023. “The less I type, the better”: How AI Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 830, 14 pages. <https://doi.org/10.1145/3544548.3581560>

[89] Stephanie Valencia, Jessica Huynh, Emma Y Jiang, Yufei Wu, Teresa Wan, Zixuan Zheng, Henny Admoni, Jeffrey P Bigham, and Amy Pavel. 2024. COMPA: Using Conversation Context to Achieve Common Ground in AAC. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[90] Eva AM Van Dis, Johan Bollen, Willem Zuidema, Robert Van Rooij, and Claudi L Bockting. 2023. ChatGPT: five priorities for research. *Nature* 614, 7947 (2023), 224–226.

[91] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S Bernstein, and Ranjay Krishna. 2023. Explanations can reduce overreliance on AI systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[92] Veniamin Veselovsky, Manoel Horta Ribeiro, Philip J. Cozzolino, Andrew Gordon, David Rothschild, and Robert West. 2025. Prevalence and Prevention of Large Language Model Use in Crowd Work. *Commun. ACM* 68, 3 (Feb. 2025), 42–47. <https://doi.org/10.1145/3685527>

[93] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, et al. 2021. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359* (2021).

[94] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design principles for generative AI applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.

[95] Krista M Wilkinson, Rick Gilmore, and Yiming Qian. 2022. Judicious arrangement of symbols on a simulated augmentative and alternative communication display optimizes visual attention by individuals with Down syndrome. *Journal of Speech, Language, and Hearing Research* 65, 2 (2022), 710–726.

[96] Krista M Wilkinson, Janice Light, and Kathryn Drager. 2012. Considerations for the composition of visual scene displays: Potential contributions of information from visual and cognitive sciences. *Augmentative and Alternative Communication* 28, 3 (2012), 137–147.

[97] Cynthia Zastudil, Christine Holyfield, Christine Kapp, Xandria Crosland, Elizabeth R Lorah, Tara Zimmerman, and Stephen MacNeil. 2024. Exploring the use of generative AI to support automated just-in-time programming for visual scene displays. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6.

[98] Cynthia Zastudil, Magdalena Rogalska, Christine Kapp, Jennifer Vaughn, and Stephen MacNeil. 2023. Generative AI in computing education: Perspectives of students and instructors. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.

[99] Xiaoyi Zhang, Harish Kulkarni, and Meredith Ringel Morris. 2017. Smartphone-Based Gaze Gesture Communication for People with Motor Disabilities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 2878–2889. <https://doi.org/10.1145/3025453.3025790>

[100] Annuska Zolyomi, Anushree Shukla, and Jaime Snyder. 2017. Technology-mediated sight: A case study of early adopters of a low vision assistive technology. In *Proceedings of the 19th international ACM SIGACCESS conference on computers and accessibility*. 220–229.

A Images and Case Studies Used for VSD Creation

VSD 1. Children Playing in a Sandbox. (see Figure 8a) “Imagine you took this photo to create a VSD for the child in the yellow shirt. He is a 7 year old on the autism spectrum. His goals in therapy are focused on building an expressive and receptive single word vocabulary, as his parents and professionals estimate that he understands about 25 words and spontaneously communicates about 10 words. He loves all forms of vehicles and playing with his sister Cara, also pictured.”

VSD 2. Children Petting a Chicken. (see Figure 8b) “Imagine the parent of the girl in the yellow shirt in this photo sent the photo to you to create a VSD of a fun moment from the weekend. She is a 4 year old on the autism spectrum. Her goals in therapy are focused on expanding her receptive and expressive single word vocabulary (estimated around 150 words) and the emergence of two-word combinations. She loves all animals and spending time with her friend Dean, also pictured.”

VSD 3. Children Playing with Blocks at a Table. (see Figure 8c) “Imagine you took this photo to create a VSD for the child in the red plaid shirt. He is a 9 year old on the autism spectrum. His goals in therapy are focused on building a single word vocabulary because his estimated receptive word knowledge to be fewer than 50 words and his spontaneous use of words expressively is limited (about five words). He loves colorful toys and all building-based toys. He also enjoys playing with his little cousin Jonas, also pictured.”

VSD 4. Children Riding a Carousel. (see Figure 8d) “Imagine the parent of the girl in the white shirt in this photo sent the photo to you to create a VSD of a fun moment from the weekend. She is a 6 year old on the autism spectrum. Her goals in therapy are focused on expanding her single word vocabulary (expressive and receptive) as she is estimated to understand about 250 words and use about 50 spontaneously. Her therapy is also focused on the emergence of two-word combinations. She loves animals and spending time with her big brother, Sam, also pictured.”

B Post-Questionnaire

Participants were asked to rate their agreement to the following statements at the conclusion of our user study. Each statement was accompanied by a 7-point Likert scale for measuring agreement (Strongly agree, Agree, Somewhat agree, Neither agree or disagree, Somewhat disagree, Disagree, Strongly disagree).

- (1) The generated vocabulary included words I wanted to use.
- (2) The generated vocabulary included words I did not want to use.
- (3) The generated vocabulary included words I would not have thought of that are relevant.
- (4) The generated vocabulary included words which were redundant.
- (5) The vocabulary generated included words I would target during educational and/or speech therapy.
- (6) Overall, the vocabulary generated was effective in helping me achieve targeted goals.



(a) VSD 1. Children playing in a sandbox.



(b) VSD 2. Children petting a chicken.



(c) VSD 3. Children playing with blocks at a table.



(d) VSD 4. Children riding a carousel.

Figure 8: All four VSD images we used in our user study.