REV INT ANDROL
Revista Internacional de Andrología

ORIGINAL RESEARCH

Can artificial intelligence chatbots provide reliable patient education for penile curvature treatment?

Murat Ozturk[1],*, Akif Koc[1], Metin Kilic[2], Abdullah Gul[1], Salim Zengin[2], Ali Rıza Turkoglu[1], Soner Coban[1], Atilla Satir[2]

[1]Department of Urology, Bursa Faculty of Medicine, Health Sciences University, 16310 Bursa, Turkey
[2]Department of Urology, Bursa Training and Research Hospital, 16310 Bursa, Turkey

*Correspondence
murat.ozturk@sbu.edu.tr
(Murat Ozturk)

Abstract

Background: This study evaluated the reliability and quality of information provided by four artificial intelligence (AI) chatbots regarding penile curvature treatment. As patients increasingly seek medical information online, particularly for sensitive urological conditions such as penile curvature, AI chatbots have demonstrated potential for addressing these inquiries. However, the accuracy, comprehensiveness, and readability of their responses remain uncertain. The primary objective of this study was to evaluate the reliability and quality of information provided by AI chatbots on penile curvature treatments. **Methods**: The question "How is penile curvature treated?" was asked to four different AI chatbots: ChatGPT, Perplexity, Gemini and Copilot. Responses were independently evaluated by five urologists based on readability, understandability, actionability, reliability and transparency. The DISCERN score, PEMAT-P (Patient Education Materials Assessment Tool for Print Materials) test, WRR (Web Resource Rating) scale, Coleman-Liau index and Likert scale were used for assessment. **Results**: The DISCERN score evaluation showed that Gemini provided poor-quality information, ChatGPT and Copilot offered moderate-quality information, and Perplexity provided good-quality information (Total DISCERN scores: 31, 41, 42 and 51, respectively). PEMAT-P Understandability scores were 45% for Gemini, 55% for Copilot, 64% for ChatGPT and 73% for Perplexity. PEMAT-P Actionability scores were 40% for ChatGPT, Gemini and Copilot, and 60% for Perplexity. According to the Coleman-Liau index, readability levels were required at least at the university-level education. **Conclusions**: AI chatbots can be useful tools for obtaining information on penile curvature. Given their growing utilization, educating patients on effective interactions with AI chatbots to enhance response accuracy is crucial. With continuous updates and professional oversight, these tools are expected to evolve to become more effective in the future.

Keywords

Artificial intelligence; Chatbot; Quality assessment; Treatment; Penile curvature

¿Pueden los chatbots de inteligencia artificial proporcionar una educación confiable al paciente para el tratamiento de la curvatura del pene?

Resumen

Antecedentes: Este estudio evaluó la fiabilidad y la calidad de la información proporcionada por cuatro chatbots de inteligencia artificial (IA) en relación con el tratamiento de la curvatura del pene. Dado que cada vez más pacientes buscan información médica en línea, en particular para afecciones urológicas delicadas como la curvatura del pene, los chatbots de IA han demostrado potencial para abordar estas consultas. Sin embargo, la precisión, exhaustividad y legibilidad de sus respuestas siguen siendo inciertas. El objetivo principal de este estudio fue evaluar la fiabilidad y la calidad de la información proporcionada por los chatbots de IA sobre los tratamientos de la curvatura del pene. **Métodos**: La pregunta "¿Cómo se trata la curvatura del pene?" se formuló a cuatro chatbots de IA diferentes: ChatGPT, Perplexity, Gemini y Copilot. Cinco urólogos evaluaron las respuestas de forma independiente, basándose en su legibilidad, comprensión, viabilidad, fiabilidad y transparencia. Para la evaluación se utilizaron la puntuación DISCERN, la prueba PEMAT-P (Patient Education Materials Assessment Tool for Print Materials), la escala WRR (Web Resource Rating), el índice Coleman-Liau y la escala Likert. **Resultados**: La evaluación de la puntuación DISCERN mostró que Gemini proporcionaba información de baja calidad, ChatGPT y Copilot ofrecían información de calidad moderada, y Perplexity proporcionaba información de buena calidad (puntuaciones DISCERN totales: 31, 41, 42 y 51, respectivamente). Las puntuaciones de comprensión de PEMAT-P fueron del 45 % para Gemini, del 55 % para Copilot, del 64 % para ChatGPT y del 73 % para Perplexity. Las puntuaciones de procesabilidad de PEMAT-P fueron del 40 % para ChatGPT, Gemini y Copilot, y del 60 % para Perplexity. Según el índice Coleman-Liau, los niveles de legibilidad exigidos eran al menos en el nivel de educación universitario. **Conclusiones**: Los chatbots de IA pueden ser herramientas útiles para obtener información sobre la curvatura del pene. Dado su creciente uso, es fundamental educar a los pacientes sobre interacciones efectivas con chatbots de IA para mejorar la precisión de las respuestas. Con actualizaciones continuas y supervisión profesional, se espera que estas herramientas evolucionen para ser más eficaces en el futuro.

