JSLHD 2026. Vol. 35, No. 3. 077-086
ISSN 1226-587X / eISSN 2671-7158
Journal of Speech-Language & Hearing Disorders 언어치료연구

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2026.35.3.077
http://jslhd.org/

Artificial Intelligence in Dysphagia Management: Promise and Doubt

Youngsun Kim[1*], Jaelyn Warnock[2], Molly Taylor[2], Taeok Park[3]

[1] Dept. of Hearing, Speech, and Language Sciences, Ohio University, Associate Professor
[2] Dept. of Hearing, Speech, and Language Sciences, Ohio University, Speech Language Pathology Student
[3] Dept. of Communication Sciences and Disorders, Illinois State University, Associate Professor

Purpose: Artificial intelligence (AI) presents both promising opportunities and significant uncertainties in clinical healthcare. Currently, AI technologies are used across a wide range of health applications such as medical imaging analysis and treatment recommendations. However, there are limited studies for how AI platforms can generate evidence-based recommendations for dysphagia. This study evaluated the appropriateness and accuracy of evidence-based dysphagia management recommendations generated by three AI platforms, ChatGPT, Claude, and DeepSeek.
Methods: Two real clinical case prompts were used: one involving post-stroke dysphagia and the other concerning pediatric dysphagia. Recommendations were analyzed qualitatively for physiological appropriateness and quantitatively for citation accuracy and application.
Results: It was found that all platforms generated some physiologically appropriate recommendations, but frequently misrepresented, inaccurately cited, or inappropriately applied research evidence. On average, only approximately 71% of cited sources across platforms corresponded to existing articles, with accurate citation and appropriate application rates averaging 36% and 43%, respectively.
Conclusions: While each platform provided generally useful information related to dysphagia intervention, the results indicated that clinical decision-making in this domain currently exceeds AI capabilities. The platforms frequently failed to prioritize interventions based on severity or patient-specific context, limiting their clinical applicability. As AI continues to integrate into clinical practice, it is crucial to foster a transparent system, build clinician and patient trust, and ensure robust education regarding the strengths and limitations of these tools.

Keywords: Swallowing, dysphagia, AI, decision making

Correspondence : Youngsun Kim, PhD
E-mail : kimy2@ohio.edu

Received : May 13, 2026
Revision revised : July 06, 2026
Accepted : July 31, 2026

This study did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

ORCID
Youngsun Kim
https://orcid.org/0000-0001-7549-316X
Jaelyn Warnock
https://orcid.org/0009-0003-5906-7672
Molly Taylor
https://orcid.org/0009-0008-9713-4000
Taeok Park
https://orcid.org/0000-0003-0535-8434

Ⅰ. Introduction

Artificial intelligence (AI) has introduced transformative possibilities in healthcare and carries the potential to make services more personalized, predictive, preventative, and interactive (Khan et al., 2023). AI technologies are currently employed across a wide range of healthcare applications, including medical imaging analysis, early disease detection, diagnostic support, treatment recommendations, patient engagement tools, administrative tasks, drug development, and medical image interpretation (Davenport & Kalakota, 2019; National Institutes of Health, 2024; Patel, 2024). Research suggests that AI can match or exceed human performance in select clinical tasks, such as disease diagnosis and tumor detection. Nevertheless, AI is viewed as an augmentative tool for support, rather than a replacement for human clinicians (Davenport & Kalakota, 2019).

Common forms of AI in healthcare include machine learning (ML) and natural language processing (NLP). ML is frequently used to predict treatment protocols, assess disease risks, and detect abnormalities in imaging data based on patient data and clinical context. NLP applications support the generation, interpretation, and analysis of clinical documentation and research, including transcripts, report generation, and conversational systems such as chatbots and virtual assistants (Davenport & Kalakota, 2019). These conversational AI systems, increasingly popular for their accessibility and continuous availability, offer users immediate health information without requiring a clinician's presence (China, 2024; National Institutes of Health, 2024). These systems have the potential to reduce costs and promote accessible, precise medical care.

Despite AI's many benefits, several challenges remain. Barriers include data privacy concerns, data collection bias, ethical concerns, and the lack of validated clinical trials (Alowais et al., 2023; Khan et al., 2023). Many AI tools require large and diverse datasets to function effectively and accurately predict outcomes. However, medical records are confidential, difficult to access, and vulnerable to cybersecurity threats (Khan et al., 2023). Furthermore, bias in data collection that is used to inform model development may result in poorly trained models that learn inaccurate associations between patient characteristics and outcomes. Deep learning models also offer little interpretability for their predictions, resulting in a lack of transparency in their decision-making processes, raising legal and ethical concerns when errors occur. The absence of rigorous validation studies impedes broader clinical adoption of AI.

Public and professional perceptions toward AI remain mixed. While patients often accept AI for general health inquiries, most typically prefer human clinicians for complex medical decisions (Alowais et al., 2023). Healthcare providers express concerns about AI's potential impacts on clinician-patient relationships, data privacy, and job security (Davenport & Kalakota, 2019; Robeznieks, 2024). Nonetheless, many clinicians recognize AI's utility in improving diagnostic accuracy, efficiency, and patient outcomes, though human input remains invaluable (Davenport & Kalakota, 2019; Robeznieks, 2024).

