# Exploring AI-Based Support in Speech-Language Pathology for Culturally and Linguistically Diverse Children

Aaleyah Lewis
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
alewis9@cs.washington.edu

Aayushi Dangol
Human Centered Design & Engineering
University of Washington
Seattle, Washington, USA
adango@uw.edu

Hyewon Suh
Human Centered Design & Engineering
University of Washington
Seattle, Washington, USA
hyewon25@uw.edu

Abbie Olszewski
College of Education and Human Development
University of Nevada, Reno
Reno, Nevada, USA
aolszewski@med.unr.edu

James Fogarty
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
jfogarty@cs.washington.edu

Julie A. Kientz
Human Centered Design & Engineering
University of Washington
Seattle, Washington, USA
jkientz@uw.edu

## Abstract

Speech-language pathologists (SLPs) provide support to children with speech and language difficulties through delivering evaluation, assessment, and interventions. Despite growing research on how Artificial Intelligence (AI) can support SLPs, there is limited research examining how AI can assist SLPs in delivering equitable care to culturally and linguistically diverse (CLD) children with disabilities. Through interviews with 15 SLPs and a two-part survey study with 13 SLPs, we report on SLP challenges in delivering responsive care to CLD children with disabilities (i.e., unrepresentative materials, unreliable translation, insufficient support for language variations), areas for AI-based support, evaluations of how available AI performs in addressing these challenges, and bias assessments of AI-generated materials. We discuss implications of contextually unaware AI, the range of care in AI-prompting, tensions and tradeoffs of AI-based support, and honoring diverse representations in AI-generated materials. We offer considerations for SLPs using AI-based tools and general-purpose AI in their practice.

## CCS Concepts

• **Human-centered computing → Empirical studies; Artificial Intelligence**.

## Keywords

Speech and language difficulties, Artificial Intelligence, Generative AI, AI harms, Human-centered AI

**ACM Reference Format:**

Aaleyah Lewis, Aayushi Dangol, Hyewon Suh, Abbie Olszewski, James Fogarty, and Julie A. Kientz. 2025. Exploring AI-Based Support in Speech-Language Pathology for Culturally and Linguistically Diverse Children. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3714131


*CHI '25, Yokohama, Japan*

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714131

## 1 Introduction

Speech-language pathologists (SLPs) play a pivotal role in providing support to more than half of U.S. children aged 3 to 17 [6], including support via the Individuals with Disabilities Education Act. Their services encompass screening, evaluation, assessment, and interventions for speech and language difficulties, delivered across diverse settings (e.g., schools, hospitals, residential care, private practices) in rural, suburban, and urban regions [2]. Challenges that SLPs encounter in their practice have been well-documented, including high caseload and extensive workload [24], much of which is dedicated to administrative tasks (e.g., preparing therapy materials, writing progress reports, setting individualized goals, preparing Individualized Education Plan meetings) [24, 79]. However, these challenges are amplified when SLPs provide care to culturally[1] and linguistically diverse (CLD) children with disabilities, who represent a substantial proportion of children with speech and language difficulties [6]. Despite their numbers, CLD children with disabilities do not always receive equitable care compared to White and monolingual counterparts. This disparity is influenced by multiple factors, including the lack of representative evaluation, assessment, and therapy materials, resources, and training [15, 26, 32, 47, 68] as well as demographic homogeneity of the profession. Notably, 90.5% of SLPs in the U.S. identify as White [1], and 91.7% are monolingual, primarily English-speaking [5]. This lack of diversity, combined with heavy workloads and insufficient access to representative materials, resources, and support impedes the ability of SLPs to provide equitable care for CLD children with disabilities. These challenges

[1]For the purpose of this paper, we use the term "culture" to refer to race, ethnicity, and religion, but we acknowledge that culture encompasses many additional dimensions. We refer to race as a social construct that categorizes people based on several factors, including but not limited to ancestry, physical features, social and behavioral qualities, and self-identification [64, 73].

highlight the need for solutions to better support SLPs in their efforts to deliver responsive and effective care.

Artificial intelligence (AI) has made significant strides, particularly with Generative AI and the emergence of Large Language Models. This has sparked a surge in interest and adoption of AI across professions, including in exploring AI's potential to enhance speech-language pathology services by supporting, rather than replacing, SLPs in various stages of their workflow [31, 36, 79]. Despite such work, there remains limited exploration of how AI can support SLPs in providing responsive and equitable care to CLD children with disabilities [32]. Addressing this gap is critical, as many existing AI tools fail to account for cultural, linguistic, and disability-related nuances that are essential for equitable and effective care.

To address this gap and explore opportunities and challenges in AI-based tools supporting SLPs in delivering more equitable care for CLD children with disabilities, we conducted a two-part study. We report findings from semi-structured interviews and a subsequent survey study with SLPs who work with CLD children with disabilities, focusing on three research questions:

- **RQ1:** What are current experiences and challenges experienced by SLPs in providing culturally, linguistically, and disability responsive practices?
- **RQ2:** What opportunities do SLPs see for AI technologies supporting their culturally, linguistically, and disability responsive practice, and what do they perceive as potential tensions?
- **RQ3:** How does available AI (i.e., ChatGPT-4o) perform in addressing these challenges and opportunities in supporting culturally, linguistically, and disability responsive practices, and how do SLPs perceive the performance of that AI?

In summary, our paper makes three primary contributions:

- We present an interview study with 15 SLPs, reporting participant-identified challenges in delivering responsive care to CLD children with disabilities and participant-identified opportunities for AI-based support.
- We present a subsequent survey study with 13 SLPs, evaluating AI-generated materials to support their delivery of responsive care to CLD children with disabilities. Informed by prior bias literature, we additionally identify and assess cultural, linguistic, and disability biases in these AI-generated materials.
- We discuss implications of contextually unaware AI, the range of care in AI-prompting, tensions and tradeoffs in AI-based support for responsive speech-language pathology, the necessity of honoring diverse representations in AI-generated materials, and design considerations for AI-based support.

## 2 Related Work

We first overview the roles and responsibilities of speech-language pathologists, highlighting challenges in responsive care. Next, we review literature on AI-based support in speech-language pathology. Lastly, we review prior research on pervasive biases in AI and prior methods to uncover and evaluate these biases.

### 2.1 Speech-Language Pathologists and Challenges in Responsive Care

Defined by the American Speech-Language-Hearing Association (ASHA), an SLP is a "professional who engages in professional practice in the areas of communication and swallowing across the life span" [18]. Within their practice, SLPs face significant challenges, including high caseload and extensive workload [24, 79], largely focused on administrative tasks. These challenges are magnified when providing care to culturally and linguistically diverse children with disabilities. Well-documented challenges include assessment bias (e.g., linguistic bias, content bias, disproportionate representation in sampling of standardized evaluations) [15, 26, 32, 38, 47, 56], supporting diverse language development [53], delivering services in diverse languages, accessing training for collaborating with CLD families, and finding sufficient time to implement additional best practices recommended for working with diverse families [26, 43, 52, 55, 62, 65, 78, 87, 88].

Through lenses of DisCrit [14] and Critical Race Theory [33], Harris-Johnson critically examines the intersection of racism and ableism in speech-language pathology, "Racism and ableism is maintained and neutralized in invisible ways throughout speech pathology" [47]. Similarly, through a Critical Race Theory approach [33], Privette examines how systemic racism and intersecting language and disability ideologies shape inequities in speech-language pathology [68]. Overwhelming demographic homogeneity among SLPs (i.e., 90.5% White [1], 91.7% monolingual [5]) contributes to inequities in the field of speech-language pathology, both among practitioners and in the quality of care CLD children with disabilities receive. Dagenais and Stallworth [30] found that SLPs demonstrate bias toward speakers of their own dialect, making them more likely to assign lower ratings in speech and language to speakers from different linguistic backgrounds. Such biases cause challenges in SLPs differentiating between language variations and genuine speech and language disorders, which can result in misidentification of CLD children with disabilities (i.e., over-identification or under-identification) [4, 20, 37, 47, 66, 68] and can reinforce disproportionality in schools and healthcare settings [4, 47, 68, 77]. Moreover, the field's homogeneity affects the diversity of therapy materials. Harris-Johnson critiqued that materials often fail to capture nuances of culture, language, and disability, instead presenting reductive, one-dimensional portrayals [47]. Additionally, in a study exploring SLP selection of representative materials, researchers found that selection of diverse books is related to the race of the SLP rather than their caseload composition — Black SLPs were more likely than White SLPs to report selecting books with diverse representation [46]. Our work builds on these findings, further examining the lack of representative materials in speech-language pathology and investigating how AI can both mitigate and perpetuate such disparities.

### 2.2 AI-Based Support in Speech Language Pathology

Researchers have employed AI and machine learning in speech-language pathology to enhance therapy outcomes, automate speech analysis, provide real-time feedback, and enable personalized treatment plans. For example, Jia et al. analyzed child speech patterns, detecting errors in pronunciation and identifying other phonological issues through pattern-recognition [51]. Bílková et al. [23] used a convolutional neural network in combination with augmented reality to monitor lip, tongue, and teeth movements during speech therapy exercises, providing insights into client articulation patterns. Similarly, Ng et al. [63] and Sztahó et al. [80] developed an automated speech therapy tool offering visual feedback on elements like intensity, intonation, and rhythm for children with hearing impairments. Real-time feedback in such approaches aims to help children understand and address speech errors. Complementary approaches assess speech patterns and progress to create personalized learning plans tailored to a child's specific needs, thereby aiming to enhance engagement and learning outcomes [71, 72, 84]. Data-driven analyses of a child's speech patterns, progress, and difficulties can also provide therapists, families, and educators with information about the success of interventions and can support decision-making in a child's treatment plan [89]. AI-driven applications can also reduce gaps in access to speech therapy by expanding availability or increasing engagement. For example, Lee et al. developed a virtual AI-based speech therapist application to assist children on demand, guiding them through exercises, simulating correct pronunciation, and providing immediate feedback [57]. Utepbayeva et al. [85] conducted a 5-month deployment study of AI-based speech intervention tools (i.e., Fluency SIS, Articulation Station, Apraxia Farm), showing improved speech performance and acceptance. AI-driven games and activities can offer interactive speech therapy exercises that encourage speech practice through play [10, 29]. Such approaches can be valuable in situations where access to SLPs is limited, as they can offer affordable tools that can be used at home or in educational settings.

Furthermore, the integration of Generative AI, particularly Large Language Models (e.g., ChatGPT), has opened new avenues for speech-language pathology. For example, Suh et al. [79] explored opportunities and challenges for AI-based support for SLPs (e.g., creating personalized therapy content, generating vocabulary lists, translating therapy materials, providing bilingual therapy). Du et al. [36] suggested how text-to-image models (e.g., DALL·E) might replace traditional materials, providing more interactive and versatile learning experiences. Moreover, they discuss opportunities to generate contextually appropriate and culturally competent content to enhance therapy effectiveness.

Amidst growing interest in AI to support SLPs, including that SLPs are already using available AI in their current practice, research overlooks the needs and challenges present when working with CLD children with disabilities. This motivates the importance of examining how AI can support SLPs in providing care to CLD children with disabilities.

