Exploring the use of Generative AI to Support Automated Just-in-Time Programming for Visual Scene Displays

CYNTHIA ZASTUDIL, Temple University, USA
CHRISTINE HOLYFIELD, University of Arkansas, USA
CHRISTINE KAPP, Temple University, USA
XANDRIA CROSLAND, Western Governors University, USA
ELIZABETH LORAH, University of Arkansas, USA
TARA ZIMMERMAN, University of Arkansas, USA
STEPHEN MACNEIL, Temple University, USA

Millions of people worldwide rely on alternative and augmentative communication devices to communicate. Visual scene displays (VSDs) can enhance communication for these individuals by embedding communication options within contextualized images. However, existing VSDs often present default images that may lack relevance or require manual configuration, placing a significant burden on communication partners. In this study, we assess the feasibility of leveraging large multimodal models (LMM), such as GPT-4V, to automatically create communication options for VSDs. Communication options were sourced from a LMM and speech-language pathologists (SLPs) and AAC researchers (N=13) for evaluation through an expert assessment conducted by the SLPs and AAC researchers. We present the study's findings, supplemented by insights from semi-structured interviews (N=5) about SLP's and AAC researchers' opinions on the use of generative AI in agumentative and alternative communication devices. Our results indicate that the communication options generated by the LMM were contextually relevant and often resembled those created by humans. However, vital questions remain that must be addressed before LMMs can be confidently implemented in AAC devices.

CCS Concepts: • **Human-centered computing** → **Accessibility technologies**; **Accessibility systems and tools**.

Additional Key Words and Phrases: AAC, autism, visual screen displays, VSDs, generative AI, just-in-time programming

ACM Reference Format:
Cynthia Zastudil, Christine Holyfield, Christine Kapp, Xandria Crosland, Elizabeth 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 *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24), October 27–30, 2024, St. John's, NL, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3663548.3688502

1 INTRODUCTION & BACKGROUND

Visual scene displays (VSDs) are a form of augmentative and alternative communication (AAC) which use photographs or other images with interactive "hotspots" placed on them to represent language concepts [3, 28, 34] (see the example in Figure 1). VSDs have proven especially useful for beginning communicators (i.e., communicators who are learning their first words) because they incorporate personally relevant imagery [26, 34], maintain the relationship between people and objects [25], combine subjects or activities within a single visual context [26], and reduce the visual cognitive demands typically associated with AAC by aligning with natural visual processing [28].

Manuscript submitted to ACM

While VSDs offer numerous benefits for beginning communicators, several challenges can reduce their effectiveness, impacting both their adoption and retention. One significant issue is that the default imagery and communication options (COs) provided by VSDs are only relevant in specific settings, which limits their effectiveness [32]. This issue has led to the development of just-in-time (JIT) programming [6, 8, 16, 31], where communication partners manually create COs for VSD users in real-time, tailored to the specific image or naturally occurring scene. While JIT programming can improve communication outcomes for users [16], it requires clinicians to be present and continuously reconfiguring the user interface to capture contextually relevant and engaging scenarios [16].

Prior work has evaluated the potential to automatically generate COs for users. For example, Holyfield et al. investigated the potential to augment a grid display using communication partner speech input, increasing participation of young children using the device [18]. Prior work has also investigated incorporating automatically generated COs based on a photograph into an AAC device by creating topic-specific grid displays [10, 11]. However, it is crucial to carefully examine the content, focus, and relevance of AI-generated communication options (COs), particularly considering the potential for racial and gender biases inherent in AI systems [1, 7, 19, 29]. This raises questions about whether AI-generated communication options can serve as a useful support for beginning communicators.

In this work, we used a LMM to generate COs for VSDs intended for use by young children on the autism spectrum or with other developmental disabilities. We compared these COs to options created by speech-language pathologists (SLP) and AAC researchers (N=13) to compare how the relevance and focus of the COs differ between the two sets of COs. Expert VSD researchers (N=5) then conducted an evaluation to compare the quality of human- and LMM-generated COs. Lastly, we conducted semi-structured interviews with expert VSD researchers (N=5) to better understand the implications of these models on AAC devices.