Dysphagia, a condition characterized by difficulty swallowing, often results from structural or neurological impairments and is particularly common among young children and older adults. It increases the risk of aspiration, pneumonia, malnutrition, and dehydration, frequently contributing to prolonged hospitalizations and increased mortality risk (Girardi et al., 2023; Kim et al., 2024). Despite its prevalence, dysphagia frequently goes undiagnosed (Gugatschka et al., 2024). Although videofluoroscopic swallow studies (VFSS) are considered the "gold standard" for diagnosis, they are resource-intensive, reliant on clinician expertise, and cannot be done frequently due to repeated radiation exposure risks. Fiberoptic endoscopic evaluation of swallowing (FEES) and other diagnostic methods also face challenges in tracking long-term changes in patient condition (Kim et al., 2024).

Recent studies have explored AI-enhanced approaches to improve dysphagia diagnosis. AI models have demonstrated rapid and highly accurate interpretation of VFSS data for the diagnoses, with some platforms achieving 97.2% accuracy in classifying image frames and 93.2% accuracy in video files when detecting aspiration (Jeong et al., 2024). However, these models have failed to identify the underlying physiological deficits, limiting their usefulness in treatment planning. AI-supported evaluations in FEES have improved accuracy among inexperienced clinicians, and novel approaches, such as voice analysis-based dysphagia screening, show potential for at-home monitoring (Imaizumi et al., 2024; Kim et al., 2024). Despite encouraging results, clinical reasoning and patient-specific interventions have been shown to remain domains where human expertise is indispensable.

Broadly, AI has shown promise in enhancing diagnostic accuracy, expediting swallowing assessments, and facilitating targeted interventions, regardless of a clinician's skill level. It is even suggested that, with robust validation and interdisciplinary collaboration, AI-enhanced VFSS may become the "cornerstone" of future dysphagia management (Girardi et al., 2023). Although AI platforms have been studied in dysphagia screening and instrumental evaluations, less is known about whether general AI platforms can generate evidence-based recommendations for clinical management. This study explored how three AI platforms, ChatGPT, Claude, and DeepSeek, responded to real clinical prompts involving adult and pediatric dysphagia. The research question was whether the AI platforms developed dysphagia management and treatment recommendations that have citation accuracy, clinical appropriateness, and alignment with evidence-based practice.

II. Methods

1. Three AI platforms

We chose to investigate how AI is defined, which AI platforms are most widely used, and AI's implications within healthcare. After reviewing various platforms, three were selected for this study: ChatGPT, Claude, and DeepSeek. We utilized the paid version of two AI platforms, ChatGPT-4o and Claude 3.5 Sonnet, while

DeepSeek was free.

ChatGPT, developed by OpenAI, is a generative AI model based on the GPT architecture designed for natural language processing and human-like text generation (Gregersen, 2024). Released by OpenAI in 2022 and built upon earlier GPT-2 and GPT-3 models, ChatGPT quickly gained popularity for its human-like, contextually appropriate responses (Ray, 2023). The GPT-4 version has demonstrated diagnostic accuracy comparable to human physicians in medical applications (Reis et al., 2024). Despite its strengths, ChatGPT is known to 'hallucinate,' or generate plausibly sounding but inaccurate information, like other AI platforms, posing concerns for its use in scientific and medical contexts (Alkaissi & McFarlane, 2023; Gregersen, 2024).

Claude is a large language model developed by Anthropic that employs deep learning and constitutional AI to align outputs with human values (Priyanshu et al., 2024). Unlike models partnered with search engines, Claude is self-contained, capable of document processing, image analysis, and basic coding. Claude has demonstrated powerful performance in complex decision-making tasks. It outperformed ChatGPT-4o in surgical, chemotherapy, and radiation therapy recommendations for head and neck cancer using the artificial intelligence performance instrument (AIPI) scale, and it scored similarly in clinical explanation and summarization (Schmidl et al., 2024). Additional comparisons in complex medical scenarios also favored Claude over ChatGPT, Bard, and Perplexity (Uppalapati & Nag, 2024).

DeepSeek is a China-based AI company offering flagship large language models. DeepSeek allows researchers and developers to study, modify, and commercially apply its models (Bhattacharya et al., 2025; Gibney, 2025). DeepSeek R1 uses a "chain of thought" reasoning method, generating step-by-step responses that enhance performance on complex tasks such as coding, mathematics, and chemistry, which is on par with top models like ChatGPT-o1 (Gibney, 2025; Poo, 2025). Founded in 2023, the company quickly gained attention for producing competitive models at a fraction of the cost of leading alternatives, while offering lower usage fees to developers (Poo, 2025). In addition, DeepSeek offers unlimited usage, large file upload capacity (up to 100 MB), real-time web search across over 100 sources, and productivity-enhancing customization features such as prompt templates (Bhattacharya et al., 2025).