## 2.3 Bias in Generative AI and Evaluations

*2.3.1 Bias in Generative AI.* Growing adoption of Generative AI has prompted critical examinations of bias, misrepresentation, and the exclusion of marginalized communities perpetuated by language and image models. Research has found AI rendering bias and negative stereotypes across various identity dimensions, including race [58], ethnicity [11, 34, 58], religion [58, 61], language variations [42, 76], disability [40, 61, 86], and intersectional dimensions [61, 81]. Such research contributes to the broader discourse on AI fairness, particularly concerning ethical implications of AI-generated material across race, ethnicity, religion, disability, and the intersection of identity dimensions [21, 44, 45, 48, 54, 61, 83].

*2.3.2 Evaluations for Identifying and Measuring Bias.* Researchers have developed methods to identify and measure harmful biases in language and image models, often exploring how communities are represented in AI-generated content according to various prompts. Among such methods, automated evaluations use metrics and tools to evaluate model performance [27]. This is commonly used to examine demographic characterizations depicted in language generation. Sheng et al. [75] assessed demographic biases (i.e., race, gender, sexuality) in natural language generation across two language models using pre-trained classifiers. Similarly, Hutchinson et al. used pre-trained classifiers to assess disability bias in language models in regards to toxicity and sentiment [49]. Bianchi et al. conducted a comparison between CLIP [70] embeddings of generated images and a dataset of images with self-identified race and gender [22]. Cho et al. quantified biases in generated images for text prompts describing various professions by computing the distribution of skin tone and gender using an automated classifier and human-evaluations [28]. In contrast to automated methods, human evaluations take a more qualitative and nuanced approach to examining subtleties of individual perception [27]. Gadiraju et al. adopted a qualitative approach to assess disability bias in large language models [40]. Qadri et al. solicited perspectives from community experts through focus groups to assess South Asian cultural bias in text-to-image models [69]. Mack et al. conducted focus groups with people with disabilities to examine disability representation in text-to-image generation, finding reductive archetypes produced by available models [60].

In a literature review, Chang and Wang et al. summarized both automated evaluation criteria and human evaluation criteria [27]. Building on principles of the 3H rule (i.e., Helpfulness, Honesty, and Harmlessness) [17], Chang and Wang et al. expand human evaluation criteria into six categories: accuracy, relevance, fluency, transparency, safety, and human alignment [27]. Our work leverages five of Chang and Wang et al.'s criteria (i.e., accuracy, relevance, transparency, safety, human alignment) to explore participant perceptions on the performance of available AI in supporting their responsive practices.

# 3 Methods

To examine SLP experiences and challenges in providing responsive care to CLD children with disabilities and corresponding opportunities for AI-based support, we conducted semi-structured interviews, followed by a two-part survey study to evaluate available AI in addressing these needs.

## 3.1 Semi-Structured Interviews

*3.1.1 Participants.* We disseminated a brief study description and enrollment survey via email to a participant pool established by the National AI Institute for Exceptional Education [8]. Inclusion criteria included: (1) currently practicing as a SLP, and (2) current or prior experience working with CLD children with disabilities.

Of 15 participants, the majority identify as female, White, and monolingual (see Table 1), consistent with SLP demographics reported by the American Speech-Language-Hearing Association (ASHA) Member and Affiliate Profile [1] and the Profile of ASHA Multilingual Service Providers [5]. Participants had a range of professional expertise, with 8 having over 10 years of experience in their practices and 7 having 9 or fewer years of experience. Many participants (P1, P2, P4, P6-P8, P11, P12, P14, P15) reported prior experience using AI tools for general purposes (e.g., modifying the tone of an email (P7)), and some (P1, P6, P7, P8, P11, P12, P14, P15) reported prior experience using available AI tools to personalize materials (e.g., tailoring stories to align with student reading levels (P11)). Only P6, P12, P14 reported prior experience addressing cultural, linguistic, or disability adaptations (e.g., analyzing Spanish influenced English grammar (P6)). Remaining participants reported no prior engagement with AI tools (P3, P5, P9, P10, P13). Table 1 provides additional self-reported participant information.

Participants had experience working with a diverse range of children characterized by varying racial, ethnic, linguistic, and disability backgrounds. Table 2 provides additional participant-reported information about the demographics of children they reported working with. Because SLP participants come from different practice settings (e.g., schools, hospitals), we generally refer to their clients as "child" or "children". Specific participants sometimes use more precise language (e.g., "student"), in which case we often use that same term in order to mirror participant language.

*3.1.2 Procedure.* We conducted semi-structured interviews with 15 SLPs who have experience working with CLD children with disabilities. Semi-structured interviews each lasted approximately 1 hour, in which we asked participants to:

(1) Discuss their experiences and challenges creating responsive care for CLD children with disabilities, their current technology usage and limitations, and any experience using AI to support their practice.
(2) Discuss their experiences supporting diverse language and dialect development, including tensions in supporting multilingual and multi-dialectal development. Participants also shared how current technology supports and hinders diverse development. We then asked participants how they envision AI supporting the development of language variations.
(3) Discuss their experiences and challenges collaborating with families and caregivers of CLD children with disabilities, including disseminating carryover practice and producing progress reports. We then asked participants how they envision AI supporting them when collaborating with families and caregivers.

All interviews were conducted by the first author and recorded using Zoom. Our university's Institutional Review Board reviewed and approved this research, and participants were compensated with a $30 USD gift card for their time.

*3.1.3 Data Analysis.* We transcribed each interview using Rev, a secure audio transcription service. We analyzed interviews through a codebook thematic analysis approach [25], where we applied a combined deductive-inductive approach to coding using codes from prior research [79]. Authors developed additional inductive codes by reviewing and open-coding 15 interview transcripts, and then discussed and revised as a group among the first three authors to produce an initial codebook to guide coding for all interview transcripts. The first three authors then independently coded subsets of the interview transcripts and discussed codes throughout the coding process to resolve discrepancies, reach a consensus, and organize our findings into high-level themes.

## 3.2 Surveys

Building upon findings from our interviews, the second part of our study used a two-part survey to examine the performance of a commercially available AI model (i.e., ChatGPT-4o) in generating culturally, linguistically, and disability related materials to support SLPs in delivering responsive care. We scoped our research to this widely adopted and publicly available model because many participants with prior AI experience (SP1-SP4, SP6-SP8, SP10, SP12, SP13) had previously used a version of it. We employed a survey to be a more efficient with SLP time and allow them to individually reflect on their experiences with AI. Participants were compensated with a $20 USD Amazon gift card for their time.

*3.2.1 Participants.* We recruited 13 participants, including 9 from our initial interviewees and 4 from a broader participant pool established by the National AI Institute for Exceptional Education [8]. All 13 participants completed the initial survey (Section 3.2.2) and 8 participants (SP1-SP8) completed the follow up evaluation survey (Section 3.2.3). The majority of participants identified as female, White, and monolingual. Many participants (SP1-SP8, SP10, SP12, SP13) reported prior experience using AI tools, while the remaining indicated no prior engagement with AI (SP5, SP9, SP11) (see Table 3).

*3.2.2 Survey Protocol for AI Prompt Crafting and Material Generation.* Using a survey, we collected three prompts from each participant to then generate AI-based materials relevant to their current practice with CLD children with disabilities. This survey had five parts: (1) we introduced concepts of generative AI, including how they are trained and used; (2) we provided example images illustrating how inputs and outputs function when prompting AI systems; (3) informed by challenges and opportunities that SLPs identified in earlier interviews, we provided examples of prompt areas for participants to explore for their practice (e.g., Write a story that uses the 'th' phoneme about a 6 year old Chinese girl celebrating Chinese New Year with her family, create an image of a South Asian princess who is deaf); (4) we asked participants to create two text-based prompts and to provide a brief description of how these prompts are relevant to

**Table 1: Overview of self-reported participant information, including race, gender, spoken language(s), practice setting, and years of professional experience. Participants were additionally asked to report their prior experience using AI in their practice, which is classified into four categories: (G) General Purpose, (P) Personalization, (C) CLD Adaptations, and (N) No Use.**

| ID | Reported Race | Reported Gender | Reported Language | Practice Setting | Years of Experience | AI Use (G/P/C/N) |
|---|---|---|---|---|---|---|
| P1 | White / Caucasian | Female | English | Public School | 1.5 | G, P |
| P2 | White / Caucasian | Female | English | Public School | 13 | G |
| P3 | White / Caucasian | Female | English | Early Intervention | 23 | N |
| P4 | White / Caucasian | Female | English, Spanish | Early Intervention, Pediatric Hospital | 8 | G |
| P5 | White / Caucasian | Female | English | Public School | 10 | N |
| P6 | White / Caucasian | Female | English, Spanish | Public School | 4 | G, P, C |
| P7 | White / Caucasian | Female | English, ASL | Public School | 14 | G, P |
| P8 | White / Caucasian | Female | English | Public School | 5 | G, P |
| P9 | White / Caucasian | Female | English | Private School | 24 | N |
| P10 | White / Caucasian | Female | English | Public School | 2 | N |
| P11 | White / Caucasian | Female | English | Public School | 13 | G, P |
| P12 | Asian | Female | English | Early Intervention, Pediatric Hospital | 20 | G, P, C |
| P13 | White / Caucasian | Female | English | Public School | 27 | N |
| P14 | White / Caucasian | Female | English | Public School | 5 | G, P, C |
| P15 | Hispanic or Latino or Spanish Origin, White / Caucasian | Female | English, Spanish | Public School | 3 | G, P |

**Table 2: Overview of self-reported participant caseload demographics. Participants reported working with children across a range of racial, ethic, linguistic, and disability identities.**

| Category | Reported Children Demographics |
|---|---|
| Race / Ethnicity | Asian, Black or African American, Hispanic or Latino or Spanish Origin, Indigenous American or Alaskan Native, Multiracial, Native Hawaiian or Other Pacific Islander, White / Caucasian |
| Language | American Sign Language, Arabic, English, Spanish, Korean, Mandarin, Ukrainian, Gujarati, Hindi, French, Bulgarian, Urdu, Lao, Berber, Cantonese, Vietnamese, Amharic, Kazakh, Oromo, Marshallese, Tagalog, Portuguese |
| Disability | Apraxia, Articulation Disorders, Autism, Deaf or Hard-of-Hearing, Developmental Disabilities, Dyslexia, Central Auditory Processing Disorder, Cerebral Palsy, Developmental Language Disorders, Fluency, Language Learning Disability, Learning Disabilities, Motor Speech Related, Phonological Processing Disorder, Receptive-Expressive Language Disorder, Selective Mutism, Social Communication Disabilities, Speech Disabilities, Stuttering |

their practice; and (5) we asked participants to then also create one image-based prompt and to provide a brief description of how this prompt is relevant to their practice. We asked participants to create prompts that were based on (1) real-world materials they currently used and would like to adapt for culture, language, and disability, or (2) materials they did not have access to that were relevant to their specific caseload and would like customized to support their practice with CLD children with disabilities.

Although prompting can be iterative, we developed a method in which the research team generated participant materials using participant-provided prompts to: (1) minimize SLP time and burden of participation, (2) ensure the most advanced version of ChatGPT (i.e., ChatGPT-4o), which required a paid subscription at the time of the study, and (3) focus on eliciting participant evaluations of AI-generated material results, rather than assessing the learnability of the AI system. Section 6 then notes an opportunity for future work further examining learnability of such interactions for SLPs.