In this work, we investigate the following research questions: **(RQ1)** How do communication options generated by SLPs, AAC researchers, and a LMM compare in terms of perceived relevance, topics of focus, and quality? and **(RQ2)** What are the perceptions of SLPs and researchers who use VSDs on the use and ethics of the use of LMMs for just-in-time programming of VSDs?

We found that generative AI created communication options align well with those created by SLPs and AAC researchers. However, we also discovered challenges. Specifically, SLPs draw on a deep understanding of the individual contexts and backgrounds of their clients to provide tailored support, a level of personalization that generative AI currently lacks. Additionally, it is unclear how harmful developmentally inappropriate communication options that may be generated would be on their language development.

2 STUDY 1: COMPARING HUMAN- AND LMM-GENERATED COMMUNICATION OPTIONS

We created a corpus of COs for multiple scenarios sourced from people and LMMs. We collected the human-created COs by surveying SLPs and AAC researchers. We created the LMM-generated COs using OpenAI's GPT-4V[1], the most popular LMM at the time this work was done. We then conducted a comparison between the human- and LMM-generated COs using a combination of deductive coding, part-of-speech (POS) analysis, and expert analysis.

2.1 Communication Option Collection

We collected COs from SLPs and AAC researchers (N=13) via a survey. The range of professional or clinical experience for our participants was between 1 and 25 years ($\bar{x}$ = 9.3). The majority (7/13) of our participants were frequent users

[1] https://openai.com/gpt-4

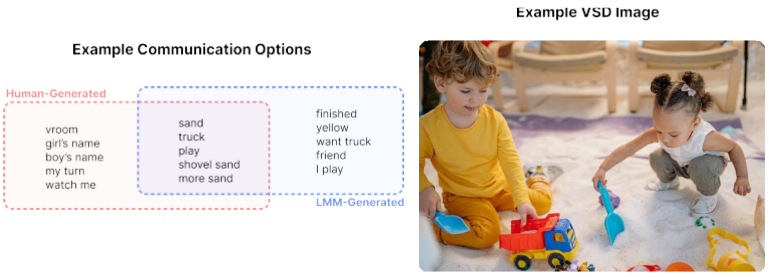

Fig. 1. An example image which could be used in a VSD. Example COs generated by human participants and by the LMM are provided in the figure. The COs can be embedded within the image as clickable "hotspots" or as buttons presented on the display.

or researchers of VSDs. Participants created COs for three contexts: playing, reading a storybook, and retelling a past activity. These contexts were selected because they are common use cases for VSDs [2, 5, 22]. We selected six different images, two for each context. See Figure 1 for a sample image shown to participants. For each context, participants created COs for one image per context which was randomly chosen between the two image variants in order to create a variety of COs for future comparison. Vignettes, which are common in AAC research [16], were provided to participants to provide context about how the VSD would be used and the linguistic abilities of the fictional child using the VSD. Participants were provided with two different vignettes for different communication stages for each context: children working on building engagement in interactions and the emergence of words (i.e., pre-linguistic) and children focused on beginning to combine words (i.e., multiword). We refer to these communication stages as pre-linguistic and multiword, respectively. A sample vignette for the playing context for each of the communication stage ((A) pre-linguistic and (B) multiword) is included below: "You took this photo to create a VSD for the child in the yellow shirt in this photo. Please write out the COs you would program for them if you were focused on [(A) building engagement in interactions and the emergence of words (B) beginning to combine words]."

We prompted GPT-4V to generate COs for the same contexts as our human participants. We generated an equal number of sets of COs as we obtained from our survey participants. Our prompt to the model contained instructions for generating COs, including the communication stage and a vignette similar to what the survey participants saw. An example prompt (using the image in Figure 1) for the pre-linguistic context is provided here: "You're an assistant to generate vocabulary for pre-linguistic communicators on the autism spectrum who use AAC devices in the form of visual screen displays. This photo was taken to be used in a visual screen display for the child in the yellow shirt in the picture. Please write out the most contextually relevant communication options you would program for them if you were focused on building engagement in interactions and the emergence of words."