2. Experimental procedures: Prompt development

Two clinical case studies were developed using the Ohio University Swallowing Research Laboratory (SRL) database of videofluoroscopic swallowing studies (VFSS). All collected at the SRL were approved by the Internal Review Board (IRB) at Ohio University. Demographic information, clinical exam findings, and physiological findings of VFSS were described in the prompt. The following represents the cases used:

- Stroke survivor with dysphagia: John, a 75-year-old man, had multiple bilateral strokes three days ago. A bedside evaluation found poor nutrition, reduced tongue strength, dysphonia, and a wet voice quality. The VFSS results showed that John had difficulty swallowing both thick and thin liquids, requiring more effort with the thick liquids. During swallows with thin liquids, the VFSS revealed penetration, silent aspiration, and residue in the tongue base, pharyngeal wall, vallecula, and pyriform sinus. Delayed pharyngeal swallow, reduced laryngeal closure, and reduced UES opening were also observed. John attempted to clear the bolus with repeated swallows.
- Pediatric infant with disordered swallowing: An 8-month-old full-term male presented with hypoxic-ischemic encephalopathy (HIE). He showed exterior signs of aspiration: he widened his eyes while feeding, had a lower blood oxygen content, and was seen grunting. Upon completing a VFSS, premature spillage, pharyngeal delay, aspiration, rocking motion of tongue (swallowing apraxia), weak tongue, and poor coordination were observed.

Prompt to AI platforms: The investigators purchased subscriptions to two AI platforms, utilizing the free version of another, and delivered identical prompts to all AI platforms on the same data: February 21, 2025. Responses of AI platforms were recorded in Microsoft Word for further analysis. Two investigators at the Swallowing Research Laboratory, who took the graduate-level dysphagia class and previously participated in the VFSS research projects, reviewed the accuracy and relevance of references and recommendations based on the current literature review, respectively. After each review, the principal and the other two investigators discussed and solved the differences.

Ⅲ. Results

All three AI platforms proposed dysphagia management plans for both adult and pediatric patients. AI responses included compensatory strategies, rehabilitation exercises, and caregiver education with corresponding references, some of which are physiologically reasonable and consistent with clinical experts.

For Case 1, an adult with dysphagia post-stroke, all AI platforms offered multiple management recommendations targeting key physiologic deficits observed in the case: poor tongue coordination, delayed pharyngeal swallow, reduced laryngeal closure, and reduced UES opening. See Table 1 for a summary of the dysphagia management recommendations for all three AI platforms for Case 1.

ChatGPT produced some responses that were physiologically aligned, especially when follow-up prompts directed the model to address specific impairments. Of the seven sources cited, only two (29%) were real, and none were both accurately cited and appropriately applied. For example, the platform cited Robbins et al. (2005). To support tongue strengthening and resistance exercises: 'Exercises using resistance (e.g., with devices such as the Iowa Oral Performance Instrument) target the tongue's strength and endurance, thereby improving bolus propulsion, and structured resistance training protocols have been shown to improve both strength and coordination of the tongue, which contributes to more effective bolus propulsion and reduced pharyngeal residue.' The cited article was fabricated and did not exist; however, a similarly titled publication by the same author supported the intervention. Further, Robbins et al. (2008) were referenced to justify the Mendelsohn Maneuver for delayed pharyngeal swallow and reduced laryngeal closure. The response read, 'By teaching the patient to sustain laryngeal elevation at the peak of the swallow, this strategy prolongs airway closure and improves the coordination of swallowing timing (Robbins et al., 2008).' Although the source was real, the citation was inaccurate, and the described content was absent from the article. Another fabricated example was Cohen and Kim (2007) to support sensory cueing for swallowing apraxia and poor coordination: 'Providing strong, consistent cues and even sensory stimuli (e.g., sour bolus) may help improve the timing and coordination of the swallow.' The article cited does not exist. Overall, while ChatGPT's responses contained seemingly authentic recommendations, the outputs were undermined by unreliable evidence.

Claude generated the most comprehensive response for Case 1, listing over treatment strategies and citing 38 sources, 36 (95%) of which were real, though only 33% were both accurately cited and appropriately applied. Some recommendations were physiologically relevant but lacked alignment with the described physiological impairment. A clear example of misapplication was the citation of Park et al. (2012): 'When applied specifically to tongue musculature, neuromuscular electrical stimulation (NMES) can address both poor coordination and apraxic movements by activating and reeducating muscle patterns.' Although valid and correctly cited, this article discussed NMES targeting infrahyoid muscles, not the tongue musculature. Following a description of oral myofunctional therapy (OMT) using a different source, Claude referenced Hägg and Anniko (2010) to support the use of OMT for poor tongue coordination: 'Research demonstrates improved tongue movement patterns and reduced compensatory behaviors in patients with neurogenic dysphagia.' While this is a legitimate publication with an accurate citation, the cited article does not contain information regarding this therapeutic approach. Conversely, tongue pressure resistance training was supported with accurate and appropriately cited references from Steele et al. (2016) and Robbins et al. (2007). Claude referenced the indicated articles for using devices like the Iowa Oral Performance Instrument (IOPI) for progressive resistance training to improve tongue strength and coordination. This specifically addresses both poor tongue coordination and the rocking motion by providing targeted feedback about tongue movement patterns (Steele et al., 2016). Studies show improvement in tongue motion control and coordination during the oral phase of swallowing (Robbins et al., 2007). While Claude's responses were detailed and rich with citations, over two-third of the references were either inaccurately linked to articles or misused.