*3.2.3 Survey Protocol for Generated Material Evaluation.* We employed ChatGPT-4o to generate materials based on participant prompts and shared the initial results with participants to evaluate. Participants completed a 30-minute, four-part survey assessing the utility and effectiveness of their AI-generated materials. First, participants shared their overall impressions, perceived effectiveness of materials, perceived areas of improvement for the output, and shifts in their perception of how AI can support their practice. The subsequent three parts asked participants to qualitatively evaluate their three AI-generated materials based on five established Large Language Model evaluation criteria: alignment, accuracy, relevance, safety, and transparency [27]. We adapted the definitions of the five criteria to align with the context of this study as follows: *'Alignment'* assessed the degree to which generated material aligns with participant expectations as an SLP; *'Accuracy'* assessed precision and correctness of generated materials compared to participant prompts; *'Relevance'*

**Table 3: Overview of self-reported survey participant information, including race, gender, spoken language(s), years of professional experience, and any prior experience using AI in their practice. For survey participants who also took part in interviews, we additionally indicate their interview participant ID.**

| ID | Reported Race | Reported Gender | Reported Language | Years of Experience | AI Use | Interview Participant ID |
|---|---|---|---|---|---|---|
| SP1 | White / Caucasian | Female | English, Spanish | 4 | Yes | P6 |
| SP2 | White / Caucasian | Female | English, Spanish | 8 | Yes | P4 |
| SP3 | White / Caucasian | Female | English | 1.5 | Yes | P1 |
| SP4 | White/ Caucasian | Female | English | 5 | Yes | P14 |
| SP5 | White / Caucasian | Female | English | 27 | No | P13 |
| SP6 | White / Caucasian | Female | English | 12 | Yes | - |
| SP7 | White / Caucasian | Female | English | 8 | Yes | - |
| SP8 | White / Caucasian | Female | English | 5 | Yes | P8 |
| SP9 | White / Caucasian | Female | English | 34 | No | - |
| SP10 | Hispanic or Latino or Spanish Origin, White / Caucasian | Female | English, Spanish | 3 | Yes | P15 |
| SP11 | White / Caucasian | Female | English | 2 | No | P10 |
| SP12 | White / Caucasian | Female | English | 11 | Yes | - |
| SP13 | White / Caucasian | Female | English | 13 | Yes | P2 |

assessed appropriateness and how well generated materials addressed participant prompts; *'Safety'* assessed potential harm or unintended consequences that may arise from generated material; and *'Explainability and Transparency'* assessed how well the AI informed participants of its decision-making process in generating the material, including references for each of the AI's outputs. Participants additionally described any modifications they would make to their prompts to enhance their results, and any biases they detected in their AI-generated materials.

*3.2.4 Data Analysis.* We adopted a codebook thematic analysis approach [25], starting with the first author closely reviewing and open-coding each survey response. The first three authors discussed and revised the codes as a group to produce an initial codebook. The first author then refined the codes and applied the codebook to all of the survey data. Complementary to participant evaluation of their generated materials, and to further address RQ3, the research team conducted additional analysis on prompts and the resulting generated materials. This analysis aimed to identify and evaluate the behavior of the AI system, any additional bias it perpetuated in generated materials, and potential impacts in SLP delivery of care. We assessed for cultural bias, linguistic bias, and disability bias, informed by prior literature exploring bias in language and image models [40, 60], AI fairness [21, 45, 54, 61], and critical racial equity, disability, and intersectional frameworks [19, 64, 82].

## 3.3 Positionality

Our analysis and writing are informed by our experiences and identities. Our research team is composed of scholars with and without disabilities, graduate students, a postdoctoral researcher, and senior academic faculty. Our team includes individuals who identify as Black American, Korean, Nepalese, White American, and individuals who are monolingual (i.e., English) and multilingual (i.e., Hindi, Korean, Nepali, Newari). Two of the authors have extensive experience studying accessibility, one author has extensive experience studying technology support for health, education, and families, and one author has clinical experience in speech-language pathology. One of the authors is also a parent of two children with learning differences. We acknowledge that our scope of culture, language, disability, and related biases is contextualized to the U.S. based on our collective positionality and experiences.

# 4 Results

We begin with findings from our semi-structured interviews, including (1) challenges participants described encountering in their practices, and (2) participant descriptions of desired AI-based support in response to these challenges. We then report participant reactions to AI-generated materials addressing their described challenges and desires.

## 4.1 SLP Challenges in Responsive Care

*4.1.1 Unrepresentative Material.* Participants emphasized fostering practices that celebrate the diverse backgrounds of CLD children with disabilities. Despite commitment and efforts to provide responsive care, SLPs reported insufficient culturally, linguistically, and disability representative text-based and image-based materials to support CLD children.

Participants commonly employed low-tech text-based materials (e.g., storybooks, worksheets, flashcards) to foster development of speech and language skills. However, participants reported challenges *"finding materials that feel like they represent our students"* (P10), highlighting the disconnect between available resources and needs of CLD children with disabilities. P7 echoed this sentiment, describing how their available text-based resources lack diversity: *"there was no diversity in there at all."* P5 voiced similar dissatisfaction with current high-tech resources providing text-based materials, stating, *"I haven't been super impressed with a lot of online speech therapy materials."*

Participants also used a combination of low-tech and high-tech image-based materials (e.g., boom cards, picture books, coloring sheets, sequencing cards). Much like their experiences with text-based materials, participants frequently encountered challenges finding representative materials. P2 recalled how this lack of representation negatively impacted children's engagement: *"We did a superhero activity where it was like coloring sheets, and one girl was like, 'but none of these have curly hair'."* P13 described online materials often defaulting to White and Euro-centric representation, *"In the picture symbols or the actual photographs of kiddos doing things [...] it was just all White kids."* P11 further critiqued materials that included diversity in skin tone but failed to go beyond this singular form of representation, remarking, *"they do a good job of having different skin tones within the images, but outside of that, they're pretty typical American images that don't allow for a lot of diversity."*

Participants described strategies they employed to navigate this challenge, but such workarounds often came with their own costs. For example, P1 described creating their own materials, which resulted in extra effort due to this disproportionality: *"there's not the same ready-to-print and go material, so you have to make it from scratch, which is a big challenge."* P2 and P13 described requesting families to provide materials for practice (e.g., books, pictures), but this approach placed additional burdens on families.

Some participants (P1, P6, P7, P8, P11, P12, P14, P15) reported using available AI tools (e.g., ChatGPT, DALL-E, MagicSchoolAI) to personalize materials (e.g., generate images, create humorous stories, craft social stories to help children anticipate situations, generate text with specific phonemes, simplify text to specific reading levels). Such participants shared positive sentiments regarding the usefulness of these tools. However, only three participants (P6, P12, P14) reported using AI to create materials tailored to CLD children with disabilities. When attempting to generate an image for a hard-of-hearing child, P12 shared, *"I tried to get a teddy bear with a bone conduction hearing aid [...] but it was not giving me a bone conduction [aid], and then it wasn't in the right place."* P14 observed that AI defaulted to White-presenting representations, *"You can definitely see where there's biases. If you're working in an image-generating tool [...] you may get only people with one color skin [...] There's a lot of assumptions that are made."*

*4.1.2 Unreliable and Inefficient Machine Translation.* Participants expressed concerns about reliability and accuracy of machine translation (e.g., Google Translate) for translating assessments and therapy material. Concerns were largely attributed to participant inability to verify translations, due to lacking proficiency in the target languages. This lack of linguistic expertise made participants apprehensive of potential inaccuracies in automated translation. For example, P10 expressed mistrust of Google Translate for therapy materials: *"I can't trust that."* P1 expressed similar concerns, *"Is it really conveying what I want it to?"* Frustration was echoed by P6, who described Google Translate as *"clunky"* and described colleagues excusing students from assignments or expecting students to bear the cognitive load of both translating material and completing assignments.

Participants additionally highlighted challenges translating critical documents (e.g., progress reports, IEPs, home practice activities), particularly when working with multilingual families not fluent in English (P3, P7, P9, P11, P13). They reported translation resources beyond Google Translate were often limited to Spanish, neglecting the linguistic diversity of the children and families they serve. P15 described this with their current IEP system, *"it translates parent rights and procedural safeguards in Spanish, but it doesn't translate any other documents into any language other than English."* Participants further described challenges translating assessment materials, hindering accurate evaluation and risking misdiagnosis or inappropriate intervention. For example, P13 recalled Google Translate failing to deliver accurate translations during an assessment conducted alongside an interpreter, stating, *"We are finding that it's not translating exactly, and it's not communicating as accurately as it needs to."*

Translating materials for multilingual children and their families presented time challenges for some SLPs, adding to already high workloads. P14 described cumbersome translation of materials from Ed Discussion, *"Ed doesn't offer the translation. I have to take what I write and put it in a Google Doc or MagicSchool to do the proper translation, then copy it back into Ed. [...] So it takes a lot of time."* P10 expressed frustration with their district's delayed translation processes, impacting the delivery of progress reports to families: *"In this particular district, it's been months and months and months before even progress notes have been translated into Spanish. So we're pretty backed up [...] It's obscene."*

Participants described collaborating with families, interpreters, and English Language Specialists as strategies for navigating language barriers and unreliable machine translation, but these workarounds resulted in additional costs. P11 recalled instances where translation alternatives were unavailable, leading to challenges in communication and additional work for family members: *"We unfortunately do not have a Spanish version of that form. And so we've kind of relied on the older brother to step through that report with her."* Other participants described using interpreters, but time, availability, and financial costs were additional burdens. For example, P4 shared, *"I have an Uzbek family right now, and Uzbek interpreters are really hard to find."* P6 shared, *"It's really frustrating because interpreters cost a lot of money. [At the hospital] we have a virtual video interpreter, but it's not easily accessible throughout the hospital [because of] wifi. So that's also a barrier."*

*4.1.3 Lack of Materials and Resources for Supporting Language Variations.* Participants emphasized preserving and honoring children's language variations (e.g., home languages and dialects). However, linguistic homogeneity among many participants presented challenges. Participants reported a scarcity of available resources on language development in less commonly represented languages. P15 noted that while resources for Spanish are available, many languages remain under-resourced. P9 described difficulties of conducting effective assessments without being able to evaluate a child in both English and their home language, *"the most difficult thing is the assessment because I know ideally it would be best to be able to assess them in both languages, but I'm not able to because I'm not bilingual."* Participants described *"issues with over- and under-identification"* (P6) of children with speech or language difficulties due to language variations. [...] Honestly, this kid didn't have a language disorder. He was bilingual and had some behavior issues, and I don't know who the SLP was who originally assessed him, but she probably shouldn't have." These challenges also extend beyond assessment to therapy. P11 expressed concerns in therapy for a potential child for whom they were not able to conduct an assessment because of language barriers, *"that would be an issue that I would have to figure out as we went, if he were to qualify for services, how would I work on the process of doing English second language services?"*

Participants discussed strategies for distinguishing between language variations and speech and language difficulties (e.g., ASHA website, Diagnostic Evaluation for Language Variation (DELV) resources, informal consultations with other professionals), but also described these as *"time-consuming"* and *"often difficult."* P5 shared their experience evaluating a student who speaks Marshallese: *"[evaluating] involves trying to find the phonological rules of Marshallese and the sound inventory of Marshallese. I need to ensure that I'm not marking items on the test incorrectly and that I'm avoiding treating accent modification as a disorder. This process is very time-consuming."* P5 also described *"picking through online"* content, often requiring *"double-checking with the teacher or the parent,"* which further increased time costs.