2.2 Communication Option Analysis

First, we conducted a POS analysis to determine how the content and structure of the COs aligned across the human- and LMM-generated COs. In order to ensure no parts-of-speech were missed, we split every CO into single words. Additionally, to evaluate differences and similarity in the focus of generated COs, we conducted a deductive coding process for both the human- and LMM-generated COs. We used the ''Four Functions of Communication" framework [23] as our coding scheme: expressing wants or needs (intended to make requests), information transfer (intended to share information with others), social closeness (intended to develop or maintain relationships), and social etiquette (intended to convey polite terms (e.g., "thank you")). We also included an "Other" category to handle COs which did not clearly

Part of Speech	Human	LMM
Adjectives	7.60%	10.10%
Adverbs	5.00%	8.90%
Interjections	1.70%	2.30%
Nouns	39.2%	33.8%
Particles/Determiners/Conjunctions	6.40%	1.20%
Prepositions	5.20%	5.10%
Pronouns	6.70%	6.90%
Verbs	25.8%	31.7%

Table 1. The distribution of part-of-speech for COs.

Code	Human	LMM
Expressing Wants and Needs	16.7%	16.9%
Information Transfer	69.9%	79.4%
Social Closeness	6.2%	0.8%
Social Etiquette	2.0%	1.8%
Other	5.2%	2.3%

Table 2. The results of the deductive coding using Light's [23] Four Functions of Communication framework.

align with these four functions. Two researchers performed the coding and inter-rater reliability was computed. The inter-rater reliability score was 0.65 indicating substantial agreement between raters [20].

After we compared the content and focus of the COs, we conducted a follow-up survey with a subset of experts (N=5) (i.e., more than 5 years of experience and extensive experience using and configuring VSDs) participants from our first survey. Each participant was shown COs for the same contexts as our first survey. Participants did not rate the COs they previously created. Participants were not aware that any options had been LMM-generated. Participants rated LMM- and human-generated COs on a scale from 1 to 5 (1 being the worst). Each participant rated between 68 and 80 sets of COs for a total of 364 ratings.

2.3 Results

In total, human participants created 306 COs across all contexts and the LMM generated 379 (see Figure 1 for examples). From the results of our POS analysis and deductive coding show that the content and focus of human- and LMM-generated COs are very similar. Table 1 shows the POS frequencies for the COs created by survey participants and the LMM. There is generally alignment in the structure and content of the COs. Nouns and verbs are the most commonly used parts-of-speech for both people and the LMM.

In addition to content and structure, the focus of COs was fairly similar for human- and LMM-generated COs (see Table 2). People focused primarily on information transfer and expressing wants and needs, which is congruent with existing AAC research [14, 17, 27]. Similarly, the LMM also focused mostly on information transfer and the expression of wants and needs. However, the LMM generated very few COs for social closeness. Additionally, participants generated 2.2 times as many COs which belonged to the other category which were commonly sound effects (e.g., "vroom", "bawk bawk"). We observed a similar focus on social etiquette between the human- and LMM-generated COs.

Based on our expert evaluation, LMMs can generate COs of similar, if not better quality than humans (see Figure 2). For the playing context, the human-generated COs were generally preferred over the LMM-generated COs. For the reading of a storybook and retelling of a past activity, however, the LMM-generated COs were generally preferred over the human-generated COs. While there are differences in the average ratings, the human- and LMM-generated COs perform similarly across all contexts.