DeepSeek provided fewer, more concise responses for Case 1. While some recommendations (e.g., Shaker exercise, tongue strengthening, thermal stimulation) aligned with the physiological deficits, the platform lacked elaboration and clinical rationale. All eight cited sources were real, but only 50% of those were accurately cited and appropriately applied. For example, Rosenbek and Donovan (2006) was cited to support thermal-tactile stimulation: 'Rosenbek and Donovan (2006) demonstrated that thermal-tactile stimulation improves swallow initiation

Table 1. Summary of the dysphagia management recommendations of all three AI platforms for Case 1

Platform	ChatGPT	Claude	DeepSeek
Summary	Some physiologically appropriate strategies (e.g., effortful swallow, Mendelsohn) Improved responses with targeted prompts (e.g., "Please address tongue apraxia")	Most expansive response with 10+ treatment strategies Frequently over-prescribed interventions without clear physiologic relevance	Concise recommendations, some aligned with physiological deficits Minimal rationale for suggested treatments
Citation accuracy	Cited 7 sources; 2 were real 0 correctly cited For example, cited Robbins et al. (2005), for tongue exercises with resistance. This article did not exist, though a similarly titled one supported the recommendation	Cited 38 sources; 36 were real and 12 were correctly cited For example, cited Park et al. (2012) to support tongue coordination treatment, but the article focused on NMES for infrahyoid muscles; real source, but misused	Cited 8 sources; all were real 4 accurately cited For example, cited Rosenbek and Donovan (2006) to recommend thermal-tactile stimulation. The article was real, but the date was incorrect, and the technique wasn't mentioned
Overall assessment	Lacked depth and citation accuracy; included misleading or unsupported evidence	Provided exhaustive recommendations for each physiology, but lacked citation accuracy, and included unsupported evidence	Provided best citation accuracy but included misrepresentation of sources and lacked elaboration

Table 2. Summary of the dysphagia management recommendations of all three AI platforms for Case 2

Platform	ChatGPT	Claude	DeepSeek
Summary	Some physiologically appropriate pediatric recommendations (e.g., pacing, side-lying) Specific instructions for both caregivers and clinicians	Detailed responses, some physiologically appropriate	Limited, general recommendations based on physiology
Citation accuracy	Cited 8 sources; 7 were real 4 accurately cited 6 matched the cited content For example, cited Krüger et al. (2019) to support tongue strengthening exercises; the article was real and related to HIE, but did not include this intervention	Cited 6 sources; 2 were real 0 accurately cited 1 matched cited content For example, cited Arverdson et al. (2020) for oral-motor interventions, but the article was published in 2010, and its content did not fully support the claim	Cited 5 sources; 4 were real 3 accurately cited 2 matched the cited content For example, cited Clark et al. (2009) twice to support paced feeding and bottle modifications, but the article did not exist; a similarly titled article by Clark et al. (2009) included only adults
Overall assessment	Improved citation use compared to Case 1, but still included misapplied evidence	Elaborate recommendations and instructions, but several inaccurate or unsupported	Lacked elaboration and citation integrity; included unsupported information

and coordination.' This article exists but was cited with an incorrect data, and the intervention was not discussed in the article. It also cited Robbins et al. (2007) in support of using smaller bolus sizes (e.g., smaller bites or sips) to improve tongue coordination and reduced residue. While this article exists, it uses the incorrect data and digital object identifier (DOI) in the full citation. More notably, the article discussed reduced residue due to smaller bolus sizes, without mentioning coordination. One correctly cited and appropriately applied article was Shaker et al. (2002), to support the Shaker Exercise to improve UES opening: 'This involves lifting the head while lying flat to strengthen the suprahyoid muscles, improving UES opening and bolus clearance.' This article was real, accurately cited, and used appropriately (Table 1).

For Case 2, an infant with dysphagia secondary to HIE, all platforms recognized the relevance of common management strategies such as thickened liquids, pacing techniques, and positioning adjustments. See Table 2 for a summary of the dysphagia management recommendations of all three AI platforms for Case 2. It is important to note that pediatric dysphagia management is highly context-dependent and may involve discussion of risks, age, medical diagnosis, and stability. Any recommendations were appropriate under clinical or caregiver supervision.