Palabras Clave

Inteligencia artificial; Chatbot; Evaluación de la calidad; Tratamiento; Curvatura del pene

1. Introduction

Penile curvature (PC) may result from congenital structural anomalies, such as tunical dysplasia, or be acquired later in life due to fibrotic changes, trauma, or other pathological alterations of the penile architecture [1–3]. Congenital PC is a relatively uncommon condition, affecting less than 1% of the population [4]. Acquired PC, however, is most commonly associated with Peyronie's disease (PD). The overall prevalence of PD ranges from 0.5% to 13% [5, 6]. PD often leads to penile curvature, shortening, pain, difficulties in maintaining erections, and challenges during sexual intercourse [7]. Both medical and surgical approaches are employed in the treatment of PC. Surgical intervention is typically recommended in cases of severe deformity (curvature $>30°$) and/or erectile dysfunction [5].

The internet has become a readily accessible and frequently used resource for health-related information. For patients hesitant to consult a physician due to feelings of embarrassment, online platforms have become a significant point of reference [8]. Telemedicine services, by reducing the need for hospital visits, can save time and reduce treatment costs for both patients and healthcare providers. Additionally, their rapid accessibility can help alleviate the workload of hospitals and clinics [9]. Consequently, the reliability and quality of online information are of paramount importance. Unfortunately, some websites provide inaccurate or incomplete information, and research demonstrates a prevalence of suboptimal health information online [10, 11].

Artificial intelligence chatbots (AICs) offer the potential to address patient queries by understanding symptoms, providing personalized medical advice, and offering insights into potential treatment options. However, concerns remain regarding the accuracy, reliability, and applicability of the information provided by AICs [12]. While these chatbots exhibit remarkable capabilities have shown remarkable capabilities across a variety of disciplines, various disciplines, they also have inherent limitations and have been shown to generate incorrect information at times [13]. Moreover, there is a paucity of studies evaluating the readability, comprehensibility, and actionable quality of AIC-generated responses specific to PC.

Therefore, a comprehensive assessment of the reliability of information provided by AICs regarding PC treatments is critical before relying on them to inform treatment decisions. This study aims to evaluate the reliability and quality of information provided by AICs regarding PC treatments and to examine the potential advantages and disadvantages of integrating these tools into urological care.