3 STUDY 2: SEMI-STRUCTURED INTERVIEWS WITH CLINICIANS AND AAC RESEARCHERS

We conducted semi-structured interviews (N=5) with SLPs and AAC researchers to begin to understand potential benefits and issues with using generative AI for automation of VSDs. Participants had between 5 and 15 years of

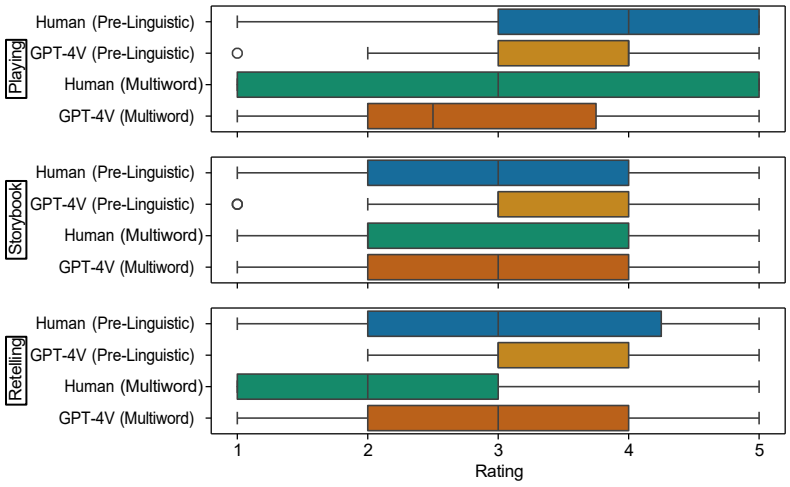

Fig. 2. A comparison of the experts' ratings of COs generated by human participants and GPT-4V. Human-generated COs were preferred for the Playing context; however, for the Storybook and Retelling contexts, LMM-generated COs were preferred.

experience ($\bar{x}$ = 9) using VSDs clinically or in intervention research and all participants have published articles on VSDs, vocabulary selection for beginning communicators, or JIT programming for AAC devices. One researcher on our team performed a thematic analysis [4] on the interview transcripts by reviewing participants' responses, coding participants' insights, and identifying themes.

3.1 Results

Multiple themes emerged through our interviews about the potential benefits of incorporating generative AI in AAC devices for automated programming, including the reduction of effort required for JIT programming by SLPs and improved accessibility of VSDs and other AAC for untrained communication partners. Two primary themes emerged regarding participants' concerns about potential harms of LMM-generated COs: lack of personalization of LMM-generated COs and developmentally inappropriate LMM-generated COs. Four participants (P1, P2, P3, P4) expressed concerns about how AI-generated COs lack the context that exists between communication partners and AAC users, and as such, will likely miss out on personally relevant COs, especially regarding culturally relevant or family-oriented COs. Four participants (P1, P2, P3, P5) were concerned about the potential risks of dynamically generated communications. Specifically, they were concerned about the harm COs which are not developmentally appropriate could pose to advancement of communication skills.

4 DISCUSSION & FUTURE WORK

Our findings indicate that the content, focus, and quality of LMM-generated COs were generally comparable to those created by humans. However, human-generated COs often included more sound effects and language aimed at social engagement. This difference represents a minor departure from best practices [9, 15, 27, 30, 33].

Furthermore, our interview study identified concerns participants had about whether the COs would be developmentally appropriate or personally relevant for users. This highlights a key value that SLPs bring to the usage of

VSDs—detailed knowledge about the user's interests and their linguistic ability. Personalization is a key aspect of early language development [13, 21, 24], and, currently, generative AI does not provide the personalization necessary for the effective use of VSDs for early language development. Future work could explore the development of personalized user models to address this issue; however, the value SLPs and communication partners bring to interactions with VSDs cannot and should not be replaced. Future work should leverage communication partners' insights to develop a better understanding of how they can monitor and edit generated COs to ensure their relevance and appropriateness. LMMs have been observed to produce harmful output which contains negative stereotypes and biases [1, 7, 19, 29]. While we did not observe any occurrences of this in our research, it is critical that future work prioritizes identifying mitigation strategies to prevent these stereotypes and biases from being present in any AAC devices relying on LMM output. Lastly, future work should also incorporate appropriate methods to reduce linguistic demands [12] allowing for VSD end users to participate in the design process.