ChatGPT provided moderately detailed, caregiver and clinician-focused guidance. Of the eight sources cited, seven (88%) were real, with 85% appropriately applied and 57% accurately cited. Citation issues persisted, such as citing Krüger et al. (2019) to support daily tongue strengthening exercises, despite the article not addressing that intervention. Similarly, Fucile et al. (2011) were cited to support its claims regarding sensory stimulation for weak tongue and poor coordination: 'Sensory stimulation can improve oral motor planning and coordination, contributing to safer swallowing.' While this cited study is legitimate, the correct publication date is 2002, not 2011. ChatGPT also cited an article by Shaker (2017) to support environmental modifications for swallowing apraxia and poor coordination: 'A controlled, quiet environment supports better oral coordination by minimizing sensory overload.' Rather than the article this claims to reference (*Cue-based feeding in the neonatal intensive care unit (NICU): Moving from a volume-driven to infant-driven approach*), this information may be referencing a different article from 2013, similarly titled with the same author (*Cue-based feeding in the NICU: using the infant's communication as a guide*), which would align with the information cited by ChatGPT. Among the articles correctly cited and appropriately used was Foster et al. (2016) to support the use of non-nutritive sucking (NNS) for swallowing apraxia and poor coordination: "NNS has been shown to stabilize breathing, improve sucking coordination, and lay the groundwork for efficient feeding." This article was genuine, correctly cited, and was used accurately by ChatGPT.

Claude also produced detailed, caregiver-oriented recommendations, some of which were appropriate for infants. However, citation integrity was notably low: of the six sources cited, only two (33%) were real, none were accurately cited, and only one was appropriately applied. Additionally, citations were often grouped at the end of explanations under 'Research Support,' leading to vague attribution. Claude misattributed the publication year of an article by Arvedson et al. (2010) as 2020 and used it to support instructions for the Modified Beckman Oral Motor Protocol; the information within the article does not align with what is cited by Claude. It is also cited in Goldfield et al. (2020) as evidence supporting caregiver instructions for pacing techniques to address premature spillage and pharyngeal delay. However, this article was fabricated. While there is a similarly titled 2010 publication by the same author, it also makes no mention of pacing techniques. Harris and Johnson (2019) was cited to support instructions for a modified bottle system to address the infant's weak tongue, but this article was hallucinated. Among the correctly used articles is Park et al. (2019), to support a modified side-lying position to address premature spillage. Despite the incorrect data used in the citation (2019 instead of 2024), this article was real and contained the information used by Claude.

DeepSeek's response to the pediatric case was the least robust. It offered general but vague suggestions and cited only five sources. Four (80% were real, three of which were accurately cited, and two were appropriately applied. For example, it cited Clark et al. (2009), to support that 'tongue strengthening exercises are effective in improving swallowing function in infants with weak tongue muscles.' This article does not exist. While a similarly titled publication by the same author exists, it only discusses tongue-strengthening exercises in adults and makes no mention of specialized nipple improving feeding outcomes for individuals with weak tongues. It also cited Lefton-Grief and McGrath-Morrow (2007) to support pacing techniques: 'Pacing techniques are effective in reducing aspiration risk in infants with dysphagia.' The referenced article mentions various management strategies, but pacing is not one of them. Among the correctly used articles is Arvedson and Brodsky (2002), cited twice to address premature spillage and pharyngeal delay using different techniques: chin tuck and thickened liquids. The response quotes the study accurately: This approach is supported by studies showing that chin tuck and thickened liquids can improve swallowing safety in infants with dysphagia, and upright positioning.' It also includes another relevant excerpt: 'Upright positioning is a well-documented strategy to improve swallowing safety in infants with dysphagia.' This source was legitimate, correctly cited, and appropriately used.

Ⅳ. Discussion

The findings of this study highlight notable variability in the accuracy and appropriateness of AI-generated dysphagia management recommendations. While all three AI platforms, ChatGPT, Claude, and DeepSeek, were capable of generating responses that demonstrated general awareness of dysphagia management techniques, the degree of citation accuracy, evidence alignment, clinical appropriateness, and responsiveness 1123to prompt specificity was highly variable. This variation reinforces the importance of appraisal by qualified clinicians before integrating AI-generated content into clinical decision-making.

Claude produced extensive treatment strategies across swallowing impairments, often citing many sources, particularly in the adult clinical case study. While some of these citations referred to real publications, the platform frequently misrepresented the content of the sources, significantly undermining its reliability. In some instances, references were attributed to statements they did not contain. The expansive recommendations suggest wide-ranging content generation; however, the frequent misapplication of literature raises concerns about the platform's appropriateness for clinical guidance.

ChatGPT demonstrated strength in formatting and clarity. Its treatment recommendations were often well-organized, with clear rationales, particularly in the pediatric case. However, if frequently struggled with citation accuracy, misdating sources, or referencing articles that did not support the presented claims, especially in the adult case. Citation fabrication, or hallucination, was also observed, consistent with other published findings, often using real author names with previous relevant publications, plausible titles, and journal formatting. Publication dates were the most erroneous, with 47% incorrect, and reference titles the least, with 34% incorrect (Gravel et al., 2023). ChatGPT often required targeted follow-up prompts to produce physiologically aligned recommendations, including that its initial outputs may not always reflect best-practice dysphagia management recommendations.

DeepSeek displayed variable performance between the two cases. In the pediatric case, it provided minimal citations, with some fabricated or not directly relevant to pediatric dysphagia. In contrast, its performance with the adult case showed improved citation relevance and the inclusion of some appropriate management strategies. Compared to the other platforms, DeepSeek's responses tended to be shorter and less nuanced, which may limit their clinical utility despite relatively higher citation accuracy in certain contexts.