2. Materials and methods

The question "How is penile curvature treated?" was posed to four artificial intelligence chatbots (ChatGPT, Perplexity, Gemini, and Copilot). The study design is summarized in Fig. 1. The chatbot responses to the prompts asked on five separate computers were recorded and independently analyzed by five urologists.
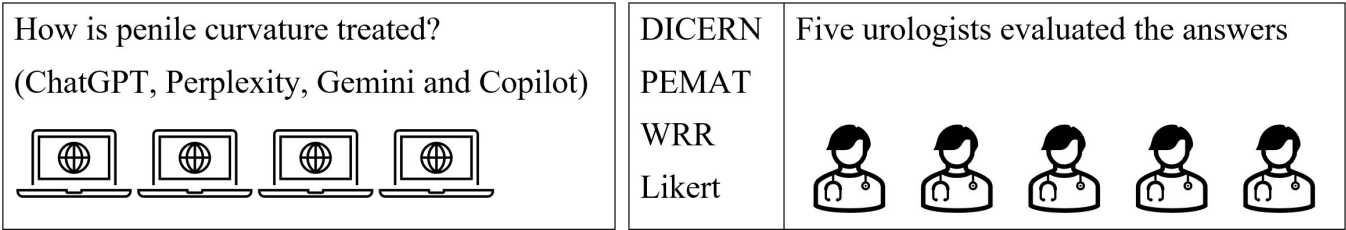
The urologists calculated the DISCERN score, a validated

F I G U R E 1. Study design. PEMAT: Patient Education Materials Assessment Tool; WRR: Web Resource Rating.

instrument used to assess the reliability and quality of medical information. DISCERN criteria are designed to evaluate the quality of patient education materials, such as brochures about PC. The tool comprises 16 questions, each rated on a scale from 1 to 5. Questions 1–8 assess the reliability and credibility of the source, while questions 9–15 evaluate the specificity and details of the treatment options presented. The final question provides an overall quality rating. Total scores are categorized as "excellent" (63–75), "good" (51–62), "moderate" (39–50), "poor" (27–38) and "very poor" (15–26) [14].

The understandability and actionability of the chatbot responses were evaluated using the Patient Education Materials Assessment Tool for Print Materials (PEMAT-P), with scores ranging from 0% to 100% [15]. The Web Resource Rating (WRR) scale, which measures the reliability and transparency of online information, was also used, with scores between 0% and 100% [16]. Text readability was assessed using the Coleman-Liau Index, where scores above 11 indicate the need for at least college-level education to comprehend the material [17].

Guideline compatibility of the treatment options was evaluated using a 5-point Likert scale. A score of 1 indicated "serious or extensive deficiencies", while a score of 5 reflected "minimal deficiencies". In alignment with prior literature, scores below 3 were classified as poor, scores between 3 and 4 as moderate, and scores above 4 as highly reliable [18].

Inter-rater reliability for the 16-question DISCERN assessment was assessed using the the Likert scale, intraclass correlation coefficient (ICC) analysis was performed, yielding an ICC of 0.801 (95% confidence interval 0.708–0.871; $p < 0.0001$), indicating strong agreement among the urologists. Other scales did not undergo ICC testing, as their scoring was based on objective data (Agree/Disagree).

Data analyses were conducted using IBM SPSS Statistics (version 25, IBM Corp., Armonk, NY, USA). Scoring systems and the consistency of treatment recommendations were reported as median values (minimum–maximum), while the word count and Coleman-Liau Index scores were presented as mean values.

3. Results

The question "How is penile curvature treated?" was posed to four different artificial intelligence chatbots (AICs): ChatGPT, Perplexity, Gemini and Copilot. Their responses were evaluated by five urologists using the latest version of the European Association of Urology (EAU) Patient Information guidelines [16]. The study's findings, encompassing DISCERN score, Coleman-Liau Index results, PEMAT-P scores, WRR score, word counts, and the Likert scale assessments regarding treatment guideline compatibility, are summarized in Table 1.

Evaluation of the treatment information quality, as assessed by DISCERN scores, revealed that Gemini provided information of poor quality, while ChatGPT and Copilot generated information of moderate quality. Perplexity scored higher in certain areas due to the quality of its references, clear presentation of treatment options, and the inclusion of supportive images and videos. In contrast, while ChatGPT and Copilot presented treatment options, their lack of explanatory images and tables resulted in lower scores. Gemini, which had the lowest DISCERN score among the four AICs, provided limited and superficial information with brief headings. PEMAT-P Understandability scores were 45% for Gemini, 55% for Copilot, 64% for ChatGPT and 73% for Perplexity. PEMAT-P Actionability scores were 40% for ChatGPT, Gemini and Copilot, and 60% for Perplexity. The reliability of the re-

T A B L E 1. Evaluation of quality, readability, understandability and reliability of chatbots.

