

Zinoubia Hasasna

**Philadelphia College of Osteopathic Medicine
USA**

The Role of AI in Foot and Ankle Patient Education: A Comparative Analysis of ChatGPT and AOFAS Resources

Abstract:

ChatGPT, an AI-powered chatbot, utilizes natural language generation to provide human-like responses across various topics, including medicine. It continuously learns and improves through advanced algorithms. ChatGPT offers up-to-date information on diseases, symptoms, and treatments in healthcare. Its latest iterations, GPT-3.5 and GPT-4.0, enhance accuracy and performance. This study evaluates ChatGPT's patient education capabilities by comparing GPT-3.5 and GPT-4.0 in explaining fifteen foot and ankle orthopedic conditions. Additionally, Foot and Ankle Orthopedic Surgeons assess the accuracy of the provided information.

Fifteen foot and ankle orthopedic conditions were presented to ChatGPT (GPT-3.5 and GPT-4.0) using the standardized prompt: "I have been diagnosed with [condition]. Can you tell me more about it?" The generated responses were compared to publicly available information from the American Orthopaedic Foot & Ankle Society (AOFAS) Foot Care MD patient information sheets within the same time frame. Each inquiry was conducted in a separate chat session to prevent ChatGPT from referencing prior responses. To evaluate the accuracy of ChatGPT's information, two fellowship-trained foot and ankle orthopedic surgeons (A.B. and G.I.P.) independently reviewed the outputs. They categorized their accuracy into four tiers: <50%, 50–74%, 75–99%, or 100% accurate, with each category reflecting an estimated level of correctness.

Compared to the AOFAS FootCare MD website, ChatGPT-4.0 provided a more comprehensive description of symptoms, whereas ChatGPT-3.5 included more risk factors. In contrast, AOFAS offered more detailed treatment options. Regarding accuracy, the majority of conditions evaluated using ChatGPT-3.5 (12 of 15) and ChatGPT-4.0 (13 of 15) were rated as primarily accurate (75%–99%) by both reviewers. Notably, both surgeons consistently classified one condition generated by ChatGPT-3.5 as mostly inaccurate (<50% accuracy). Interobserver agreement for accuracy ratings was poor, with a Cohen's kappa coefficient of -0.02, indicating a lack of consistency between reviewers.

ChatGPT (GPT-3.5 and GPT-4.0) demonstrates a relatively high degree of accuracy in generating patient education materials for foot and ankle orthopedic conditions. However, its ability to provide detailed, condition-specific information remains limited. Specialty organizations, such as the American Orthopaedic Foot & Ankle Society (AOFAS), are the most accurate and reliable sources for comprehensive musculoskeletal information related to foot and ankle pathology.

Biography

Zinoubia Hasasna is affiliated with the Philadelphia College of Osteopathic Medicine, USA, with expertise in medical research and education, contributing to advancements in osteopathic medicine and healthcare practice.