Challenges extended into complexities of dialectal variations. Despite resources like DELV, participants reported difficulty differentiating between English and non-English dialects versus genuine speech and language difficulties. P9 highlighted inadequacy of current resources: *"Sometimes the directions are a little vague... so it's hard to know how much of a kid's dialect is actually their dialect and how much of it's [related to language difficulties]."* P13 stated, *"the knowledge about different dialectical differences is not as readily accessible as I would like it to be."* Time-intensive parsing of dialectal variations further exacerbated these challenges, as participants often found this detracted from other important aspects of sessions: *"I often find myself spending hours and hours analyzing responses to determine if the productions are consistent with patterns of Spanish-Influenced English, African American Vernacular English, or Southern English"* (P13).

Participants summarized challenges in terms of disparities in quality and equity of care provided to CLD children with disabilities. P10 stated, *"I'm a little worried about the difference between how we're serving our white population and our non-white population and the equity of it"*, that service gaps represent *"a problem bigger than"* any single practitioner can address, and acknowledged *"it's a systemic thing"* rooted in racist and ableist structures and ideologies.

## 4.2 SLP Identified Areas for AI-Based Support in Responsive Care

*4.2.1 Personalized Representative Material Generation.* Participants expressed enthusiasm for AI-based tools to quickly generate materials tailored to the cultural, linguistic, and disability backgrounds of CLD children with disabilities. They indicated a strong desire for customized materials reflecting children's needs and identity dimensions, but highlighted that time constraints and competing responsibilities make this difficult to accomplish [79]. P11 described generating personalized materials relevant to children's families and background: *"target words and target images that are appropriate to their family home and the things that they would see at home, I think would be very, very helpful."* P2 expressed excitement about AI assistance for generating customized books: *"I can pick who the characters are and what they look like and their dialect, their accent."* P2 emphasized potential impact of such technology in allowing their time to be better focused on students, *"I just think of that as being really, really powerful to then give me time as a human to be able to work with those students."* P4 echoed this, noting AI could streamline administrative tasks, *"If there was some way that some computer could generate [multilingual] word lists with cute pictures, that would make my life a lot easier."*

*4.2.2 Accurate and Efficient Translation for Materials and Family Collaboration.* Participants expressed a strong desire for AI-based tools to provide accurate and efficient translation of materials and for collaboration with families in communication and documents (e.g., progress reports, IEPs). P5 envisioned AI eliminating the need to move between tools, wanting *"a way to easily check in without having to go to Google Translate and then copy and paste it into an email."* P6 explained, *"I would just be able to type what I want in an email, and then [the AI] would type the Spanish right below it."* In addition to streamlining communication, participants envisioned translation accurately capturing intent, with P7 wanting, *"translation that accounts for the nuances of language."* P1 imagined accurate translation reducing the reliance on interpreters, thus minimizing logistical complications: *"If I could just say it, have it written down, and know it was accurately translated, it would be so much easier to send parents any information or progress reports."*

*4.2.3 Automated Speech Recognition to Support Language Variations.* Participants envisioned AI assisting in differentiating between language variations, including both English and non-English dialectal variations, versus genuine speech and language difficulties. They imagined automated speech recognition enhancing accuracy in detection to reduce misidentification and time required. P13 elaborated, *"After administering a receptive/expressive language assessment and/or gathering an informal language sample, I would like a way to enter the responses and determine if the utterances were consistent with language differences versus disordered language [...] I believe a tool like that would increase scoring accuracy and analysis accuracy. It would also decrease over-identification."* P11 echoed this desire for AI-based support in accurate identification, *"I think the biggest breakthrough would be an assessment tool that would pick up on those different pronunciations and those different language forms that are appropriate in different dialects and different languages."*

*4.2.4 Responsive Planner Generation.* Participants expressed a desire for AI-based tools to facilitate cultural, linguistic, and disability-responsive planning at a larger scale. P6 envisioned a tool that generates a tailored *"plan of resources"* and acts as a *"roadmap"* to provide foundational support based on demographic input, aligning with a child's English proficiency, cultural background, and disability. They shared, *"staff members aren't sure even quite where to start. It's not that they don't want to help. I just think that they're overwhelmed with the amount of work that they have. So, if they had a little bit more of a guide to help them start out, maybe that would be helpful and useful."* Similarly, P8 imagined a tool that could provide personalized recommendations for child-specific goals, selecting appropriate techniques, and monitoring progress to reduce the workload of developing individualized plans: *"basically do all of the things that I think about with my brain, and then it spits it out to me."*

## 4.3 SLP Reactions to AI-Based Adapted Materials

In the second part of this study, we invited SLPs to complete a two-part survey exploring ChatGPT-4o in addressing reported challenges (Section 4.1) and opportunities (Section 4.2) for AI-based support (i.e., RQ3). Prompts generated by survey participants included story generation, articulation word lists, oral narrative analysis, progress report generation, text simplification, translation, differentiating between dialectal difference and difficulty, and image generation. A full list of participant prompts is provided in Supplementary Materials.

*4.3.1 AI as a Good Start.* Survey participants described AI-generated text-based materials as providing a *"good starting place"* (SP1), indicating outputs were not flawless but offered a valuable foundation for supporting their practice with CLD children with disabilities. Survey participants were highly satisfied when AI-generated materials, many of which were stories, aligned with specifics of their prompt and were accurate, relevant, and safe. In evaluating their generated story to deliver social learning lessons on friendship and social engagement for students with Down Syndrome, SP4 wrote, *"The output provided appeared accurate and the content was well organized and detailed, reflecting a school dance scenario–despite being delivered in a different format than requested. The output was relevant and delivered practical insights and suggestions."* SP4 also explained how this material provided tips that would encourage extended discussion and learning. Based on SP8's prompt for a social story *"for a student in first grade who is intellectually disabled and whose primary language is Spanish about starting at a new school,"* SP8 believed the material would help them support an incoming student in acclimating, writing, *"The social story for the student whose primary language is Spanish was perfect."* Several survey participants (SP1, SP2, SP5, SP6) expressed positive shifts in their perception of AI-based support: *"my perception of how AI could support my practice working with CLD children positively changed"* (SP5).

Despite positive reception to AI as offering an effective foundation, survey participants expressed reservations rooted in outputs lacking depth and specificity necessary for their practice. When evaluating a prompt asking for *"a progress report for an autistic child who is making progress towards his goals"*, SP7 observed, *"There was too much generic information."* SP1 expressed optimism about AI's potential while highlighting the need for refinement: *"I was reasonably impressed with the materials at first glance, but there are definitely tweaks that need to be made."*

When generated image-based materials closely matched specificities of a prompt, participants found them accurate, relevant, and aligned with their expectations. In response to their image intended for a social story to help children in their predominately military family caseload prepare for a move to a new duty station, SP5 shared, *"I was impressed with how aligned the image was to my prompt."* Similarly, SP6 shared their image was *"relevant for a [therapy] session, and better than a google image in [ethnic] representation of [a] child."* In contrast, some participants (SP1, SP2, SP3) perceived image-based AI-generated materials as less effective in meeting their needs and expectations. SP3 commented, *"I did not find the image generation to be helpful. I could see how the style of the images could be off-putting."* Similarly, SP2 noted, *"The picture has an oddness to it that may be challenging to implement clinically."*

*4.3.2 Contextually Unaware, Insensitive, and Biased AI.* Survey participants expressed concerns about AI's lack of contextual awareness, particularly in relation to speech-language pathology practices and the cultural, linguistic, and disability identities of children. For example, SP7 highlighted AI's lack of awareness of SLP standards, noting inefficiencies in generated progress reports: *"public school progress reports should be shorter"* and *"AI should make reports easier. Reading all that information would take more time than if I had just written it myself."* SP7 further emphasized gaps in AI's understanding of SLP technical jargon in progress reports and expressed a need for AI to better grasp domain-specific knowledge to help families better understand their child's progress: *"I wish ChatGPT knew more SLP jargon and could explain it to parents in simple terms."*

Survey participants pointed to instances where the AI failed to demonstrate awareness and understanding of cultural, linguistic, and disability identities. SP4 captured this in regards to cultural identities, commenting, *"Overall, I found the outputs to be well-intended responses to my prompt. However, I can see where the outputs could be limited by the lack of cultural intelligence of ChatGPT."* In evaluating their prompt for a list of children's books written by Asian American and Pacific Islander (AAPI) authors, SP5 wrote, *"There were potential biases in the AAPI ethnic groups and countries represented in the book list. While there were a few books representing the people from the Hmong culture and the countries of Taiwan, Philippines, and Korea, the majority of the books referenced China."* Although the AI defaulted to dominant cultural narratives, SP5 said it provided valuable recommendations, *"Based on the book reviews that I read, the books were regarded as culturally accurate, age appropriate, and culturally sensitive"* (SP5). Survey participants expressed instances where AI's lack of diverse linguistic understanding intersected with its limited knowledge of SLP contexts. For example, SP4 highlighted the AI's difficulty in distinguishing between English phonemes and those from other languages, *"ChatGPT could have done a better job understanding that the way the letter j/phoneme j is pronounced in the English language and Spanish language are different."* SP1 reflected on errors in their text-based prompt intended to assist with analyzing speech sound productions to differentiate between dialectal variations and speech and language difficulties. They stated, *"I did notice some errors [...] For example, s-blends occur across syllables in Spanish, but not in the beginning of words in English. So, this would be an example of Spanish Influenced English."* SP1 highlighted that the AI's lack of linguistic sensitivity could *"result in a SLP overidentifying a student."* They elaborated on harmful repercussions that could arise if such outputs were implemented in practice, explaining *"[This] may pull the student* *out of class when it is not necessary."* However, SP1 said the AI was *"able to correctly identify the other differences vs. disorders correctly.*

Survey participants additionally highlighted challenges in disability representation. SP2 discussed results of their prompt: *"Create an image of children playing - some should be on a playground and some should be playing with toys. The children should be racially diverse with at least half being Black/African American."*, critiquing the lack of disability representation in the AI-generated image. They noted, *"All of the children are [non-disabled]."* Similarly, SP9 critiqued AI assumptions around assistive technology in their generated story intended to help a new student who is visually impaired acclimate to their class, *"The story about the student with visual impairment wasn't quite right because he does not use a cane."* SP3 evaluated results for their prompt intended to encourage families to incorporate AAC devices into routine activities: *"Generate a series of realistic images including African American families and Latin American families where adults are modeling use of augmentative and alternative communication devices for children during mealtimes, while watching television, and at bedtimes,"* saying, *"Most of the pictures didn't show modeling of the devices and none of the individuals in them were actually looking at the devices."* SP4 discussed ableist assumptions in their AI-generated image, stating, *"There is a bias that the AAC would have only pictures and not text, perhaps hinting that students with Down Syndrome do not develop literacy skills such as for reading and writing and navigating written language."*