	ChatGPT	Perplexity	Gemini	Copilot
Total DISCERN score, median (min–max)	41 (32–44)	51 (47–59)	31 (23–41)	42 (29–52)
PEMAT-P Understandability	64 (50–67)	73 (50–87)	45 (31–60)	55 (40–79)
PEMAT-P Actionability	40 (40–80)	60 (0–100)	40 (20–80)	40 (20–80)
WRR	7.14	46.40	30.35	33.20
Word Count	542	332	108	254
Coleman Liau Index	15.81	17.27	15.28	16.45
Likert score	4 (3–4)	4 (3–5)	3 (1–3)	4 (2–4)

Abbreviations: PEMAT-P: the patient education materials assessment tool-printable materials; WRR: web resource rating; min: Minimum; max: Maximum.

sponses was assessed using the Web Resource Rating (WRR) scale, which indicated that all responses had low reliability (7.14%–46.4%). Among the AICs, Perplexity had the highest WRR median score, while ChatGPT had the lowest.

According to the Coleman-Liau Index, the readability of the responses was deemed highly challenging, requiring a professional-level understanding. All AICs scored above 11 on the index, indicating that the information required at least a university-level education to comprehend.

Assessment of the treatment options' compatibility with clinical guidelines, as measured by the Likert scale, demonstrated moderate adherence across all AICs, with scores ranging from 3 to 4. Perplexity was found to be the most guideline-compliant chatbot.

4. Discussion

This study aimed to evaluate and compare the reliability and effectiveness of four different artificial intelligence chatbots (AICs) in presenting treatment options for penile curvature (PC), with a particular focus on patient guidance and education. It's crucial to reiterate that AICs are not intended to supplant clinical expertise; the treatment plans generated by these tools are not direct equivalents of established urological guidelines. The primary focus of this research was to assess the information provided to patients rather than treatment planning itself. The findings revealed significant variations in performance across multiple criteria for each AIC. In this discussion, a critical analysis is presented by synthesizing the study results with relevant literature.

For the past two decades, a substantial amount of medical information has been disseminated through the internet, particularly via platforms such as YouTube [19]. While content from healthcare professionals can be beneficial, information from individuals lacking medical expertise carries the risk of misinformation [19, 20]. In the digital age, patients increasingly self-diagnose and explore treatment options online. The widespread availability of ChatGPT v3.5 since 2022 has further facilitated access to health-related information, enabling individuals to obtain answers to complex medical inquiries rapidly. This accessibility has contributed to growing reliance on AICs for healthcare-related decision-making. A study investigating patient experiences with AICs reported that 84.1% of users found them useful, and 91.4% indicated they would use them again [21].

Christopher J. Warren et al. [22] demonstrated that AICs have the potential to offer provide high-quality information regarding PD; however, the quality of the generated content is contingent on the user's input. Furthermore, the citations provided by AICs were found to be unreliable, with only 42% being accurate, and the readability of responses exceeded the comprehension level of the average American [22]. In the present study, an evaluation of the quality of treatment-related information indicated that among the four AICs assessed, Gemini provided poor-quality information, ChatGPT and Copilot delivered moderate-quality content, and Perplexity offered high-quality information.

According to the DISCERN scores, Perplexity demonstrated the most robust performance of treatment information quality.

This superior performance can be attributed to its accurate references, comprehensive explanations, and the inclusion of supporting visuals like images and videos. DISCERN scores are widely recognized in the literature as an objective and reliable measure for evaluating the quality of patient education materials [23]. While ChatGPT and Copilot presented treatment options, the lack of explanatory images and tables may have contributed to their lower scores. Conversely, Gemini, providing brief and superficial content, received the lowest DISCERN score.

