AI-Enhanced Speech-Language Intervention Documentation: Opportunities and Design Goals

Qingxiao Zheng[1(✉)], Abhinav Choudhry[2], Zihan Liu[2], Parisa Rabbani[2], Yuting Hu[1], Abbie Olszewski[3], Yun Huang[2(✉)], and Jinjun Xiong[1(✉)]

[1] University at Buffalo, Buffalo, NY, USA
{qingxiao,yhu54,jinjun}@buffalo.edu
[2] University of Illinois at Urbana-Champaign, Champaign, IL, USA
{ac62,zihanl18,rabbani8,yunhuang}@illinois.edu
[3] University of Nevada, Reno, USA
aolszewski@med.unr.edu

Abstract. Speech-Language Pathologists (SLPs) support individuals with communication difficulties, but high caseloads and heavy documentation demands often lead to strain and burnout. This research examines how AI-generated intervention documentation can support SLPs and how their insights can inform future AI-enhanced documentation systems. Through a formative study with 17 SLPs, we used an iterative, human-centered approach to examine how AI-generated documentation aligned with their professional practices. Our findings reveal four key opportunities for using AI in documentation and propose three fluidity-focused design goals—contextual, assessment, and individual—for future systems. These goals aim to balance individualization in special education with high-quality documentation. This study advances understanding of how AI can support SLP workflows and inform the design of documentation tools across educational settings where structured documentation is critical.

Keywords: human-centered AI · special education · vision-language model · behavior analysis · design space

1 Introduction

In the U.S., 3.4 million children are served under the Individuals with Disabilities Education Act, with over half requiring support from speech-language pathologists (SLPs) [20]. SLPs assess and treat speech and language difficulties [1], but increasing demand and high caseloads often lead to burnout [24]. One of the most time-consuming tasks is documenting intervention sessions, typically through SOAP (Subjective, Objective, Assessment, Plan) notes, which document the performance, analysis, and planning for subsequent sessions [22]. Alleviating documentation work could allow SLPs to devote more time to the most meaningful aspects of their work, such as enhancing educational quality.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. I. Cristea et al. (Eds.): AIED 2025, LNAI 15882, pp. 132–140, 2025.
https://doi.org/10.1007/978-3-031-98465-5_17

Recent advances in generative AI, particularly vision-language models (VLMs), show promise for automating video comprehension [5,25] and text summarization [13]. These capabilities present a potential opportunity to automatically document intervention session details. While current AI and ML tools focus on domain-specific tasks in special education [2,4], the individualized nature of interventions in special education calls for human-centered design in developing AI for education (AIED) systems [14,21,23,30]. To explore AI's role in supporting SLP workflows, we ask, **RQ**: *What opportunities do SLPs identify in AI-generated intervention session documentation, and how do their insights inform the design goals of future AI-enhanced documentation systems?*

Using a human-centered approach, we conducted a formative study with 17 SLPs, who reviewed AI-generated documentation from real intervention session videos and conducted a think-aloud activity. Through thematic analysis, we identified AI's opportunities and perceived limitations in supporting intervention documentation. We iteratively refined an AI prompt based on each participant's insights, incorporating SLP documentation needs. With this iterative process, we synthesized a set of Prompt Sheets to guide future AI-supported documentation in speech-language education. Our findings reveal key opportunities for using generative AI to address the documentation challenges of SLPs in special education. We also identify three design goals for future AI-enhanced systems that align with SLPs' needs and support professional practice in education settings.

2 Related Work

Speech-language pathologists (SLPs) rely on detailed documentation of client behaviors, such as attention-seeking, engagement, emotional state, and verbal expression, to guide intervention [3]. This documentation, often considered part of the "invisible workload", is essential for evaluation and planning but is time-consuming and labor-intensive beyond service hours [24]. HCI researchers have explored participatory design in developing AI-driven speech therapy tools [8,29], while recent AIED studies highlight the importance of human-centered design in educational systems [14,21,23,30]. Our work builds on this foundation.

AI and machine learning have been applied to personalized treatment plans and real-time feedback, with models analyzing body movements to infer intent, face-touching linked to thought or imitation [6,10]. Existing behavior documentation tools focus on general behavior detection: ConverSense captures audio-based social signals in clinical settings [4], Patel et al. detect nonverbal cues for empathic feedback, and Arakawa and Yakura analyzes gaze to identify anomalies in coaching. Group behavior tools include MeetingCoach's behavioral dashboards [28] and analyses of team cohesion [19]. Generative models have been explored for clinical use, e.g., Du et al. [9] found it helpful for creating therapy materials, including vocabulary and bilingual support. However, current tools offer limited support for domain-specific needs in speech-language special education contexts.