*4.3.3 SLPs as Prompt Engineers.* Survey participants often attributed shortcomings of AI responses to prompts, saying more *"carefully crafted"* (SP5) and *"intentional prompts"* (SP4) might achieve better results. As such, survey participants expressed a desire to re-craft prompts in pursuit of more desirable outcomes that better represent children in their caseloads. For example, SP4 shared this desire to adjust the prompt for a story-generation result that better captured their intent, *"I see that I should have provided more in depth and specific details in the prompt language to provide tighter constraints and arrive at the output I was seeking."* For a story prompt intended to model the production of the English /j/ phoneme for a Spanish speaking student learning English, SP4 shared a desire to adjust their prompt because they were *"not specific enough,"* thus *"the text could confuse the learner and undermine the intention of the lesson/activity."*

*4.3.4 Explainability, Transparency, and Cross-Checking.* Survey participants expressed concerns about a lack of explainability and transparency, which led to reservations about reliability of generated material. Only one participant (SP5) received materials that included any form of explanations (i.e., cited sources), which they appreciated because the accurate links and citations facilitated quick verification. However, the majority of participants indicated generated materials lacked any explanation of how AI derived the cultural, linguistic, and disability content. This absence reduced confidence and created significant challenges in assessing accuracy and appropriateness of AI-generated outputs. As SP3 remarked, *"I have no idea what the decisions made were based on."* SP6 added, *"There's no information about how the algorithm makes its choices."* SP2 shared that generated material they intended to use to improve a student's home language, *"Identify 10 words in Uzbek that a 6 year old child may use at home,"* was highly aligned with their expectations, but it was, *"hard to judge the accuracy as I don't speak Uzbek."*

Survey participants highlighted the need for explainability and transparency in AI-generated materials to validate accuracy and contextual appropriateness, particularly regarding cultural, linguistic, and disability relevance. SP5 provided the prompt, *"Create a list of activities considered inappropriate for individuals who practice the Jehovah's Witness religion,"* to better understand beliefs of children in their caseload and inform session planning. Evaluating the result, SP5 emphasized the importance of explainability and the risks of its absence, stating, *"Because the information provided did not include citations or sources, the potential for bias is strong. Without citing evidence from reputable sources, the information generated could continue negative stereotyping about Jehovah's Witnesses."* SP3 emphasized additional research to ensure generated materials were culturally appropriate, *"I think the story generated would definitely work, but I would need to do more research to verify."* SP6 noted an absence of sources made it more difficult to *"fact check, so [it] could be entirely fictional."* AI opacity not only complicated fact-checking and increased the risk of perpetuating biases, but also undermined participant trust. SP6 shared their reservations, *"I don't feel like I know where the words came from, and not speaking Spanish I'm just trusting they are what they say they are."* SP3 described distrust of AI-generated materials and the need for cross-checking due to hallucinations [12, 50], explaining, *"I have no idea what the decisions were made based on and wouldn't trust AI-generated citations anyway since it's been documented that ChatGPT will make them up."*

## 4.4 Range of Care in AI Prompting

We observed a range of care in how survey participants crafted AI prompts, likely influenced by variations in participant AI literacy, even among those with prior AI experience. This ranged from prompts not including any cultural, linguistic, or disability nuances to prompts that meticulously included all three identity categories. Yet, even when prompts included one or more identity dimensions, AI-generated outputs still exhibited biases, both within specified categories and in omitted identity dimensions. We report these biases, arising from the interplay between variations in participant prompt crafting and AI's lack of contextual awareness, to further evaluate how they affect the AI's performance in addressing challenges and opportunities in supporting responsive practices (i.e., RQ3).

*4.4.1 Cultural Bias.* Cultural bias in AI-generated image-based materials was observed through a lack of diverse cultural inclusion and sensitivity, manifesting in two ways: defaulting to (1) White-presenting characters, and (2) Euro-centric aesthetics (see **??**). When race and ethnicity were not specified, AI consistently generated images featuring White-presenting characters, revealing biases that position Whiteness as the default [69, 82]. When non-White racial identities were specified, AI-generated images did include corresponding characters, but almost always alongside White-presenting characters, again prioritizing White-presenting characters as a standard. As above, this inclusivity was generally not reciprocated when racial or ethnic identities were not specified. Additionally, cultural bias was exhibited through strong Euro-centric representations, prioritizing European and Western norms, standards, and perspectives in settings, clothing, and other contextual elements. In text-based generated materials, cultural bias manifested through AI tendency to perpetuate reductive stereotypes, particularly in story generation. When prompted to generate stories with culturally significant details, such as *"Generate a 3rd grade level story about a Latin American family's trip to the beach"* (SP3) or *"Write a story that uses the 'j' phoneme about a 15-year-old Spanish-speaking girl celebrating her Quinceañera with her family"* (SP4), AI frequently resorted to narratives that simplified to cultural stereotypes and tokens, including in names (e.g., Javier, Juan, Maria), foods (e.g., empanadas, carne asada, enchiladas, arepas), and sports (e.g., soccer). Such outputs reinforce monolithic narratives that fail to capture diversity within cultural experiences [69, 82].

*4.4.2 Disability Bias.* Disability bias was frequently observed in AI-generated image-based materials, manifesting in five ways: (1) defaulting to non-disabled characters, (2) reinforcement of disability tropes, (3) defaulting to specific types of assistive technology, (4) incorrect depictions of assistive technology usage, and (5) addition of extraneous hardware to assistive technology (see ??). Reinforcement of disability tropes was evident through portrayal of people with disabilities as appearing to be sad. Although it is crucial to acknowledge the diversity of disabled experiences (e.g., empowerment, joy, grief), the overemphasis of a singular narrative can be harmful and risk reducing multifaceted experiences. Disability bias also manifested through overrepresentation of certain assistive technology, such as wheelchairs and high-tech augmentative and alternative communication (AAC) devices. For example, the prompt *"create an image of a physically disabled white girl asking for help in a bodega"* (SP9) generated an image featuring a wheelchair, thus perpetuating a narrow view of disabilities and tools used for access [40, 60]. Additionally, there was an overwhelming presence of inaccurate depiction of assistive technology. In several instances, AI rendered AAC devices without text and backwards, demonstrating improper use, or showing non-use despite explicit prompting for use. Although shifting focus from assistive technology to the individual can be valuable [60], omitting use despite explicit prompting undermines the image's intended purpose. Lastly, addition of extraneous hardware in AI-generated images of assistive technology further reflects lack of awareness and understanding of disability. For example, in prompts specifying use of AAC devices, AI-generated images frequently depicted additional and unnecessary equipment such as headpieces and harnesses that are not typically associated with AAC.

Similarly, disability bias manifested in text-based generated materials through (1) defaulting to non-disabled characters, (2) the omission of disabilities-specific nuances, (3) defaulting to specific assisitve technology, and (4) ableist language (e.g., special). In story generation prompts that lacked explicit mention of disability, AI defaulted to non-disabled characters. When provided the prompt: *"Write a story about a student with visual impairment and minimal verbal speech for his general education peers about how to include and support him"* (SP8), AI defaulted to including a cane. Ableist language was also present in AI-generated text-based material, as when the prompt *"Write a social story about a Colombian boy that uses hearing aids. The story should be about starting in a new classroom and answering questions from peers about his hearing aids"* (SP13) resulted in AI using the term *"special"* as a descriptor for the character and the hearing aids (i.e., a term viewed as patronizing and offensive within many disability communities [7]).


**Figure 1: SP1's image-based material generated from the prompt: *"Create an image of vocabulary items and key terms that are labeled in Spanish and English for the American Revolution."* Although it displays images capturing some terms in both Spanish and English, it presents linguistic bias through often prioritizing English words while Spanish words are more likely to be incorrect.**

*4.4.3 Linguistic Bias.* Linguistic bias, though less prevalent than disability and cultural biases, was observed through prioritization of English, specifically Standard American English, over alternative languages specified in prompts. For example, the prompt *"Create an image of vocabulary items and key terms that are labeled in Spanish and English for the American Revolution"* (SP1) yielded AI-generated output exhibiting clear disparities in language representation. English words were more prominent, whereas Spanish words were less common, illegible, and sometimes not actual words in the Spanish language (see Figure 1).

## 5 Discussion

Our findings reveal that although AI holds great potential to support SLPs in providing equitable and responsive care to CLD children with disabilities, these systems remain a conduit for the perpetuation and amplification of systemic cultural, linguistic, and disability biases (see Section 4.3.2 and Section 4.4). Such biases in AI outputs risk perpetuating the historical continuum of

**Table 4: Participant prompts, AI-generated image-based materials, and identified cultural bias in images.**

| Prompt | Generated Image | Identified Bias |
|---|---|---|
| "A middle school aged boy with Down Syndrome using an ipad to verbally communicate" (SP4) |  | Defaulting to White-presenting character |
| "An autistic child experiencing sensory difficulties while his classmates and teachers of different ethnic backgrounds look at him with concern. Include a social story for the classmates to understand the difficulties some autistic children experience in a classroom setting" (SP7) |  | Defaulting to White-presenting character |
| "A series of realistic images including African American families and Latin American families where adults are modeling use of augmentative and alternative communication devices for children during mealtimes, while watching television, and at bedtimes" (SP3) |  | Defaulting to White-presenting characters; Defaulting to Euro-centric representations |

racism and ableism in speech-language pathology, compromising equitable care and the representation of the diverse identities of the children served. Furthermore, our findings demonstrate available AI often lacks nuanced understanding of SLP contexts of use, producing text-based and image-based outputs that fail to meet the standards and expectations of their practice and inadequately address participant reported challenges (see Section 4.1) and desires for AI-based support (see Section 4.2).

## 5.1 Implications of Range of Care in AI Prompting

Participants demonstrated a range of care in how they crafted AI prompts (Section 4.4), with some creating prompts that omitted one or more cultural, linguistic, or disability-related identity dimensions and some carefully incorporating all three identity categories. The range of care observed in our study likely reflects a broader trend among SLPs who are using or considering personalized AI-generated materials in their practice, demonstrating implications for the quality and equity of care for CLD children with disabilities. SLPs with greater sensitivity, and perhaps greater AI experience, are more likely to create prompts that effectively capture cultural, linguistic, and disability-specific nuances, resulting in AI-generated materials that better meet child needs. This was largely the case for participants in our survey study, where the nature of the study also encouraged considering these identity dimensions when creating prompts. However, SLPs may not always think to incorporate these contextual nuances in prompts, as also evidenced by participants in our study. This may be influenced by SLPs being unaware of AI's limitations and the need for specificity, lacking understanding of biases embedded in AI, or feeling it was unnecessary to specify these dimensions. To further explore how AI performs when the range of care in prompt construction is minimal, we systematically removed cultural, linguistic, and disability-specific contexts from SLP prompts. Without these contextual cues, the AI often erased these identities, leading to an overwhelming trend of White, Euro-centric, and non-disabled representation in re-generated materials (see Table 6). Given AI's limitations in contextual awareness (see Section 4.3.2), when SLPs are not aware or otherwise do not take the extra time and effort to craft specific prompts for CLD children with disabilities, they risk perpetuating inequities in care, reinforcing systemic biases, and widening of the gap in outcomes between CLD children with disabilities and their peers.