A previous study by Erkan et al. [10] assessed the reliability and quality of urogenital cancer treatment information provided by ChatGPT, Gemini and Copilot. Their findings showed that ChatGPT and Gemini offered moderate-quality information, whereas Copilot provided low-quality content [10]. In the current study on PC treatment, Gemini's information quality was assessed as poor, while ChatGPT and Copilot performed at a moderate level, and Perplexity demonstrated superior quality. These findings suggest that the effectiveness of AICs may vary depending on the medical domain, underscoring the need for further research to identify the most suitable AICs for different health conditions. Moreover, consistent with previous research, our study revealed that the readability of the AIC responses was challenging, requiring a professional-level understanding.

PEMAT-P scores revealed notable differences in the understandability and actionability of the information provided by the AICs. Perplexity outperformed other AICs in both categories, suggesting its capacity to empower patients to comprehend and apply complex treatment information. However, the moderate performance of ChatGPT and Copilot highlights the limitations of these AICs in effectively conveying medical information. Literature suggests that the comprehensibility of patient education materials plays a crucial role in treatment adherence [15, 24]. In Erkan et al.'s [10] study, PEMAT-P understandability scores were low (40%), and actionability scores were moderate only for Gemini (60%), while the other AICs scored lower (40%). In our study, PEMAT-P Understandability scores were 45% for Gemini, 55% for Copilot, 64% for ChatGPT, and 73% for Perplexity. PEMAT-P Actionability scores were 40% for ChatGPT, Gemini, and Copilot, and 60% for Perplexity. In a study conducted on urology patients, Cocci et al. [25] found that ChatGPT produced low-scoring information.

The WRR scores indicated that the reliability of AIC responses was generally low, highlighting the limitations of AICs in accessing and accurately interpreting evidence-based information. Among the assessed AICs, Perplexity demonstrated relatively better reliability scores, suggesting higher accuracy and source credibility compared to others. Across both studies, the WRR scores remained low, ranging from 7.14% to 46.4%, underscoring the need for continuous monitoring and improvement of AIC algorithms.

According to the Coleman-Liau Index, the readability of the responses was found to be significantly high, indicating that they require at least a university-level education for comprehension. This finding aligns with the broader literature emphasizing the need for health education materials to be accessible to the general public [26–28].

When evaluating the consistency of treatment recommendations with EAU Patient Information guidelines, the AICs demonstrated moderate adherence (Likert 3–4), with Perplexity achieving the highest compatibility. This finding aligns with previous research emphasizing the importance of evidence-based practices in guiding patient care [17, 29]. On the other hand, most AICs were considered to provide moderately consistent information when evaluated according to the objective booklet. The discrepancy noted was due to incomplete content rather than incorrect information. The rate of incorrect information was not specifically measured in our study. This can be considered a limitation. Measuring this rate in future studies may add value to the research. Perplexity showed the most consistent and error-free information among the AICs we evaluated. The rate of incorrect information was not specifically measured in our study.

The integration of AI into healthcare is transforming patient care, with AI-powered chatbots emerging as valuable tools in telemedicine. In sexual medicine, where privacy and patient comfort are critical, these technologies have the potential to complement telehealth consultations. AI chatbots can improve telemedicine services by streamlining pre-consultation processes, gathering patient histories, assessing symptoms and assisting in triage. Post-consultation, they can reinforce treatment adherence, provide psychoeducation, and address common patient inquiries. The anonymity offered by AI-driven systems may encourage disclosure of sensitive issues, promoting a more comprehensive clinical assessment. However, their limitations, including potential misinformation, lack of human empathy, and ethical concerns regarding patient safety and data security, must be carefully evaluated.

The effectiveness of telemedicine in sexual health care depends on patient perceptions and accessibility. Factors such as privacy, convenience, and comfort in discussing sensitive topics remotely need to be analyzed. Barriers to adoption, including technological limitations, trust issues, and cultural considerations, must be explored. Additionally, telemedicine's role in expanding access to sexual healthcare for underserved populations, such as those in rural areas or with disabilities, should be investigated to determine its broader impact.