In Case 1, the adult case, DeepSeek performed strongest in citation realism, with 100% of sources being real and 50% accurately cited and correctly applied. Claude cited mostly real sources (95%) but had low rates of accurate citation and correct application (33% each). ChatGPT performed weakest in this case, with only 29% real sources and no accurately cited or applied references. In Case 2, the pediatric case, ChatGPT's performance improved substantially, with 88% of sources being real, 57% accurately cited, and 85% appropriately applied. DeepSeek maintained moderately strong performance with 80% real sources, 75% accurate citations, and 40% correct application. Claude's citation quality declined sharply in the pediatric case, with only 33% real sources, none accurately cited, and just 50% appropriately applied.

These patterns indicate several key contrasts on a case-by-case basis. ChatGPT was the least reliable in the adult case (Case 1) but achieved the highest percentage of accurate and correctly applied citations in the pediatric case. While in some instances, misapplying and inaccurately citing real references, DeepSeek demonstrated the most consistent citation realism across both cases. Claude, despite producing the largest number of citations, exhibited variable citation realism and frequent inaccuracies in citations and applications. It is relatively easy to ensure the accuracy of citations to research articles. However, clinical appropriateness in evidence-based practice needs to have clinical judgments and expertise. It is important to develop more clinical research data and guidelines in dysphagia management.

Similar patterns have been documented in other studies, where ChatGPT was found to provide more accurate responses to definitional questions (8 out of 12) but was less reliable for reasoning and methodological inquiries. ChatGPT also included citations that did not exist in the literature (Shiferaw et al., 2024). Bhattacharyya et al. (2023) reported that ChatGPT references were 47% fabricated, 46% authentic but inaccurate, and only 7% authentic and accurate. The likelihood of the fabricated references was significantly different based on prompt variations, yet the rate of authentic and accurate references remained low in all cases. These inaccuracies highlight a critical risk: AI-generated recommendations could mislead clinicians and delay intervention, potentially causing them to miss the critical window for recovery or cause harm to patients.

Across all platforms, there were recurring limitations in citation reliability, with frequent instances of fabricated and misused clinical reasoning as well as improperly formatted references. While many recommendations appeared superficially relevant to dysphagia management, none of the AI platforms consistently demonstrated structured clinical reasoning or prioritized treatment strategies in alignment with physiological principles. Inconsistencies were also found in recommendations for identical impairments. For example, in addressing poor tongue coordination in Case 1, all three platforms suggested the effortful swallow, targeting pharyngeal muscles, but their additional recommendations varied. ChatGPT proposed cueing and sensory enhancement, and volume control. DeepSeek recommended tongue strengthening and range-of-motion exercises, the Masako maneuver, thermal stimulation, chin tuck, and diet modification. Claude suggested Orofacial Myofunctional Therapy (OMT), sEMG, NME, oral apraxia treatment, and prompt therapy, some of which are unrelated to tongue coordination deficits. These discrepancies are similar to findings from other investigations of AI in healthcare.

To move forward responsibly, public dialogue about AI in healthcare must grow to reflect both patient and professional perspectives. As AI becomes more integrated into clinical practice, trust-building, transparent processes, and robust patient education will be essential (Khan et al., 2023; Kiseleva et al., 2022).

V. Conclusion

While all AI platforms may generate treatment recommendations for patients with dysphagia, the accuracy and integrity of the proposed interventions are doubtful (Alkaissi & McFarlane, 2023). In this study, all three AI platforms, ChatGPT, Claude, and DeepSeek, proposed a comparable number of interventions, yet their recommendations varied in content and quality. Although each platform provided research references, many were inaccurate, misleading or fabricated, which is a phenomenon known as “hallucination,” where AI systems produce misleading or inaccurate information, particularly in citations (Gregersen, 2024).

AI platforms may be useful for brainstorming general intervention tools, generating patient education drafts, or reviewing intervention techniques. However, AI should not be used as an independent source for patient-specific decision-making in dysphagia management. AI platforms can support clinical thinking in dysphagia management by suggesting common evidence-based treatments. However, current models frequently misattribute evidence, lack depth in interpreting clinical scenarios, recommend techniques without appropriate clinical justification, and fail to prioritize interventions based on severity or patient-specific context. As such, verification of AI-generated recommendations through expert clinical judgment and evidence-based practice is essential.

While AI platforms can serve as valuable tools for accessing updated clinical information, they should not replace clinician expertise, evidence-based practice, or individualized patient care. The development of clear guidelines for the safe and appropriate integration of AI into healthcare practice will be critical to maximizing benefits while minimizing risks.

Reference

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cuerus, 15*(2), e35179. doi:10.7759/cureus.35179

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., . . . Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education, 23,* 689. doi:10.1186/s12909-023-04698-z

Arvedson, J. C., & Brodsky, L. (2002). *Pediatric swallowing and feeding: Assessment and management* (2nd ed.). Singular Publishing Group / Thomson Learning.

