Medical Student Research Symposium   School of Medicine

March 2024

Are ChatGPT Responses to Urologic Health Inquiries Readable and Supported by American Urological Association Guidelines?

Chase Morrison
Wayne State University, hi2717@wayne.edu

Jack Vercnocke MD
jack.vercnocke@med.wayne.edu

Isaiah Dennis
dennisisaiah65@gmail.com

Michael Cher MD
mcher@med.wayne.edu

Steven Lucas MD
smlucas@med.wayne.edu

See next page for additional authors

Recommended Citation

Authors

Chase Morrison, Jack Vercnocke MD, Isaiah Dennis, Michael Cher MD, Steven Lucas MD, John Cochrane MD, Aron Liaw MD, and Kevin Ginsburg MD

Are ChatGPT Responses to Urologic Health Inquiries Readable and Supported by American Urological Association Guidelines?

Chase Morrison, Jack Vercnocke, Isaiah Dennis, Michael Cher, Steven Lucas, John Cochrane, Aron Liaw, Kevin Ginsburg

Introduction and Objective:
ChatGPT is widely accessible to patients, but it is unknown if the responses provided by the chatbot are supported by corresponding American Urological Association's (AUA) guidelines. Our primary objective was to assess if cGPT's responses to a hypothetical patient's question regarding their urologic conditions were supported by the AUA guidelines.

Methods:
Thirty prompts, each addressing one or more AUA Guideline, were written from the perspective of a patient. Responses were assessed by 8 evaluators consisting of practicing Urologists and current Urology Residents. Each response was evaluated using a Support Score (SS) and Quality Score (QS) (Table 1). A median SS score of ≥3 was considered "Supported" by the AUA Guidelines and a median QS score of ≥4 was considered "High Quality". Each response was analyzed for readability using Flesh-Kincaid Readability Grade Level (FKRGL) scores. The kappa coefficient was calculated to determine interrater agreement.

Results:
Overall, 20/30 (66%) responses were supported by the AUA guidelines. Responses to oncology questions were unlikely to be supported by AUA guidelines (5/12 questions supported) while questions pertaining to infertility (4/4), erectile issues (2/2), and reconstruction (1/1) were uniformly supported by the guidelines. Only 11/30 (37%) responses were deemed high quality. Questions in most domains of urology generally did not have high quality responses, while ¾ infertility questions were deemed high quality. The average FKR Grade Level Score was 18 which is equivalent to a college graduate reading level. Response grade reading levels ranged from 13 (college level) to 23 (professional). There was "slight" interrater agreement among SS (k=0.15, p=<0.001) and QS (k=0.06, p=0.004).

Conclusion
While cGPT can provide responses supported by the AUA guidelines, the majority missed minor or nuanced elements which may influence patient care. Most responses lacked quality and accessibility in terms of reading level. Responses from cGPT about urologic conditions are not a suitable alternative to consulting with medical professionals.