A secure photo-based AI assessment tool could integrate into Electronic Health Records (EHRs) to provide personalized recommendations but could still create significant privacy concerns. In terms of privacy concerns, a connection to the EHR could be established through a secure app that will be developed in the future.

A mixed-methods research approach, incorporating both quantitative surveys and qualitative interviews with patients and healthcare providers, can provide a comprehensive understanding of these issues. The findings can inform future digital health strategies, optimize telemedicine services in sexual medicine, and support the ethical and effective integration of AI chatbots into patient care.

This study adds to the growing body of research evaluating the ability of AICs in providinformation regarding penile curvature management. Among the examined AICs, Perplexity demonstrated superior performance across multiple criteria. However, the underscore the need for substantial improvements in the patient-centered presentation of information across all AICs. AICs must deliver accurate information in a manner that is readily understandable to a broader audience.

5. Conclusions

It is essential to recognize that current AICs should serve as supportive tools rather than complete substitutes for healthcare professionals. While they offer valuable information, they lack the empathy, clinical insight, and experience of human healthcare providers. The ultimate goal should be to strategically integrate AICs in collaboration with healthcare professionals to enhance patient education and optimize clinical outcomes.

To maximize the potential of AICs, reliable data sources and robust algorithms must be integrated, rigorously tested, and regularly updated to maintain performance and accuracy. Furthermore, AIC-generated information should be presented in a manner that is easily comprehensible to a broader audience. For their efficient utilization, both healthcare professionals and patients should receive adequate training and support. In addition to their role in providing information, the development of AI-supported applications holds promise for use in disease diagnosis, such as for penile curvature, especially where visual assessment is important. However, utmost care must be taken to rigorously protect patient privacy throughout all stages of development and implementation. Future studies should focus on evaluating AIC performance across various medical conditions to continuously their capabilities and ensure their safe and effective implementation in clinical practice.

AUTHOR CONTRIBUTIONS

MO and AK—designed the research study; wrote the manuscript. MO, AK, MK, SZ and AG—performed the research. SC, ART and AS—provided help and advice on the research; analyzed the data. MO—wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Nyirády P, Kelemen Z, Bánfi G, Rusz A, Majoros A, Romics I. Management of congenital penile curvature. The Journal of Urology. 2008; 179: 1495–1498.

[2] Hsieh JT, Liu SP, Chen Y, Chang HC, Yu HJ, Chen CH. Correction of congenital penile curvature using modified tunical plication with absorbable sutures: the long-term outcome and patient satisfaction. European Urology. 2007; 52: 261–267.

[3] Appiah Boakye S, Obeng F. Clinical outcomes of circumcisions and prevalence of complications of male circumcisions: a five-year retrospective analysis at a teaching hospital in Ghana. PLOS Global Public Health. 2025; 5: e0003788.

[4] Vorobev V, Beloborodov V, Sherbatykh A, Kalyagin A, Baklanova O, Popov S, et al. The effectiveness of penile curvature treatment by cavernous body rotation and plication of the tunica albuginea. Basic and Clinical Andrology. 2023; 33: 10.

[5] Sasso F, Vittori M, D'Addessi A, Bassi PF. Penile curvature: an update for management from 20 years experience in a high volume centre. Urologia Journal. 2016; 83: 130–138.

[6] Paulis G, De Giorgio G, Paulis A. Clinical presentation of peyronie's disease: a retrospective study of 564 cases. Diagnostics. 2024; 14: 1125.

[7] Howell S, Green T, Khalafalla K. Current insights and future directions in Peyronie's disease management: a narrative review. UroPrecision. 2024; 3: 3–16.

[8] Gul M, Diri MA. YouTube as a source of information about premature ejaculation treatment. The Journal of Sexual Medicine. 2019; 16: 1734–1740.