Arvedson, J., Clark, H., Lazarus, C., Schooling, T., & Frymark, T. (2010). Evidence-based systematic review: Effects of oral motor interventions on feeding and swallowing in preterm infants. *American Journal of Speech-Language Pathology, 19*(4), 321-340. doi:10.1044/1058-0360(2010/09-0067)

Bhattacharya, K., Bhattacharya, S., Bhattacharya, N., & Bhattacharya, N. (2025). Deepseek versus ChatGPT in surgical practice. *Indian Journal of Surgery, 87,* 845-848. doi:10.1007/s12262-025-04368-y

Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus, 15*(5), e39238. doi:10.7759/cureus.39238

China, C. R. (2024, January 10). *Artificial intelligence advantages & disadvantages.* IBM. https://www.ibm.com/think/insights/artificial-intelligence-advantages-disadvantages

Clark, H. M., O'Brien, K., Calleja, A., & Corrie, S. N. (2009). Effects of directional exercise on lingual strength. *Journal of Speech, Language, and Hearing Research, 52*(4), 1034-1047. doi:10.1044/1092-4388(2009/08-0062)

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal, 6*(2), 94-98. doi:10.7861/futurehosp.6-2-94

Foster, J. P., Psaila, K., & Patterson, T. (2016). Non-nutritive sucking for increasing physiologic stability and nutrition in preterm infants. *Cochrane Database of Systematic Reviews, 2016*(10), CD001071. doi:10.1002/14651858.CD001071.pub3

Fucile, S., Gisel, E. G., McFarland, D. H., & Lau, C. (2011). Oral and non-oral sensorimotor interventions enhance oral feeding performance in preterm infants. *Developmental Medicine & Child Neurology, 53*(9), 829-835. doi:10.1111/j.1469-8749.2011.04023.x

Gibney, E. (2025). China's cheap, Open AI model DeepSeek thrills scientists. *Nature, 638,* 13-14. doi:10.1038/

d41586-025-00229-6

Girardi, A. M., Cardell, E. A., & Bird, S. P. (2023). Artificial intelligence in the interpretation of videofluoroscopic swallow studies: Implications and advances for speech-language pathologists. *Big Data and Cognitive Computing, 7*(4), 178. doi:10.3390/bdcc7040178

Gravel, J., D'Amours-Gravel, M., & Osmanlliu, E. (2023). Learning to fake it: Limited responses and fabricated references provided by ChatGPT for medical questions. mayo clinic proceedings. *Digital Health, 1*(3), 226-234. doi:10.1016/j.mcpdig.2023.05.004

Gregersen, E. (2024, October 16). *ChatGPT*. Encyclopædia Britannica. https://www.britannica.com/technology/ChatGPT

Gugatschka, M., Egger, N. M., Haspl, K., Hortobagyi, D., Jauk, S., Feiner, M., & Kramer, D. (2024). Clinical evaluation of a machine learning-based dysphagia risk prediction tool. *European Archives of Oto-Rhino-Laryngology, 281,* 4379-4384. doi:10.1007/s00405-024-08678-x

Hägg, M., & Anniko, M. (2010). Influence of lip force on swallowing capacity in stroke patients and in healthy subjects. *Acta Oto-Laryngologica, 130*(11), 1204-1208. doi:10.3109/00016481003745550

Imaizumi, M., Weng, W., Zhu, X., & Murono, S. (2024). Effectiveness of FEES with artificial intelligence-assisted computer-aided diagnosis. *Auris Nasus Larynx, 51*(2), 251-258. doi:10.1016/j.anl.2023.11.004

Jeong, C.-W., Lee, C.-S., Lim, D.-W., Noh, S.-H., Moon, H.-K., Park, C., & Kim, M.-S. (2024). The development of an artificial intelligence video analysis-based web application to diagnose oropharyngeal dysphagia: A pilot study. *Brain Sciences, 14*(6), 546. doi:10.3390/brainsci14060546

Khan, B., Fatima, H., Qureshi, A., Kumar, S., Hanan, A., Hussain, J., & Abdullah, S. (2023). Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomedical Materials & Devices, 1,* 731-738. doi:10.1007/s44174-023-00063-2

Kim, J. M., Kim, M. S., Choi, S. Y., & Ryu, J. S. (2024). Prediction of dysphagia aspiration through machine learning-based analysis of patients' postprandial voices. *Journal of NeuroEngineering and Rehabilitation, 21,* 43. doi:10.1186/s12984-024-01329-6

Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in healthcare as a multilayered system of accountabilities: Between legal requirements and technical limitations. *Frontiers in Artificial Intelligence, 5,* 879603. doi:10.3389/frai.2022.879603

Krüger, E., Kritzinger, A., & Pottas, L. (2019). Oropharyngeal dysphagia in breastfeeding neonates with hypoxic-ischemic encephalopathy on therapeutic hypothermia. *Breastfeeding Medicine, 14*(10), 718-723. doi:10.1089/bfm.2019.0048

Lefton-Greif, M. A., & McGrath-Morrow, S. A. (2007). Deglutition and respiration: Development, coordination, and practical implications. *Seminars in Speech and Language, 28*(3), 166-179. doi:10.1055/s-2007-984723