Generative AI and Vision-Language Models (VLMs) like GPT-4o, LLaVA, and Gemini offer new opportunities for intervention documentation [32]. These models can interpret complex verbal and non-verbal behaviors by analyzing text, video, and audio to generate detailed behavioral descriptions [11,15]. With strong in-context learning, VLMs adapt to new tasks with minimal data and update behavior *"on the fly"*. Yet challenges remain in multi-modal understanding, including interpreting domain-specific social cues and mitigating bias [31]. We examines their potential and limitations in supporting SLPs.

3 Method

Study Procedure. Participants watched two group-based speech-language intervention videos using a think-aloud protocol, pausing to share observations. They were prompted with guiding questions (e.g., *What behaviors stood out? Why were they significant?*) to aid interpretation. Afterward, they reviewed AI-generated analyses and responded to follow-up questions (e.g., *What did the AI get right or miss?*). They also wrote a "wishlist" for AI's analysis in the future. We recruited 17 special education professionals (SLPs, SLPAs, BCBAs, etc.) via a special education listserv. Participants (16F, 1M), aged 25–44, had 2.5–33 years of experience (M = 13.9). Each 1.5-h interview was IRB-approved and compensated with a $50 Amazon gift card. We used thematic analysis [7] to identify AI opportunities and concerns. Two independent coders iteratively grouped participant utterances into sub-themes and broader themes. Prompt revisions were tracked and later synthesized into Prompt Sheets. Coding discrepancies were discussed and resolved collaboratively.

Preparation of AI's Analysis of Videos. We selected 20 two-minute videos from the SKILL book [12], showing SLPs working with children, with author permission. Each participant viewed two randomly assigned videos. To evaluate AI's support for intervention analysis, we used an iterative prompting approach with three inputs: (1) a base prompt, (2) video frames, and (3) transcripts. For the first video, all participants received the same AI output generated from this default prompt: *"You are a special education expert in Speech-Language Pathology. I want you to see some frames taken in sequence and read the transcript of an intervention session, which you should analyze."* To improve AI analysis, we iteratively refined the base prompt using participant feedback. After each session, we incorporated suggestions, such as emphasizing cueing strategies, adding quantitative metrics, and refining grammar assessments, into a revised prompt for the second video. The first video used a fixed prompt to establish a baseline. An excerpt from the final refined prompt: *"Add numbers to your analysis... Note whether children answered 'what,' 'who,' 'where,' etc., questions correctly... Include cues required, any avoidance or attention-seeking behavior, and other relevant quantitative observations."* This iterative process helped us identify effective prompting strategies, later synthesized into Prompt Sheets shared in our findings. AI outputs were generated using GPT-4V. Since the model cannot process videos directly, we input batches of 10 still frames and researcher-edited transcripts based on Whisper AI output. Transcripts were manually corrected for missing words, speaker labels, and vocal cues.

4 Result

We reported the potential use cases of using AI to assist SLP documentation. Example AI outputs can be viewed in Table 1.

Table 1. SLPs' Perceived AI Use Cases and Example Outputs

Opportunities	Example AI Output in the Study (with Participant ID)
Track Intervention Sessions	**(P11) Answers given**: 2 times (e.g., responds "No" to a direct question and provides a specific answer about the story setting) **Behavioral Cues**: The child seems to be engaged but exhibits signs of being slightly distracted or unsure, as indicated by her body language, like touching her face and looking away from the expert **Stimming Behavior**: Fidgeting may signal comfort-seeking or focus management
Discover Unnoticed Patterns	**(P10) Middle Child**: Retold the story (timestamps 15.0–41.0) with interjections from the expert. **Right Child**: Answered questions on story sequence (timestamps 52.0–76.0) **Left Child**: Contributed to storytelling (timestamps 80.0–111.0)
Identify Disparities in Group	**(P16) Behavior**: Somewhat less engaged verbally and possibly showing signs of shyness or hesitation compared to the right and the middle child. There was also a mention of indistinct speech, which might indicate a need for focusing on articulation or confidence in speaking
Inspire for Self-Improvements	**(P14) Cues worked**: Initially, the left child seemed to forget an element of the story but then engaged well after encouragement. She showed signs of being thoughtful, possibly reflecting on the expert's feedback when looking for Tyson's shoe and when adding details to the story **Cues needed**: The child needs verbal affirmation, as indicated by the repeated use of cuing: "elaborate." Positive reinforcement from the expert, such as repeating "slithered," encourages the child to focus on the use of descriptive language