ACKNOWLEDGMENTS

We are grateful to Abi Roper for her guidance and expertise which helped refine the framing of this poster. This research was partially funded by the Temple University College of Science and Technology Research Scholars Program and Convergence Accelerator Grant (National Science Foundation grant number ITE-2236352).

REFERENCES

[1] Lena Armstrong, Abbey Liu, Stephen MacNeil, and Danaë Metaxa. 2024. The Silicone Ceiling: Auditing GPT's Race and Gender Biases in Hiring. *arXiv preprint arXiv:2405.04412* (2024).

[2] Naima Bhana, David McNaughton, Tracy Raulston, and Ciara Ousley. 2020. Supporting Communication and Participation in Shared Storybook Reading Using Visual Scene Displays. *TEACHING Exceptional Children* 52, 6 (July 2020), 382–391. https://doi.org/10.1177/0040059920918609 Publisher: SAGE Publications Inc.

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

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[5] 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.

[6] Jessica Caron Christine Holyfield and Janice Light. 2019. Programing AAC just-in-time for beginning communicators: the process. *Augmentative and Alternative Communication* 35, 4 (2019), 309–318. https://doi.org/10.1080/07434618.2019.1686538 arXiv:https://doi.org/10.1080/07434618.2019.1686538 PMID: 31790292.

[7] Zijian Ding, Arvind Srinivasan, Stephen Macneil, and Joel Chan. 2023. Fluid Transformers and Creative Analogies: Exploring Large Language Models' Capacity for Augmenting Cross-Domain Analogical Creativity. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 489–505. https://doi.org/10.1145/3591196.3593516

[8] Kathryn D. R. Drager, Janice Light, Jessica Currall, Nimisha Muttiah, Vanessa Smith, Danielle Kreis, Alyssa Nilam-Hall, Daniel Parratt, Kaitlin Schuessler, Kaitlin Shermetta, and Jill Wiscount. 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 (July 2019), 321–336. https://doi.org/10.3109/13668250.2017.1326585 Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/13668250.2017.1326585.

[9] Larry Fenson, Philip S. Dale, J. Steven Reznick, Elizabeth Bates, Donna J. Thal, Stephen J. Pethick, Michael Tomasello, Carolyn B. Mervis, and Joan Stiles. 1994. Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development* 59, 5 (1994), i–185. http://www.jstor.org/stable/1166093

[10] 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*. ACM, Athens Greece, 1–18. https://doi.org/10.1145/3517428.3544805

[11] 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 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 910, 16 pages. https://doi.org/10.1145/3613904.3642080

[12] Christopher Frauenberger, Judith Good, and Alyssa Alcorn. 2012. Challenges, opportunities and future perspectives in including children with disabilities in the design of interactive technology. In *Proceedings of the 11th International Conference on Interaction Design and Children* (Bremen, Germany) *(IDC '12)*. Association for Computing Machinery, New York, NY, USA, 367–370. https://doi.org/10.1145/2307096.2307171

[13] Bethany J Frick Semmler, Allison Bean, and Laura Wagner. 2024. Examining core vocabulary with language development for early symbolic communicators. *International Journal of Speech-Language Pathology* 26, 1 (2024), 28–37.

[14] 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.

[15] Erika Hoff. 2006. How social contexts support and shape language development. *Developmental review* 26, 1 (2006), 55–88.

[16] 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.

[17] Christine Holyfield, Kathryn DR Drager, Jennifer MD Kremkow, and Janice Light. 2017. Systematic review of AAC intervention research for adolescents and adults with autism spectrum disorder. *Augmentative and alternative communication* 33, 4 (2017), 201–212.

[18] Christine Holyfield, Stephen MacNeil, Nicolette Caldwell, Tara O'Neill Zimmerman, Elizabeth Lorah, Eduard Dragut, and Slobodan Vucetic. 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. *American Journal of Speech-Language Pathology* 33, 3 (2024).

[19] Hadas Kotek, Rikker Dockum, and David Sun. 2023. Gender bias and stereotypes in Large Language Models. In *Proceedings of The ACM Collective Intelligence Conference* (Delft, Netherlands) *(CI '23)*. 12–24.

