



ORIGINAL RESEARCH

Assessing the Ability of Artificial Intelligence-Driven Language Processing Frameworks to Create Patient-Oriented Medical Education Material on Hypothermia

Adam Schwartz¹ and Alfred Urba²

¹Rosalind Franklin University of Medicine and Science (US) and ²Illinois Farm Bureau (US)

Abstract

Introduction: Artificial Intelligence-Driven Language Processing Frameworks (AI-LPFs) such as ChatGPT, Grok, and Gemini are increasingly being explored for their ability to generate patient-oriented medical education material (PEM). While prior studies have assessed AI-generated PEM in various medical fields, their applicability to operational medicine remains understudied. Given the significance of hypothermia in operational and civilian settings, this study evaluates the quality and readability of AI-generated PEM on hypothermia. **Methods:** Three AI-LPFs (ChatGPT-4, Grok-3, and Gemini 2.0 Flash) were prompted to generate PEM on hypothermia. Readability was assessed using the Flesch-Kincaid reading grade level and Flesch Reading Ease Score (FRE). Additional text metrics included PEM length, the proportion of complex words and sentences, and average sentence and word length. The quality of AI-generated PEM was scored using the CDC Clear Communication Index (CCI), and content accuracy was assessed through fact-checking against the Wilderness Medical Society guidelines. A benchmark PEM from the American Red Cross was included for comparison. **Results:** Readability analysis showed that the PEM from Gemini and the American Red Cross met NIH recommendations for an 8th-grade reading level, whereas ChatGPT and Grok were slightly above this threshold. Grok generated the most comprehensive PEM, uniquely categorizing hypothermia into mild, moderate, and severe, aligning with Wilderness Medical Society guidelines. Unlike the other AI-generated PEM, it also addressed both EMS activation and CPR. The PEM from Grok scored the highest on the CDC CCI, outperforming the other AI-generated PEMs and the benchmark from the American Red Cross. A manual review confirmed that all AI-generated PEM were factually accurate. **Conclusion:** AI-LPFs successfully produced factually accurate PEM on hypothermia, with Grok generating the most comprehensive material. These findings suggest AI-LPFs have potential for enhancing public education on operational medicine topics. Further refinement of AI-generated PEM to improve readability and adherence to established guidelines may enhance their utility as reliable educational tools.

Key words: artificial intelligence, hypothermia, patient education

Introduction

Over the past several years, there has been an explosion of interest in the applications of Artificial Intelligence-Driven Language Processing Frameworks (AI-LPF), such as ChatGPT, to patient-oriented medical education¹. Recent studies have shown promise in the ability of AI-LPFs to create quality patient education materials (PEM) on topics ranging from dermatological diseases to men's health.^{1,2} Despite the recent interest in this topic, there has been a lack of assessments on the applications of AI-LPFs on topics within the realm of operational medicine. As both operational personnel and the general public often face extreme weather conditions, this study aims to assess the quality of PEMs on the topic of hypothermia, which

were created by several popular AI-LPFs (ChatGPT-4, Grok-3, and Gemini 2.0 Flash).³⁻⁵

Methods

Three popular AI-LPFs (Open AI's ChatGPT-4, xAI's Grok-3, and Google's Gemini 2.0 Flash) were given the prompt: "Create a patient oriented educational material on the topic of hypothermia which addresses symptoms, causes, and treatment of hypothermia."

Content analysis was then performed by manual review. This included fact-checking the PEMs and comparing them to the Wilderness Medical Society guidelines for accidental hypothermia.⁶ A PEM on hypothermia from the American Red Cross, an established and authoritative source for

Table 1. Comprehensive summary of CDC CCI scores, readability scores, and raw text statistics

Source	CDC CCI	Total Words	Total Sentences	Average Syllables per Word	Average Words per Sentence	Sentences with >30 Syllables (%)	Words with >4 Syllables (%)	FK Reading Grade Level	FRE Score
American Red Cross	84	865	86	1.6	10.1	12 (14%)	15 (2%)	7.66	58.04
ChatGPT-4	81	294	38	1.8	7.7	1 (3%)	11 (4%)	8.90	45.03
Gemini 2.0 Flash	74	276	33	1.7	8.4	1 (3%)	13 (5%)	7.42	56.73
Grok-3	89	686	60	1.6	11.4	9 (15%)	21 (3%)	8.31	55.87
<i>CDC CCI: Center for Disease Control and Prevention Clear Communication Index</i>						<i>FK: Flesch-Kincaid</i>		<i>FRE: Flesch Reading Ease</i>	

high-quality public-oriented health information, was also analyzed as a comparative benchmark.^{7,8}

Readability analysis software was utilized to obtain the Flesch-Kincaid reading grade level and the Flesch Reading Ease Score (FRE). FRE scores range from 0-100, with higher scores representing easier-to-read material. Raw text statistics were also obtained. This included PEM length in sentences and words, the proportion of words over 4 syllables and sentences over 30 syllables, and average sentence and word length. The quality of these PEMs was then further analyzed using the CDC Clear Communication Index (CDC CCI). This tool scores educational material from 0-100, with higher scores indicating better written educational material.⁹

Results

Upon manual review, no false statements were identified in any of the AI-LPF-generated materials. All PEMs mention the general signs, symptoms, and treatments for hypothermia in a manner consistent with those outlined in the Wilderness Medicine Society guidelines. Another topic addressed in the Wilderness Medicine Society guidelines is deciding when to contact emergency medical services (EMS) and provide CPR. These topics were addressed in the American Red Cross and Grok PEMs. ChatGPT mentioned the need to contact EMS but not CPR, while the Gemini PEM did not address either. The Grok-generated PEM was the only material analyzed that stratified hypothermia into mild, moderate, and severe conditions, mirroring the approach taken in the Wilderness Medicine Society guidelines. A comprehensive summary of CDC CCI scores, readability scores, and raw text statistics can be found in Table 1.