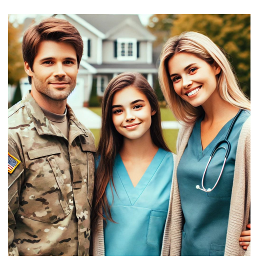
**Table 5: Participant prompts, AI-generated image-based materials, and identified disability bias in images.**

| Prompt | Generated Image | Identified Bias |
|---|---|---|
| "Physically disabled white girl asking for help in a bodega" (SP9) |  | Defaulting to certain assistive technology (i.e., wheelchair) |
| "A Latino child using an eye-gaze controlled AAC device" (SP10) |  | Incorrect depiction and modeling of assistive technology usage; Defaulting to certain assistive technology (i.e., high-tech AAC) |
| "A middle school aged boy with Down Syndrome using an iPad to verbally communicate" (SP4) |  | Incorrect depiction and modeling of assistive technology usage; Defaulting to certain assistive technology (i.e., high-tech AAC) |
| "A series of realistic images including African American families and Latin American families where adults are modeling use of augmentative and alternative communication devices for children during mealtimes, while watching television, and at bedtimes" (SP3) |  | Incorrect depiction and modeling of assistive technology usage; Defaulting to certain assistive technology (i.e., high-tech AAC); Unnecessary addition of extraneous hardware to assistive technology |

### 5.2 Tensions and Tradeoffs in Using AI for Speech Language Pathology and its Impact on Equitable Care

Using AI in speech-language pathology introduces tensions and tradeoffs that impact both quality and equity of care for CLD children with disabilities, as well as cognitive and temporal demands on SLPs. Although AI holds potential (Section 4.2), its integration raises concerns about exacerbating existing disparities, creating uneven distribution of AI's benefits, and increasing burdens on SLPs already managing high workloads [24]. In this subsection, we explore tensions around the need for personalized and responsive care (Section 4.1.1 and Section 4.2) and AI's current limitations in being biased, falling short in understanding SLP contexts and identity-related nuances, and lacking explainability and transparency (Section 4.3 and Section 4.4).

**Table 6: Re-generated image-based materials with associated prompts, removed identity dimensions, and identified bias.**

| **Original Prompt and Generated Image** | **Removed Identities** | **Prompt and Generated Image Without Identities** | **Identified Bias** |
|---|---|---|---|
| **Prompt:** "2nd grade classroom with a female teacher that wears a **hijab**. The students should be wearing clothing of a **diverse range of cultures**" (SP11)  | Culture | **Prompt:** 2nd grade classroom with a female teacher   | Defaulting to White and Euro-centric representation |
| **Prompt:** "A **Latino** child using an eye-gaze controlled AAC device" (SP10)   | Culture | **Prompt:** A child using an eye-gaze controlled AAC device   | Defaulting to White and Euro-centric representation; Disability Tropes (i.e., sad portrayal, lonely portrayal) |
| **Prompt:** "A **physically disabled** white girl asking for help in a bodega" (SP9)  | Disability Type | **Prompt:** A disabled white girl asking for help in a bodega  | Defaulting to certain assistive technology (i.e., wheelchair) |
| **Prompt:** "A 3-person family that includes an **African-American** father wearing a U.S. Army uniform, a **Samoan mother** wearing nurse's scrubs, and a 14 year old **African-American Samoan** female with **Down Syndrome**" (SP5)   | Culture & Disability | **Prompt:** A 3-person family that includes a father wearing a U.S. Army uniform, a mother wearing nurse's scrubs, and a 14 year old female  | Defaulting to White and Euro-centric representation; Defaulting to non-disabled identities |

*5.2.1 Getting it Right versus Saving Time.* We see this tension manifest through the additional role that SLPs are increasingly required to take as prompt engineers, as indicated by survey participant desire to re-craft prompts according to identified shortcomings (Section 4.3.3). In addition to the AI needing to understand the context of use, the ability for current AI to produce relevant and effective materials is largely contingent upon language of the provided prompt, thus pressuring SLPs to master prompt engineering and an ability to craft precise and contextually rich prompts. However, SLPs are generally not trained in prompting, which led to participants expressing dissatisfaction with initial AI-generated results that lacked necessary cultural, linguistic, disability, and speech-language pathology specific details. Many participants therefore wanted to revise their prompts, adding more nuanced information in hopes of producing better-aligned, accurate, relevant, and safe outcomes. Unfortunately, such a process of iterative prompt refinement introduces significant challenges for SLPs. First, it risks creating disparities in the quality of care provided to CLD children with disabilities, as SLPs with less experience and familiarity with AI may struggle to produce materials that address the diverse needs of their children. Second, the task of prompt engineering to obtain accurate and sensitive materials creates additional cognitive and temporal demands. This is particularly concerning relative to high SLP workloads [24], as it could either limit SLP capacity to engage in the iterative process of refining prompts or could mean that such engagement detracts from core responsibilities in their direct interaction with children. This presents a critical tradeoff, as SLPs must choose between relying on AI to quickly generate materials (i.e., which may lack in necessary depth) or investing extra time to refine prompts to meet the specific needs of their diverse caseloads (i.e., which may compromise other areas of practice).

*5.2.2 Cross-Checking versus Saving Time.* Related but distinct from the above in prompt authoring, we also see tension manifest via cognitive and temporal demands placed on SLPs to cross-check AI-generated materials for bias (i.e., distinguishing between demands in perfecting prompts versus validating AI-generated materials). This tension highlights discrepancies between expectations of AI streamlining the creation of culturally, linguistically, and disability adapted materials and actual demands imposed on SLPs to verify the accuracy, appropriateness, and alignment of generated material to meet the needs of CLD children with disabilities. In such cases, potential time-saving benefits of AI are undermined by its lack of contextual awareness and embedded biases, introducing additional responsibilities for SLPs. A specific prompt or prompting structure that was previously effective also cannot be assumed effective in re-use (e.g., due to differences in content provided within a re-used prompting structure, due to different potential verification needs or biases an SLP may want to consider with different students, due to evolution of the underlying AI in the time since a prompt was previously used). This presents a critical tradeoff, as SLPs must choose between investing additional time to verify AI-generated materials to ensure accuracy and mitigate bias (i.e., which adds more to their already high workloads [24]) or relying on AI-generated materials without cross-checking to save time (i.e., which may reproduce and reinforce bias).

*5.2.3 Accuracy versus Transparency and Explainability.* Lastly, we see tensions between the accuracy of AI-generated materials and the transparency and explainability of AI systems. Transparency refers to the degree to which an AI system's operations, algorithms, data sources, and decision-making criteria are made visible and understandable [41]. Explainability refers to the system's ability to articulate its reasoning for individual decisions or outputs, particularly for non-experts [16, 90]. Development of AI systems that are both highly accurate and adequately transparent and explainable is a persistent and well-documented challenge in AI research [13, 67]. For example, LLMs such as ChatGPT are often described as "black-box" models, tend to be highly accurate, but are also opaque and provide insufficient post-hoc explanations for how they arrive at specific outputs [13, 67]. This challenge becomes particularly significant in the context of SLPs using AI to create personalized materials to support CLD children with disabilities. SLPs described accuracy as crucial (Section 4.3) and equally emphasized This is further complicated by inherent biases of current models. With AI not clearly communicating how it creates personalized materials to support CLD children with disabilities, and with SLPs who do not possess the knowledge or time required to identify and correct culturally, linguistically, or disability-related inaccuracies in AI-generated materials, SLPs risk unknowingly incorporating biased content into their practice.

## 5.3 Honoring Diverse Representations of Culture, Language, and Disability in AI-generated Materials

Our analysis identified cultural, linguistic, and disability AI biases (Section 4.4), consistent with prior literature in other settings [60]. Our findings demonstrate these biases manifest within speech-language pathology settings and examine impacts they can have on the quality of care provided to CLD children with disabilities. Such biases are at least in part a product of training datasets, which perpetuate systemic biases, stereotypes, and societal assumptions [39]. However, our findings prompt a critical examination of whether biases should always be categorized as harmful, insensitive, or offensive. This question becomes particularly salient when considering different regional contexts, such as outside of the United States or beyond Western societal contexts. The conceptualization of, understanding of, and experience of cultural, linguistic, and disability identities is shaped by a wide range of factors and can vary widely across individuals, let alone across different regions, as can the attitudes, values, language, and beliefs associated with these identities. What is deemed acceptable or unharmful in one context may be perceived as biased or harmful in another [39]. It is thus important to recognize that biases identified in our analysis might be interpreted differently in other social contexts, which in turn underscores the importance of situating AI evaluations and bias assessments within their relevant sociocultural contexts to avoid overgeneralizing harm or insensitivity without considering the diversity of perspectives on identity and representation. Even within a specific sociocultural context, some biases and stereotypes that AI produces are not always inherently harmful and are instead the realities of people. For example, many people do use canes or wheelchairs, so although Section 4.4 highlighted it can be problematic for AI to assume these assistive technologies, it would also be problematic for AI to erase them. Similarly, Section 4.4 highlighted AI perpetuation of stereotypes in names, foods, and sports when prompted to generate stories with culturally relevant details, but those specific names, foods, or sports are not inherently harmful and many even hold cultural significance and correspond to the experiences of some people. Instead, harm arises when AI consistently represents the identities of people and communities *only* consisting of these singular, monolithic experiences.

Perhaps it is important to consider the prevalence of such biases and the act of solely associating identities with reductive narratives is what makes it harmful, because this fails to acknowledge the range of diverse experiences within identities. Adichie captures this sentiment in "The Danger of a Single Story," articulating risks when communities are reduced to a singular narrative: *"show a people as one thing, as only one thing, over and over again, and that is what they become"* [9]. However, we do acknowledge that some of the biases observed in our work can be inherently harmful and may inadvertently further marginalize certain groups [39]. Ferrara discusses the subjectivity of fairness and how varying interpretations contribute to the complexity of removing bias from AI models [39]. Achieving this requires technology developers to define what "fair" means, but this is exceptionally difficult because they must account for the array of experiences and perspectives among stakeholders [39]. Taken as a whole, when anticipating and designing for SLPs to use AI tools to provide cultural, linguistic, and disability adaptations to tailor materials to CLD children with disabilities, it is important that these be tailored at the individual level of a child's specific needs and identity dimensions (Section 4.2.1), rather than generic adaptations based on a single culture or disability. Where feasible in specific design settings, approaches to integrating AI could also include children and families in the process of tailoring.

## 5.4 Recommendations for SLPs and AI-Based Tools

AI presents challenges in supporting SLPs in delivering equitable and responsive care. Given the wide adoption of commercially available AI and its growing uptake in speech-language pathology, we offer considerations for SLPs interacting with AI and for technology designers to inform general-purpose AI. Although not explicitly provided as a recommendation, we also emphasize that efforts must extend beyond technologists. Systemic issues within speech-language pathology must be addressed (e.g., lack of diversity among SLPs). Moreover, ensuring diverse training data and mitigating underlying AI bias remains a critical priority.

*5.4.1 AI Literacy and SLP Interactions with AI.* SLPs must prioritize AI literacy [59, 74] to navigate limitations of these systems. Recognizing that AI outputs often reflect biases in their training is critical to minimizing risks of perpetuating those biases (e.g., in delivering biased materials). Additionally, SLPs should develop ethical prompt crafting skills, explicitly incorporating cultural, linguistic, and disability-specific nuances to generate more representative outputs. Moreover, structured protocols for detecting, evaluating, and mitigating biases in AI-generated outputs are essential. Training programs should equip SLPs to critically assess AI-generated materials, make informed adjustments, and maintain ethical standards in AI interactions.