Opportunity 1: Track Intervention Sessions. Participants noted AI's potential to improve documentation by providing quantitative, objective data for tracking individual progress, which addresses inconsistencies in current memory-based methods. This could support longitudinal analysis and informed decision-making (P15). They suggested tracking metrics like *"time off-task"* (P10), *"support needed to answer questions"* (P11), and *"performance accuracy across sessions"* (P17), offering structured insight into behavioral trends over time. Participants also saw AI as a tool to support human judgment, particularly in reducing assessment inconsistencies and comparing session quality across sessions (P12). As P11 noted, *"I might intuitively know a child had a better session today, but having concrete data... is invaluable."*

Opportunity 2: Discover Unnoticed Patterns. Beyond tracking known behaviors, participants valued AI's potential to uncover subtle patterns often missed by humans, such as *"recurring bathroom requests at specific times"* (P16) or *"changes in attentiveness due to room lighting"* (P10). These insights could reveal hidden environmental influences or situational factors influencing client's behavior, supporting more targeted and adaptive interventions.

Opportunity 3: Identify Disparities in Group. Participants noted AI's ability to uncover imbalances in group participation, supporting more equitable intervention. As P16 noted, *"AI can prompt us to consider if a child is answering but not initiating questions."* Such insights can reveal overlooked participation, enabling SLPs to adjust strategies and ensure quieter children receive equal opportunities to engage, supporting more inclusive and effective interventions.

Opportunity 4: Inspire Self-improvement. Participants appreciated AI's ability to offer feedback on their intervention strategies, helping refine instructional practices. For instance, P14 noted that AI could assess instruction quality and distinguish between cueing, prompting, and modeling, providing insights that aren't immediately obvious to them. AI's analyses could also reinforce professional confidence. P9 also positions it as a reflective tool for ongoing professional growth, *"The match between AI's analysis and our evaluations in certain cases can strengthen our confidence in treatment approaches."*

The Prompt Sheet (Table 2) is a structured prompt template to help AI engineers embed task-specific, developmental, and individualized context into prompts for more accurate behavior interpretation and fewer misinterpretations.

Design Goal 1: To Enhance "Contextual Fluidity" in Behavior Tracking. Participants raised concerns about AI's ability to distinguish structured practice from spontaneous use of skills, which is critical for accurate tracking. As P16 asked, *"How can it tell when it's a SKILL?"* We define this challenge as Contextual Fluidity, which refers to AI's capacity to adapt interpretations based on learning context.

Design Goal 2: To Support "Assessment Fluidity" for Flexible Evaluation. Participants noticed the need for AI to adapt evaluations to the developmental stage and learning focus. P16 noted that incorrect grammar (e.g., *"Oh, and we eated lunch earlier!"*) may not warrant correction if not the current focus: *"Sometimes AI needs to know when not to flag things."* Similarly, P14 said, *"The important thing is that they're trying, which is a part of progress."* We term this Assessment Fluidity, which refers to AI's ability to evaluate flexibly based on context, effort, and developmental goals.

Design Goal 3: To Ensure "Individual Fluidity" for Personalized Adaptation. Participants stressed the importance of AI recognizing non-standard behaviors as valid forms of engagement. P14 explained, *"I have a client who needs to rock, so we use a rocking chair for him to focus. If you see a child moving a lot, one might think they are disengaged, but for this child, it's just his way of meeting his sensory needs."* P16 added that *"direct eye contact isn't always a sign of engagement, some may not look directly at you but could still be listening attentively."* We define this adaptability as *"Individual Fluidity"*, which refers to AI's ability to interpret individual differences, such as fidgeting or avoiding eye contact, as engagement rather than disengagement.

