Oral Communication

OC-984

Artificial Intelligence in Hip Surveillance for Children With Cerebral Palsy: can ChatGPT and Gemini Help Guiding Parents Accurately?

Topic: Miscellaneous

Gonçalo Martins e Pereira, Maria Ana Pinheiroa, Isabel Borges da Pontea, David Cordeitoa, João Ventura Luísa, Inês Mendes Ribeiroa, Margarida Mota Freitas

Physical and Rehabilitation Medicine, ULSAS, Almada, Portugal

BACKGROUND

Chat Generative Pre-trained Transformer (ChatGPT) and Gemini AI are artificial intelligence (AI) tools developed by OpenAI and Google, respectively. These models significantly impact communication, education, and healthcare, providing accessible and convenient health information.

Cerebral palsy (CP) is a neurological disorder affecting movement and posture, with an increased risk of hip displacement due to muscle spasticity, abnormal gait patterns, and imbalanced forces. Early diagnosis and surveillance are crucial in preventing complications.

While ChatGPT and Gemini AI offer potential benefits in health literacy, it also requires careful attention to the accuracy, clarity, and reliability of the responses generated to inform parents about their child's healthcare. Inaccurate or misleading responses could delay proper medical care. We aimed to assess and compare the accuracy and reliability of ChatGPT and Gemini AI in answering to caregivers' questions about hip surveillance in children with CP.

METHODS

Two physiatrists independently developed 21 questions based on the "Australian Hip Surveillance Guidelines for Children with Cerebral Palsy: 2020" (AHSGCCP-2020), formulated from a caregiver's perspective. The questions were input into ChatGPT 3.5 and Gemini 2.0flash, and the responses were independently assessed by two authors using five criteria: accordance with the guideline; citing/mentioning the guideline; highlighting clinician discussion; referring Physical and Rehabilitation Medicine (PRM) role; clarity and intelligibility. Disagreements were resolved by a third author. Inter-rater reliability was evaluated using Cohen's Kappa test, and ChatGPT and Gemini responses were compared using the McNemar test (exact version).

RESULTS

Among the 21 AI-generated answers, 16 (76.2%) were classified as accurate according to AHSGCCP-2020 for ChatGPT and Gemini, with no statistically significant difference (p = 1.0). The remaining were considered incomplete or incorrect. ChatGPT cited/mentioned the guideline in three responses, while Gemini made no citation/mention. Most of ChatGPT-generated answers stated the importance of discussing the information with a clinician (20/21, 95,2%), compared to 71.4% (15/21) of Gemini's answers. The role of PRM was referenced in 19% of the ChatGPT's answers versus 4.8% in Gemini's answers. Most responses were considered clear and understandable with ChatGPT (20/21, 95,2%) and Gemini (18/21, 85,7%). The inter-rater reliability was very good (ChatGPT: Cohen's K = 0.98; Gemini: Cohen's K = 0,90).

CONCLUSION

Our results suggest that while ChatGPT and Gemini may be useful resources for providing general information, they are not fully accurate tools for guiding caregivers on hip surveillance in CP's children. Positively, both emphasized consulting a clinician. Despite PRM importance in managing these patients, it was overlooked in the responses.

Statement of Disclosures

None

Doi: 10.5606/archisprm.2026.OC-984

Corresponding Author:
Gonçalo Martins e Pereira
Physical and Rehabilitation Medicine, ULSAS, Almada, Portugal
Email: gmartinsepereira@gmail.com