*5.4.2 Prompt Assistance to Improve Care in Prompting.* Current generative AI lacks adequate support for crafting effective prompts, forcing SLPs to prompt engineer and to navigate associated cognitive and temporal costs of refinement. General-purpose AI-based tools should incorporate guided prompting approaches that allow SLPs to enable step-by-step assistance with interactive refinement, breaking down the prompt creation process into manageable steps. This could be especially useful for SLPs with limited AI experience or for those who may overlook contextual nuances. For example, dynamic templates could guide SLPs through creating prompts, allowing them to select fields and input details tailored to their task (e.g., image, story) and applicable representation details (e.g., race, ethnicity, language, disability). Through interactive refinement and follow-up questions, SLPs can specify domain-specific therapeutic goals (e.g., articulation, social skills, expressive language), and specific interests (e.g., sports, music). As SLPs input information, a system could provide real-time feedback, identifying issues and areas for improvement (e.g., missing details). Such assistance could support and improve care in prompt crafting, reducing the risk of omitting identity details of children, minimize iterative prompt refinement, and reduce cognitive and temporal burdens on SLPs.

*5.4.3 Greater Explainability and Transparency in AI-Based Support.* Current generative AI falls short in providing transparency and explainability required both for building trust and reliability with SLPs and for supporting identification and evaluation of bias. AI-based tools should provide explanations of how they develop CLD adaptations in generated materials. This could include techniques like citations to sources incorporated by an AI, but also explanations of how more individualized prompting influenced adapted outputs (e.g., how guidance provided by a specific child or their family was used in tailoring materials related to that child). Clear, accessible, and interpretable [35] explanations can help SLPs understand how a model's decisions align with their therapeutic goals and a child's specific needs. Additionally, AI should improve transparency regarding potential biases. Available AI generally offers vague disclaimers about possible inaccuracies (e.g., ChatGPT includes a small light gray statement at the bottom of its interface: *"ChatGPT can make mistakes. Check important info."*) Although this acknowledges potential errors, it lacks the specificity needed to guide people in identifying biases, understanding their implications, or fact-checking information. Although platforms like OpenAI offer documentation [3], this fails to address the specific needs of SLPs. AI-based tools should implement specific warnings about potential biases, their implications, and best practices for identifying and mitigating them. This would empower SLPs to critically assess AI-generated material before incorporating it into their practice, reducing the risk of perpetuating harmful biases.

## 6 Limitations and Future Work

Our research focused on the experiences and perspectives of SLPs and how AI might support their practice, emphasizing scenarios where an SLP is the stakeholder most directly interacting with an AI-based tool (e.g., in preparing a prompt). Our research did not directly include the perspectives of CLD children with disabilities, in part because the SLPs in our target group primarily work with very young children. Because children are the recipients of speech-language pathology services, their experiences are crucial in understanding the impact of AI on speech and language development, thus future research should include children's insights. Additionally, future research should incorporate insights from families and caregivers in gaining a more holistic understanding of the cultural sensitivity, appropriateness, and effectiveness of AI-generated materials, as they play a crucial role in child speech and language development.

Recognizing the prevalence of re-prompting in AI interactions, future work should explore this process and the learnability of such interactions for SLPs. Moreover, future work should further investigate the impact of AI literacy on SLP prompt crafting and its influence on the desirability of generated results.

Our study also focused primarily on the axes of race, ethnicity, language, and disability, thereby excluding other important identity factors (e.g., gender). Such dimensions can additionally influence how effectively AI technologies support the needs of children, so future work should more broadly examine additional intersections. Our study is also limited to U.S. speech-language pathology practices and practitioner expectations for AI support, which may vary greatly in non-U.S. settings. Additionally, definitions and experiences of race, ethnicity, and disability differ across cultural contexts, along with perceptions of bias and fairness, making it essential for future studies to explore these dynamics in additional contexts. Lastly, linguistic barriers encountered by our authors (i.e., we lack proficiency in some of the languages present in AI-generated materials) restricted our ability to analyze some materials which may have included additional embedded biases.

## 7 Conclusion

This work explores challenges faced by speech-language pathologists in providing responsive care to CLD children with disabilities, areas where SLPs would like AI-based support, SLP perceptions of the performance of available AI in addressing challenges and opportunities, and biases present in AI-generated speech-language pathology materials. We discuss implications of contextually unaware AI, the varying levels of care SLPs apply when crafting AI prompts, tensions and tradeoffs of integrating AI-based tools into responsive practices, and the necessity of honoring diverse representations in AI-generated materials. In addition to addressing systemic challenges within the profession (e.g., the lack of diversity among professionals), we offer recommendations for SLPs interacting with AI-based tools and general-purpose AI tools in their practice.

## Acknowledgments

This material is based upon work supported under the AI Research Institutes program by the National Science Foundation and the Institute of Education Sciences, U.S. Department of Education, through Award DRL-2229873 - AI Institute for Transforming Education for Children with Speech and Language Processing Challenges (or National AI Institute for Exceptional Education). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education. This work was also supported in part by the Jacob's Foundation CERES Center and by the University of Washington through the Center for Research and Education on Accessible Technology and Experiences (CREATE), the Paul G. Allen School of Computer Science & Engineering, the Population Health Initiative, and the Race & Equity Initiative.

## References

[1] Annual Demographic and Employment Data 2023 Member and Affiliate Profile. https://www.asha.org/siteassets/surveys/2023-member-affiliate-profile.pdf
[2] Employment Settings for SLPs. https://www.asha.org/students/employment-settings-for-slps/
[3] Is ChatGPT biased? https://help.openai.com/en/articles/8313359-is-chatgpt-biased
[4] Multilingual Service Delivery in Audiology and Speech-Language Pathology. https://www.asha.org/practice-portal/professional-issues/multilingual-service-delivery/
[5] Profile of ASHA Multilingual Service Providers. https://www.asha.org/siteassets/surveys/2023-profile-of-multilingual-service-providers.pdf
[6] Quick Statistics About Voice, Speech, Language. https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language#:~:text=More%20than%20half%20(55.2%25),services%20in%20the%20past%20year.
[7] Guidelines for Writing About People With Disabilities. https://adata.org/factsheet/ADANN-writing
[8] National AI Institute for Exceptional Education. https://www.buffalo.edu/ai4exceptionaled.html
[9] Chimamanda Ngozi Adichie. The Danger of a Single Story.
[10] Beena Ahmed, Penelope Monroe, Adam Hair, Chek Tien Tan, Ricardo Gutierrez-Osuna, and Kirrie J Ballard. 2018. Speech-driven Mobile Games for Speech Therapy: User Experiences and Feasibility. *International journal of speech-language pathology* 20, 6 (2018), 644–658.
[11] Jaimeen Ahn and Alice Oh. 2021. Mitigating Language-Dependent Ethnic Bias in BERT. *arXiv preprint arXiv:2109.05704* (2021).
[12] Hussam Alkaissi and Samy I McFarlane. 2023. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus* 15, 2 (2023).
[13] Alessa Angerschmid, Jianlong Zhou, Kevin Theuermann, Fang Chen, and Andreas Holzinger. 2022. Fairness and Explanation in AI-informed Decision Making. *Machine Learning and Knowledge Extraction* 4, 2 (2022), 556–579.
[14] Subini Ancy Annamma, David Connor, and Beth Ferri. 2013. Dis/ability Critical Race Studies (DisCrit): Theorizing at the Intersections of Race and Dis/ability. *Race ethnicity and education* 16, 1 (2013), 1–31.
[15] Graciela Arias and Jennifer Friberg. 2017. Bilingual Language Assessment: Contemporary Versus Recommended Practice in American Schools. *Language, Speech, and Hearing Services in Schools* 48, 1 (2017), 1–15.
[16] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI. *Information fusion* 58 (2020), 82–115.
[17] Amanda Askell, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, et al. 2021. A General Language Assistant as a Laboratory for Alignment. *arXiv preprint arXiv:2112.00861* (2021).
[18] American Speech-Language-Hearing Association et al. 2016. Scope of Practice in Speech-Language Pathology. (2016).
[19] Moya Bailey and Izetta Autumn Mobley. 2019. Work in the Intersections: A Black Feminist Disability Framework. *Gender & Society* 33, 1 (2019), 19–40.
[20] Jessica Ball and B May Bernhardt. 2008. First Nations English Dialects in Canada: Implications for Speech-Language Pathology. *Clinical linguistics & phonetics* 22,

8 (2008), 570–588.

[21] Cynthia L Bennett and Os Keyes. 2020. What is the Point of Fairness? Disability, AI and the Complexity of Justice. *ACM SIGACCESS Accessibility and Computing* 125 (2020), 1–1.

[22] Federico Bianchi, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2023. Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1493–1504.

[23] Zuzana Bílková, Adam Novozámskỳ, Michal Bartoš, Adam Domínec, Šimon Greško, Barbara Zitová, Markéta Paroubková, and Jan Flusser. 2020. Human Computer Interface Based on Tongue and Lips Movements and its Application for Speech Therapy System. *Electronic Imaging* 32 (2020), 1–5.

[24] Gordon W Blood, Jenna Swavely Ridenour, Emily A Thomas, Constance Dean Qualls, and Carol Scheffner Hammer. 2002. Predicting Job Satisfaction Among Speech-Language Pathologists Working in Public Schools. (2002).

[25] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. *Thematic Analysis*. Springer Singapore, Singapore, 843–860. https://doi.org/10.1007/978-981-10-5251-4_103

[26] Lena G Caesar and Paula D Kohler. 2007. The State of School-Based Bilingual Assessment: Actual practice versus recommended guidelines. (2007).

[27] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (2024), 1–45.

[28] Jaemin Cho, Abhay Zala, and Mohit Bansal. 2023. Dall-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generation Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3043–3054.

[29] Debra Collette, Alex Brix, Patricia Brennan, Nicole DeRoma, and Brittney C Muir. 2019. Proloquo2Go Enhances Classroom Performance in Children With Autism Spectrum Disorder. *OTJR: Occupation, Participation and Health* 39, 3 (2019), 143–150.

[30] Paul A Dagenais and Jamequa A Stallworth. 2014. The Influence of Dialect upon the Perception of Dysarthic Speech. *Clinical Linguistics & Phonetics* 28, 7-8 (2014), 573–589.

[31] Aayushi Dangol, Aaleyah Lewis, Hyewon Suh, Xuesi Hong, Hedda Meadan, James Fogarty, and Julie A Kientz. 2025. "I Want to Think Like an SLP": A Design Exploration of AI-Supported Home Practice in Speech Therapy. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[32] Chinmoy Deka, Abhishek Shrivastava, and Rishav Kumar. 2024. Towards Human-Centered AI in Speech Therapy: Perspectives from a Low-Resource Setting. (2024).

[33] Richard Delgado and Jean Stefancic. 2023. *Critical Race Theory: An Introduction*. Vol. 87. NYU press.

[34] Sunipa Dev, Tao Li, Jeff M Phillips, and Vivek Srikumar. 2020. On Measuring and Mitigating Biased Inferences of Word Embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 7659–7666.