Discussion

The PEM from Gemini and the American Red Cross met the NIH recommendation that PEM be written at an 8th-grade level or lower.¹⁰ Although above the recommended reading level from the NIH, the PEM from ChatGPT and Grok were less than a grade level above this cutoff. As 15% (9/60) of the sentences in the Grok PEM were over 30 syllables, breaking up larger sentences could be

an effective strategy to bring the reading grade level into compliance with NIH recommendations. Besides reading grade level, Grok consistently outperformed PEM from the other AI-LPF. The PEM from Grok was longer and more comprehensively written, being the only AI-LPF-generated PEM to address both the topics of recognizing the need to call EMS and performing CPR. Also, the PEM from Grok more closely mirrored Wilderness Medicine Society guidelines by categorizing hypothermia into mild, moderate, and severe forms and performed the best on the CDC CCI, even outscoring the PEM from the American Red Cross.

Conclusion

These research findings highlight the potential of AI-LPFs to be utilized as tools to inform the general population on topics in operational medicine. All AI-LPFs produced factually accurate materials that the majority of the general public could likely understand. However, the PEM from Grok appeared to more closely mirror resources from established sources, such as the Wilderness Medical Society and the American Red Cross.

Competing Interests

No competing interest is declared.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

All AI-generated content was evaluated based on publicly available guidelines and benchmarks, ensuring an objective and ethical assessment. The study adhered to principles of transparency, fairness, and responsible AI usage.

References

- Mondal H, Mondal S, Podder I. Using ChatGPT for Writing Articles for Patients' Education for

- Dermatological Diseases: A Pilot Study. *Indian Dermatol Online J.* 2023 Jun 28;14(4):482-486. doi: 10.4103/idoj.idoj_72_23. PMID: 37521213; PMCID: PMC10373821.
2. Shah YB, Ghosh A, Hochberg AR, Rapoport E, Lallas CD, Shah MS, Cohen SD. Comparison of ChatGPT and Traditional Patient Education Materials for Men's Health. *Urol Pract.* 2024 Jan;11(1):87-94. doi: 10.1097/UPJ.0000000000000490. Epub 2023 Nov 1. PMID: 37914380.
 3. Conlon KC, Rajkovich NB, White-Newsome JL, Larsen L, O'Neill MS. Preventing cold-related morbidity and mortality in a changing climate. *Maturitas.* 2011 Jul;69(3):197-202. doi: 10.1016/j.maturitas.2011.04.004. Epub 2011 May 17. PMID: 21592693; PMCID: PMC3119517.
 4. Irwin BR. A case report of hypothermia in the wilderness. *Wilderness Environ Med.* 2002 Summer;13(2):125-8. doi: 10.1580/1080-6032(2002)013[0125:acrohi]2.0.co;2.
 5. Herr RD, White GL Jr. Hypothermia: threat to military operations. *Mil Med.* 1991 Mar;156(3):140-4. PMID: 1901977.
 6. Hypothermia: Symptoms, causes, and treatment. *The American Red Cross.* (n.d.). <https://tinyurl.com/4dvdx5x>
 7. Dow J, Giesbrecht GG, Danzl DF, Brugger H, Sagalyn EB, Walpoth B, Auerbach PS, McIntosh SE, Nemethy M, McDevitt M, Schoene RB, Rodway GW, Hackett PH, Zafren K, Bennett BL, Grissom CK. Wilderness Medical Society Clinical Practice Guidelines for the Out-of-Hospital Evaluation and Treatment of Accidental Hypothermia: 2019 Update. *Wilderness Environ Med.* 2019 Dec;30(4S):S47-S69. doi: 10.1016/j.wem.2019.10.002. Epub 2019 Nov 15. PMID: 31740369.
 8. Harve MA, Li D. Social Impacts of the American Red Cross (ARC) Disaster Interventions: A Scoping Review. *Cureus.* 2024 Feb 29;16(2):e55265. doi: 10.7759/cureus.55265. PMID: 38558612; PMCID: PMC10981536.
 9. Baur C, Prue C. The CDC Clear Communication Index is a new evidence-based tool to prepare and review health information. *Health Promot Pract.* 2014 Sep;15(5):629-37. doi: 10.1177/1524839914538969. Epub 2014 Jun 20. PMID: 24951489.
 10. Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, Jagsi R, Golden DW. Readability of Patient Education Materials From High-Impact Medical Journals: A 20-Year Analysis. *J Patient Exp.* 2021 Mar 3;8:2374373521998847. doi: 10.1177/2374373521998847. PMID: 34179407; PMCID: PMC8205335.