Table 2. Prompt Sheet for Contextual, Assessment, and Individual Fluidity

Prompt Structure	SLPs' Example Wish	Example Prompt Example
Design Goal 1: Contextual Fluidity		
Track [*behavior*] within [*task*] and measure its frequency	"I wish AI could track how often a client follows directions correctly in learning practice."	Track how many times the client [follows instructions correctly] during a [listening and comprehension activity]
Track [*behavior*] in [*setting*], but exclude in [*other setting*]	"I wish AI could track engagement in classroom settings but ignore background noise."	Track the client's [engagement] in the [classroom] but exclude when there is [noise] in the background
Track [*behavior*] when [*context*] but ignore it when [*condition*]	"I wish AI could track gaze during tasks but not when distracted."	Track [gaze direction] when the client is [engaged], excluding when [distracted by sensory needs]
Design Goal 2: Assessment Fluidity		
Evaluate success based on [*criteria*], even if [*mistake*]	"I wish AI could track pragmatic skills, even if the topic drifts."	Evaluate [conversation initiation] by [number of attempts], even if [topic is off]
If [*mistake*], count as success if [*progress metric*]	"I wish AI could track engagement via body language, even without eye contact."	If the client [avoids eye contact], consider successful if [body language shows engagement]
Track [*behavior*] along [*dimension 1*] and [*dimension 2*]	"I wish AI could measure both frequency and duration of eye contact."	Track [eye contact] by [frequency] and [duration]
Design Goal 3: Individual Fluidity		
Recognize [*behavior*] as [*target*] rather than disengagement for [*client*]	"I wish AI could learn my client's floor-dropping is engagement."	Recognize [falling to the floor] as [engagement] for [Amy]
Consider [*action*] typical for [*client*] and track it as [*metric*]	"I wish AI could track window-gazing instead of assuming inattention."	Consider [window-gazing] as [typical behavior] and track [gaze pattern]
Track [*behavior*] as engagement for [*client profile*] and adjust tracking	"I wish AI could treat repetitive behaviors as engagement."	Track [repetitive touching] as [engagement], not [disengagement], adjusting for [session context]

5 Discussion: Towards AI-Enhanced Documentation

Prior research has explored AI-based tools that generate progress reports by summarizing text-based session data in educational contexts [26]. Building on this line, this exploratory study investigates the potential and design goals of using vision-language models (VLMs) in supporting special education intervention documentation. We found that SLPs recognized AI's lack the flexibility to meet real-world needs, thus the Prompt Sheet offers practical guidance to help AI engineers embed these principles into prompt design. The identified tasks can extend beyond speech-language pathology to inform research in computer science fields such as computer vision, audio assessment, and recognition. Additionally, we observe that AI can unintentionally reinforce biased outcomes, particularly when conclusions rely on statistical averages that overlook learners deviating from the norm [17]. Therefore, it's necessary for future AIED documentation systems to avoid such biases, especially for nonverbal kids.

A major ethical concern raised by participants was obtaining informed consent for recording intervention sessions, especially in group settings where approval from multiple guardians is required, a challenge echoed in broader critiques of educational surveillance [18]. Recent work highlights the importance of combining regulatory safeguards, ethical frameworks, and technical measures to mitigate such risks [16]. Participants also expressed concern about AI's impact on professional roles, warning that automation could lead to de-skilling or task shifting to non-experts, as seen in other domains [27]. As P9 noted, *"Then why are we needed? Couldn't somebody else just go in there, make up the notes, and let the AI do all the work?"* AI must augment, not replace, SLPs' core responsibilities, including direct engagement, collaboration, and professional growth. Also, future systems should provide ways for SLPs to understand how outputs are generated, assess reliability, and detect hallucinations to build trust in AI-generated documentation. These concerns are especially urgent for students with special needs, who face increased risks of stigma and exclusion; poorly designed AI systems risk reinforcing bias.

Acknowledgments. This work is supported, in part, by the National Science Foundation under Grant 2229873 - 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 sponsors.

References

1. American Speech-Language-Hearing Association: Asha - employment settings for SLPs. https://www.asha.org/students/employment-settings-for-slps/
2. Arakawa, R., Yakura, H.: REsCUE: a framework for real-time feedback on behavioral cues using multimodal anomaly detection. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2019)
3. American Speech Language Hearing Association, et al.: American speech-language-hearing association (2020)
4. Bedmutha, M.S., et al.: ConverSense: an automated approach to assess patient-provider interactions using social signals. In: CHI Conference on Human Factors in Computing Systems (2024)
5. Behn, O., Leyer, M., Iren, D.: Employees' acceptance of AI-based emotion analytics from speech on a group level in virtual meetings. Technol. Soc. **76**, 102466 (2024)