National Institutes of Health. (2024, June 17). *Artificial intelligence and your health*. National Institutes of Health. https://newsinhealth.nih.gov/2024/01/artificial-intelligence-your-health

Park, J., Thoyre, S., Smallcomb, J., McTernan, M., & Kneeland, T. (2024). Biobehavioral efficacy of the elevated side-lying position for feeding preterm infants: Study protocol. *Journal of Advanced Nursing*. Advance online publication. doi:10.1111/jan.16444

Park, J. W., Kim, Y., Oh, J. C., & Lee, H. J. (2012). Effortful swallowing training combined with electrical stimulation in post-stroke dysphagia: A randomized controlled study. *Dysphagia, 27*(4), 521-527. doi:10.1007/s00455-012-9403-3

Patel, M. (2024, June 28). What is the role of generative AI in healthcare and medicine industry? *Linearloop*. https://www.linearloop.io/blog/role-of-generative-ai-in-healthcare-and-medicine

Poo, M. (2025). Reflections on DeepSeek's breakthrough. *National Science Review, 12*(3), nwaf044. doi:10.1093/nsr/nwaf044

Priyanshu, A., Maurya, Y., & Hong, Z. (2024). *AI governance and accountability: An analysis of Anthropic's Claude*. arXiv:2407.01557. doi:10.48550/arXiv.2407.01557

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3,* 121-154. doi:10.1016/j.iotcps.2023.04.003

Reis, M., Reis, F., & Kunde, W. (2024). Influence of believed AI involvement on the perception of digital medical advice. *Nature Medicine, 30,* 3098-3100. doi:10.1038/s41591-024-03180-7

Robbins, J., Kays, S. A., Gangnon, R. E., Hind, J. A., Hewitt, A. L., Gentry, L. R., & Taylor, A. J. (2007). The effects of lingual exercise in stroke patients with dysphagia. *Archives of Physical Medicine and Rehabilitation, 88*(2), 150-158. doi:10.1016/j.apmr.2006.11.002

Robbins, J., Butler, S. G., Daniels, S. K., Gross, R. D., Langmore, S., Kazarus, C. K., . . . Rosenbek, J. (2008). Swallowing and dysphagia rehabilitation: Translating principles of neural plasticity into clinically oriented evidence. *Journal of Speech, Language, and Hearing Research, 51*(1), S276-S300. doi:10.1044/1092-4388(2009/021)

Robeznieks, A. (2024, January 12). *Big majority of doctors see upsides to using health care AI*. American Medical Association. https://www.ama-assn.org/practice-management/digital/big-majority-doctors-see-upsides-using-health-care-ai#:~:text=Also%2C%2056%25%20of%20respondents%20thought,and%2041%25%20with%20patient%20privacy

Rosenbek, J. C., & Donovan, N. J. (2006). Plasticity in the neural pathways for swallowing: The role in rehabilitation of dysphagia. *Textbook of Neural Repair and Rehabilitation, 2,* 356-367.

Schmidl, B., Hütten, T., Pigorsch, S., Stögbauer, F., Hoch, C. C., Hussain, T., . . . Wirth, M. (2024). Assessing the use of the novel tool Claude 3 in comparison to ChatGPT 4.0 as an artificial intelligence tool in the diagnosis and therapy of primary head and neck cancer cases. *European Archives of Oto-Rhino-Laryngology, 281,* 6099-6109. doi:10.1007/s00405-024-08828-1

Shaker, C. S. (2017). Infant-guided, co-regulated feeding in the neonatal intensive care unit. Part II: Interventions to promote neuroprotection and safety. *Seminars in Speech and Language, 38*(2), 106-115. doi:10.1055/s-0037-1599108

Shaker, R., Easterling, C., Kern, M., Nitschke, T., Massey, B., Daniels, S., . . . Dikeman, K. (2002). Rehabilitation of swallowing by exercise in tube-fed patients with pharyngeal dysphagia secondary to abnormal UES opening. *Gastroenterology, 122*(5), 1314-1321. doi:10.1053/gast.2002.32999

Shiferaw, M. W., Zheng, T., Winter, A., Mike, L. A., & Chan, L.-N. (2024). Assessing the accuracy and quality of artificial intelligence (AI) chatbot-generated responses in making patient-specific drug-therapy and healthcare-related decisions. *BMC Medical Informatics and Decision Making, 24,* 404. doi:10.1186/s12911-024-02824-5

Steele, C. M., Bayley, M. T., Peladeau-Pigeon, M., Nagy, A., Namasivayam, A. M., Stokely, S. L., & Wolkin, T. (2016). A randomized trial comparing two tongue-pressure resistance training protocols for post-stroke dysphagia. *Dysphagia, 31*(3), 452-461. doi:10.1007/s00455-016-9699-5

Uppalapati, V., & Nag, D. (2024). A comparative analysis of AI models in complex medical decision-making scenarios: Evaluating ChatGPT, Claude AI, Bard, and Perplexity. *Cureus, 16*(1), e52485. doi:10.7759/cureus.52485