[35] Finale Doshi-Velez and Been Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608* (2017).

[36] Yao Du and Felix Juefei-Xu. 2023. Generative AI for Therapy? Opportunities and Barriers for ChatGPT in Speech-Language Therapy. (2023).

[37] Catherine Easton and Sarah Verdon. 2021. The Influence of Linguistic Bias Upon Speech-Language Pathologists' Attitudes Toward Clinical Scenarios Involving Nonstandard Dialects of English. *American Journal of Speech-Language Pathology* 30, 5 (2021), 1973–1989.

[38] Audrey M Farrugia-Bernard. 2018. Speech-Language Pathologists as Determiners of the Human Right to Diversity in Communication for School Children in the US. *International Journal of Speech-Language Pathology* 20, 1 (2018), 170–173.

[39] Emilio Ferrara. 2023. Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. *arXiv preprint arXiv:2304.03738* (2023).

[40] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. "I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 205–216.

[41] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2018. Explaining Explanations: An Overview of Interpretability of Machine Learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*. IEEE, 80–89.

[42] Sophie Groenwold, Lily Ou, Aesha Parekh, Samhita Honnavalli, Sharon Levy, Diba Mirza, and William Yang Wang. 2020. Investigating African-American Vernacular English in Transformer-Based Text Generation. *arXiv preprint arXiv:2010.02510* (2020).

[43] Mark Guiberson and Jenny Atkins. 2012. Speech-Language Pathologists' Preparation, Practices, and Perspectives on Serving Culturally and Linguistically Diverse Children. *Communication Disorders Quarterly* 33, 3 (2012), 169–180.

[44] Wei Guo and Aylin Caliskan. 2021. Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (Virtual Event, USA) *(AIES '21)*. Association for Computing Machinery, New York, NY, USA, 122–133. https://doi.org/10.1145/3461702.3462536

[45] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a Critical Race Methodology in Algorithmic Fairness. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 501–512.

[46] Sierrah Harris and Amanda Owen Van Horne. 2021. Speech-Language Pathologist's Race, But Not Caseload Composition, Is Related to Self-Report of Selection of Diverse Books. *Perspectives of the ASHA Special Interest Groups* 6, 5 (2021), 1263–1272.

[47] Cynquetta Harris-Johnson. 2023. THROUGH A DIS/ABILITY STUDIES AND CRITICAL RACE THEORY LENS: BLACK FAMILIES'PERSPECTIVES OF SPEECH LANGUAGE PATHOLOGY. (2023).

[48] Ben Hutchinson, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyl. 2020. Social Biases in NLP Models as Barriers for Persons with Disabilities. *arXiv preprint arXiv:2005.00813* (2020).

[49] Ben Hutchinson, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyl. 2020. Unintended Machine Learning Biases as Social Barriers for Persons with Disabilities. *ACM SIGACCESS Accessibility and Computing* 125 (2020), 1–1.

[50] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv.* 55, 12, Article 248 (mar 2023), 38 pages. https://doi.org/10.1145/3571730

[51] Lianqin Jia, Mengmeng Zhang, and Jiasen Li. 2022. Research on the Application of Artificial Intelligence in the Rehabilitation Training of Children with Speech Disorders. In *International Conference on Computer Engineering and Networks*. Springer, 1463–1469.

[52] Heila Jordaan. 2008. Clinical Intervention for Bilingual Children: An International Survey. *Folia Phoniatrica et Logopaedica* 60, 2 (2008), 97–105.

[53] Kathryn Kohnert, Dongsun Yim, Kelly Nett, Pui Fong Kan, and Lillian Duran. 2005. Intervention With Linguistically Diverse Preschool Children. (2005).

[54] Youjin Kong. 2022. Are "Intersectionally Fair" AI Algorithms Really Fair to Women of Color? A Philosophical Analysis. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 485–494. https://doi.org/10.1145/3531146.3533114

[55] Effie Papoutsis Kritikos. 2003. Speech-Language Pathologists' Beliefs About Language Assessment of Bilingual/Bicultural Individuals. (2003).

[56] Sandra P Laing and Alan Kamhi. 2003. Alternative Assessment of Language and Literacy in Culturally and Linguistically Diverse Populations. (2003).

[57] Sue Ann S Lee. 2019. Virtual Speech-Language Therapy for Individuals with Communication Disorders: Current Evidence, Limitations, and Benefits. *Current Developmental Disorders Reports* 6 (2019), 119–125.

[58] Tao Li, Tushar Khot, Daniel Khashabi, Ashish Sabharwal, and Vivek Srikumar. 2020. UNQOVERing Stereotyping Biases via Underspecified Questions. *arXiv preprint arXiv:2010.02428* (2020).

[59] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[60] 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 CHI Conference on Human Factors in Computing Systems*. 1–23.

[61] Liam Magee, Lida Ghahremanlou, Karen Soldatic, and Shanthi Robertson. 2021. Intersectional Bias in Causal Language Models. *arXiv preprint arXiv:2107.07691* (2021).

[62] Sharynne Mcleod and Elise Baker. 2014. Speech-Language Pathologists' Practices Regarding Assessment, Analysis, Target Selection, Intervention, and Service Delivery for Children with Speech Sound Disorders. *Clinical linguistics & phonetics* 28, 7-8 (2014), 508–531.

[63] Si Ioi Ng, Dehua Tao, Jiarui Wang, Yi Jiang, Wing Yee Ng, and Tan Lee. 2018. An Automated Assessment Tool for Child Speech Disorders. In *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*. 493–494. https://doi.org/10.1109/ISCSLP.2018.8706577

[64] Ihudiya Finda Ogbonnaya-Ogburu, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376392

[65] Michelle Pascoe and Vivienne Norman. 2011. Contextually Relevant Resources in Speech-Language Therapy and Audiology in South Africa-are There Any? (2011).

[66] Wendy M Pearce and Cori Williams. 2013. The Cultural Appropriateness and Diagnostic Usefulness of Standardized Language Assessments for Indigenous Australian Children. *International Journal of Speech-Language Pathology* 15, 4

(2013), 429–440.

[67] Emmanuel Pintelas, Ioannis E Livieris, and Panagiotis Pintelas. 2020. A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability. *Algorithms* 13, 1 (2020), 17.

[68] Chelsea Privette. 2021. Critical race theory for speech-language pathology: How race-conscious practice mitigates disparities. In *Critical perspectives on social justice in speech-language pathology*. IGI Global, 84–104.

[69] Rida Qadri, Renee Shelby, Cynthia L Bennett, and Emily Denton. 2023. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 506–517.

[70] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[71] V Robles-Bykbaev, M Guamán-Heredia, Y Robles-Bykbaev, J Lojano-Redrován, F Pesántez-Avilés, D Quisi-Peralta, M López-Nores, and J Pazos-Arias. 2017. Onto-SPELTRA: A Robotic Assistant Based on Ontologies and Agglomerative Clustering to Support Speech-Language Therapy for Children with Disabilities. In *Advances in Computing: 12th Colombian Conference, CCC 2017, Cali, Colombia, September 19-22, 2017, Proceedings 12*. Springer, 343–357.

[72] Vladimir E Robles-Bykbaev, Martín López-Nores, José J Pazos-Arias, and Daysi Arévalo-Lucero. 2015. SPELTA: An Expert System to Generate Therapy Plans for Speech and Language Disorders. *Expert Systems with Applications* 42, 21 (2015), 7641–7651.

[73] Wendy D Roth. 2016. The Multiple Dimensions of Race. *Ethnic and Racial Studies* 39, 8 (2016), 1310–1338.

[74] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, et al. 2024. Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions. *arXiv preprint arXiv:2406.09264* (2024).

[75] Emily Sheng, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. 2019. The Woman Worked as a Babysitter: On Biases in Language Generation. *arXiv preprint arXiv:1909.01326* (2019).

[76] Genevieve Smith, Eve Fleisig, Madeline Bossi, Ishita Rustagi, and Xavier Yin. 2024. Standard Language Ideology in AI-Generated Language. *arXiv preprint arXiv:2406.08726* (2024).

[77] Shameka Stanford and Bahiyyah Muhammad. 2017. The Confluence of Language and Learning Disorders and the School-to-Prison Pipeline Among Minority Students of Color: A Critical Race Theory. *Am. UJ Gender Soc. Pol'y & L.* 26 (2017), 691.

[78] Carol Stow and Barbara Dodd. 2003. Providing an Equitable Service to Bilingual Children in the UK: A Review. *International Journal of Language & Communication Disorders* 38, 4 (2003), 351–377.

[79] Hyewon Suh, Aayushi Dangol, Hedda Meadan, Carol A Miller, and Julie A Kientz. 2024. Opportunities and Challenges for AI-Based Support for Speech-Language Pathologists. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–14.

[80] David Sztahó, Gábor Kiss, and Klára Vicsi. 2018. Computer Based Speech Prosody Teaching System. *Computer speech & language* 50 (2018), 126–140.

[81] Yi Chern Tan and L Elisa Celis. 2019. Assessing Social and Intersectional Biases in Contextualized Word Representations. *Advances in neural information processing systems* 32 (2019).

[82] Alexandra To, Angela DR Smith, Dilruba Showkat, Adinawa Adjagbodjou, and Christina Harrington. 2023. Flourishing in the Everyday: Moving Beyond Damage-Centered Design in HCI for BIPOC Communities. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 917–933.

[83] Shari Trewin. 2018. AI Fairness for People with Disabilities: Point of View. *arXiv preprint arXiv:1811.10670* (2018).

[84] Danang Trijatmiko, Aviv Yuniar Rahman, et al. 2023. Voice Classification of Children with Speech Impairment Using MFCC Kernel-Based SVM. In *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. IEEE, 703–707.

[85] Aigerim Utepbayeva, Nadezhda Zhiyenbayeva, Leila Assylbekova, and Olga Tapalova. 2024. Artificial Intelligence Applications (Fluency SIS, Articulation Station Pro, and Apraxia Farm) in the Psycholinguistic Development of Preschool Children with Speech Disorders. *International Journal of Information and Education Technology* 14, 7 (2024).

[86] Pranav Narayanan Venkit, Mukund Srinath, and Shomir Wilson. 2022. A Study of Implicit Bias in Pretrained Language Models Against People with Disabilities. In *Proceedings of the 29th International Conference on Computational Linguistics*. 1324–1332.

[87] Sarah Verdon, Sharynne McLeod, and Sandie Wong. 2015. Supporting Culturally and Linguistically Diverse Children with Speech, Language and Communication Needs: Overarching Principles, Individual Approaches. *Journal of communication disorders* 58 (2015), 74–90.

[88] Corinne J Williams and Sharynne McLeod. 2012. Speech-Language Pathologists' Assessment and Intervention Practices with Multilingual Children. *International Journal of Speech-Language Pathology* 14, 3 (2012), 292–305.

[89] Dapeng Yang, Eung-Soo Oh, and Yingchun Wang. 2020. Hybrid Physical Education Teaching and Curriculum Design Based on a Voice Interactive Artificial Intelligence Educational Robot. *Sustainability* 12, 19 (2020), 8000.

[90] Jichen Zhu, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G Michael Youngblood. 2018. Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. In *2018 IEEE conference on computational intelligence and games (CIG)*. IEEE, 1–8.