6. Beyan, C., Bustreo, M., Shahid, M., Bailo, G.L., Carissimi, N., Del Bue, A.: Analysis of face-touching behavior in large scale social interaction dataset. In: Proceedings of the 2020 International Conference on Multimodal Interaction (2020)
7. Braun, V., Clarke, V.: Thematic analysis. In: Encyclopedia of Quality of Life and Well-Being Research, pp. 7187–7193. Springer (2024)
8. Desolda, G., Lanzilotti, R., Piccinno, A., Rossano, V.: A system to support children in speech therapies at home. In: Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter, pp. 1–5 (2021)
9. Du, Y., Juefei-Xu, F.: Generative AI for therapy? Opportunities and barriers for ChatGPT in speech-language therapy (2023)
10. Feese, S., Arnrich, B., Tröster, G., Meyer, B., Jonas, K.: Quantifying behavioral mimicry by automatic detection of nonverbal cues from body motion. In: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, pp. 520–525. IEEE (2012)
11. Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., Hussain, A.: Multimodal sentiment analysis: a systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Inf. Fusion **91** (2023)
12. Gillam, S.L., Gillam, R.B.: SKILL narrative: supporting knowledge in language and literacy. Utah State University (2012)
13. Hameed, M.S., et al.: What is the educational value and clinical utility of artificial intelligence for intraoperative and postoperative video analysis? A survey of surgeons and trainees. Surg. Endosc. **37**(12), 9453–9460 (2023)
14. Holstein, K., McLaren, B.M., Aleven, V.: Designing for complementarity: teacher and student needs for orchestration support in AI-enhanced classrooms. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) AIED 2019. LNCS (LNAI), vol. 11625, pp. 157–171. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23204-7_14
15. Jain, J., Yang, J., Shi, H.: VCoder: versatile vision encoders for multimodal large language models. In: IEEE Conference on Computer Vision and Pattern Recognition (2024)
16. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., Qadir, J.: Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput. Biol. Med. **158**, 106848 (2023)
17. Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R.: Discrimination in the age of algorithms. J. Leg. Anal. **10**, 113–174 (2018)
18. Kurni, M., Mohammed, M.S., Srinivasa, K.: Ethics of artificial intelligence in education. In: A Beginner's Guide to Introduce Artificial Intelligence in Teaching and Learning, pp. 213–229. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-32653-0_12
19. Lehmann-Willenbrock, N., Hung, H.: A multimodal social signal processing approach to team interactions. Organ. Res. Methods (2023)
20. Lipkin, P.H., et al.: The individuals with disabilities education act (IDEA) for children with special educational needs. Pediatrics **136**(6), e1650–e1662 (2015)
21. Luckin, R., et al.: Designing educational systems fit for use: a case study in the application of human centred design for AIED. Int. J. Artif. Intell. Educ. **16**(4), 353–380 (2006)
22. Moore, B.J.: Documentation issues. In: Professional Issues in Speech-Language Pathology and Audiology, p. 401 (2019)
23. Mustafa, M.Y.: A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. Smart Learn. Environ. **11**(1), 1–33 (2024)

24. Paloniemi, A., Pulkkinen, J., Kärnä, E., Björn, P.M.: The work of special education teachers in the tiered support system: the Finnish case. Scand. J. Educ. Res. **67**(1), 35–50 (2023)
25. Razi, A., et al.: Deep learning serves traffic safety analysis: a forward-looking review. IET Intel. Transp. Syst. **17**(1), 22–71 (2023)
26. Reddy, S., Fox, J., Purohit, M.P.: Artificial intelligence-enabled healthcare delivery. J. R. Soc. Med. **112**(1), 22–28 (2019)
27. Sambasivan, N., Veeraraghavan, R.: The deskilling of domain expertise in AI development. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI 2022. Association for Computing Machinery (2022)
28. Samrose, S., et al.: MeetingCoach: an intelligent dashboard for supporting effective & inclusive meetings. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2021)
29. Singer, I., Klatte, I.S., de Vries, R., van der Lugt, R., Gerrits, E.: Using co-design to develop a tool for shared goal-setting with parents in speech and language therapy. Int. J. Lang. Commun. Disord. (2022)
30. Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M.J., Er, E., Khalil, M., Akçapınar, G.: Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review. Behav. Inf. Technol., 1–28 (2024)
31. Zheng, Q., et al.: Towards responsible use of large multi-modal AI to analyze human social behaviors. In: Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, pp. 663–665 (2024)
32. Zheng, Q., Rabbani, P., Lin, Y.R., Mansour, D., Huang, Y.: SOAP.AI: a collaborative tool for documenting human behavior in videos through multimodal generative AI. In: Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, pp. 87–90 (2024)