[20] 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

[21] Emily Laubscher and Janice Light. 2020. Core vocabulary lists for young children and considerations for early language development: A narrative review. *Augmentative and Alternative Communication* 36, 1 (2020), 43–53.

[22] 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.

[23] 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. https://doi.org/10.1080/07434618812331274657 arXiv:https://doi.org/10.1080/07434618812331274657

[24] Janice Light, Allison Barwise, Ann Marie Gardner, and Molly Flynn. 2021. Personalized early AAC intervention to build language and literacy skills: A case study of a 3-year-old with complex communication needs. *Topics in language disorders* 41, 3 (2021), 209–231.

[25] 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 (June 2004), 63–88. https://doi.org/10.1080/07434610410001655553 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07434610410001655553.

[26] 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.

[27] JC Light, AR Parsons, and K Drager. 2002. "There's more to life than cookies". Developing interactions for social closeness with beginning communicators who use AAC. In *Exemplary pratices for beginning communicators: Implications for AAC*. Paul H. Brookes Baltimore, 187–218.

[28] 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 (Baltimore, Md.: 1985)* 35, 1 (March 2019), 42–55. https://doi.org/10.1080/07434618.2018.1558283

[29] Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in large language models: origins, inventory, and discussion. *ACM Journal of Data and Information Quality* 15, 2 (2023), 1–21.

[30] Katherine Nelson. 1973. Structure and Strategy in Learning to Talk. *Monographs of the Society for Research in Child Development* 38, 1/2 (1973), 1–135. http://www.jstor.org/stable/1165788

[31] 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, 1 (Feb. 2016), 177–193. https://doi.org/10.1007/s10882-015-9452-2

[32] Martine Smith and N. Grove. 2003. Asymmetry in input and output for individuals who use augmentative and alternative communication. *Communicative Competence of Individuals WHo Use Augmentative and Alternative Communication* (01 2003), 163–195.

[33] Lev S Vygotsky. 2012. *Thought and language*. MIT press.

[34] Krista M Wilkinson and Janice Light. 2011. Preliminary investigation of visual attention to human figures in photographs: Potential considerations for the design of aided AAC visual scene displays. (2011).

[35] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582

A LMM PROMPTING METHODOLOGY

In order to collect the LMM-generated communication options, we used OpenAI's GPT-4V model, specfically the `gpt-vision-preview` model. All responses were collected in April 2024. We used two types of prompts. The first prompt type was intended to generate communication options for communicators working on building engagement in interactions and the emergence of words (pre-linguistic):

> *"You're an assistant to generate vocabulary for pre-linguistic communicators on the autism spectrum who use AAC devices in the form of visual scene displays. This photo was taken to be used in a visual scene display for [a child pictured in the image]. Please write out the most contextually relevant communication options you would program for them if you were focused on building engagement in interactions and the emergence of words."*

The second prompt was intended to generate communication options for communicators focused on beginning to combine words (multiword):

> *"You're an assistant to generate vocabulary for multiword communicators on the autism spectrum who use AAC devices in the form of visual scene displays. This photo was taken to be used in a visual screen display for [a child pictured in the image]. Please write out the communication options you would program for them if you were focused on beginning to combine words."*

These prompts were chosen because they included necessary contextual information about the end user of the visual scene display (e.g., diagnosis, linguistic ability), and closely mirrored the vignette which was provided to our human participants.

The above prompts results in lists of communication options with on average 20 options. In addition to using the above prompts, we used a technique called prompt chaining [35] to narrow the lists of communication options to the 5 options deemed most relevant by the model. The chaining prompt we used was:

> *"Using the communication options you generated, please identify the five most relevant communication options."*

We acknowledge that there might exist prompts which result in more contextually relevant results. The key limitation of LMMs generating vocabulary for AAC devices, which we identified in this work, is the lack of personalization of the responses, regardless of the quality of the contextually relevant communication options.

Received 3 July 2024; Accepted 9 August 2024