[9] Naik N, Hameed BMZ, Nayak SG, Gera A, Nandyal SR, Shetty DK, et al. Telemedicine and telehealth in urology—what do the 'patients' think about it? Frontiers in Surgery. 2022; 9: 863576.

[10] Erkan A, Koc A, Barali D, Satir A, Zengin S, Kilic M, et al. Can patients with urogenital cancer rely on artificial intelligence chatbots for treatment decisions? Clinical Genitourinary Cancer. 2024; 22: 102206.

[11] Plambeck B, Jiang J, Perez J, Wordekemper BE, Fu D, Colvin A, et al. Online health information for penile prosthesis implants lacks quality and is unreadable to the average US patient. Cureus. 2023; 15: e34240

[12] Talyshinskii A, Naik N, Hameed BMZ, Juliebø-Jones P, Somani BK. Potential of AI-driven chatbots in urology: revolutionizing patient care through artificial intelligence. Current Urology Reports. 2024; 25: 9–18.

[13] Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Network Open. 2023; 6: e2336483.

[14] Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. Journal of Epidemiology and Community Health. 1999; 53: 105–111.

[15] Agency for Healthcare Research and Quality. The patient education materials assessment tool (PEMAT) and user's guide. Available at: https://www.ahrq.gov/es/health-literacy/patient-education/pemat1.html (Accessed: 11 February 2025).

[16] Dobbins M, Watson S, Read K, Graham K, Nooraie RY, Levinson AJ. A tool that assesses the evidence, transparency, and usability of online health information: development and reliability assessment. JMIR Aging. 2018; 1: e3.

[17] EAU Patient Information. Penile curvature. Available at: https://patients.uroweb.org/condition/penile-curvature/ (Accessed: 11 February 2025).

[18] Warren CJ, Sawhney R, Shah T, Behbahani S, Sadeghi-Nejad H. YouTube and men's health: a review of the current literature. Sexual Medicine Reviews. 2021; 9: 280–288.

[19] Carneiro B, Dizon DS. Prostate cancer social media: in YouTube we trust? European Urology. 2019; 75: 568–569.

[20] Selvi I, Baydilli N. An analysis of misleading YouTube videos on urological conditions: what to do about the danger of spreading misinformation of the YouTube videos? World Journal of Urology. 2022; 40: 885–886.

[21] Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. Journal of Medical Internet Research. 2020; 22: e14679.

[22] Warren CJ, Edmonds VS, Payne NG, Voletti S, Wu SY, Colquitt J, et al. Prompt matters: evaluation of large language model chatbot responses related to Peyronie's disease. Sexual Medicine. 2024; 12: qfae055.

[23] Charnock D. The DISCERN handbook: quality criteria for consumer health information on treatment choices. 1st edn. Radcliffe Medical Press: Abingdon. 1998.

[24] Shoemaker SJ, Wolf MS, Brach C. Development of the patient education materials assessment tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Education and Counseling. 2014; 96: 395–403.

[25] Cocci A, Pezzoli M, Lo Re M, Russo GI, Asmundo MG, Fode M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. Prostate Cancer and Prostatic Diseases. 2024; 27: 103–108.

[26] Liu Y, Ji M, Lin SS, Zhao M, Lyv Z. Combining readability formulas and machine learning for reader-oriented evaluation of online health resources. IEEE Access. 2021; 9: 67610–67619.

[27] Coskun S, Bagcivan G. Associated factors with treatment adherence of patients diagnosed with chronic disease: Relationship with health literacy. Applied Nursing Research. 2021; 57: 151368.

[28] Severance SJ, Cohen KB. Measuring the readability of medical research journal abstracts. BioNLP 15. Beijing, China and July 2015. Association for Computational Linguistics: Beijing. 2015.

[29] MacLennan S, Duncan E, Skolarus TA, Roobol MJ, Kasivisvanathan V, Gallagher K, et al. Improving guideline adherence in urology. European Urology Focus. 2022; 8: 1545–1552.