

Use of generative artificial intelligence (AI) in medical communications for summarization tasks

Danielle L Rubin-Shepherd, Nitish Chaudhari, Frederick W Feldman, Amin Ghane, Stephanie Keys, Dana Lengel, Kristen Mourao, Angela Schlegel, Johnson Ying, Meryl Gersh

Red Nucleus, Yardley, PA, USA

Can generative AI platforms create effective plain language and therapeutic area summaries while adhering to medical communications standards for accuracy and accessibility?

Background

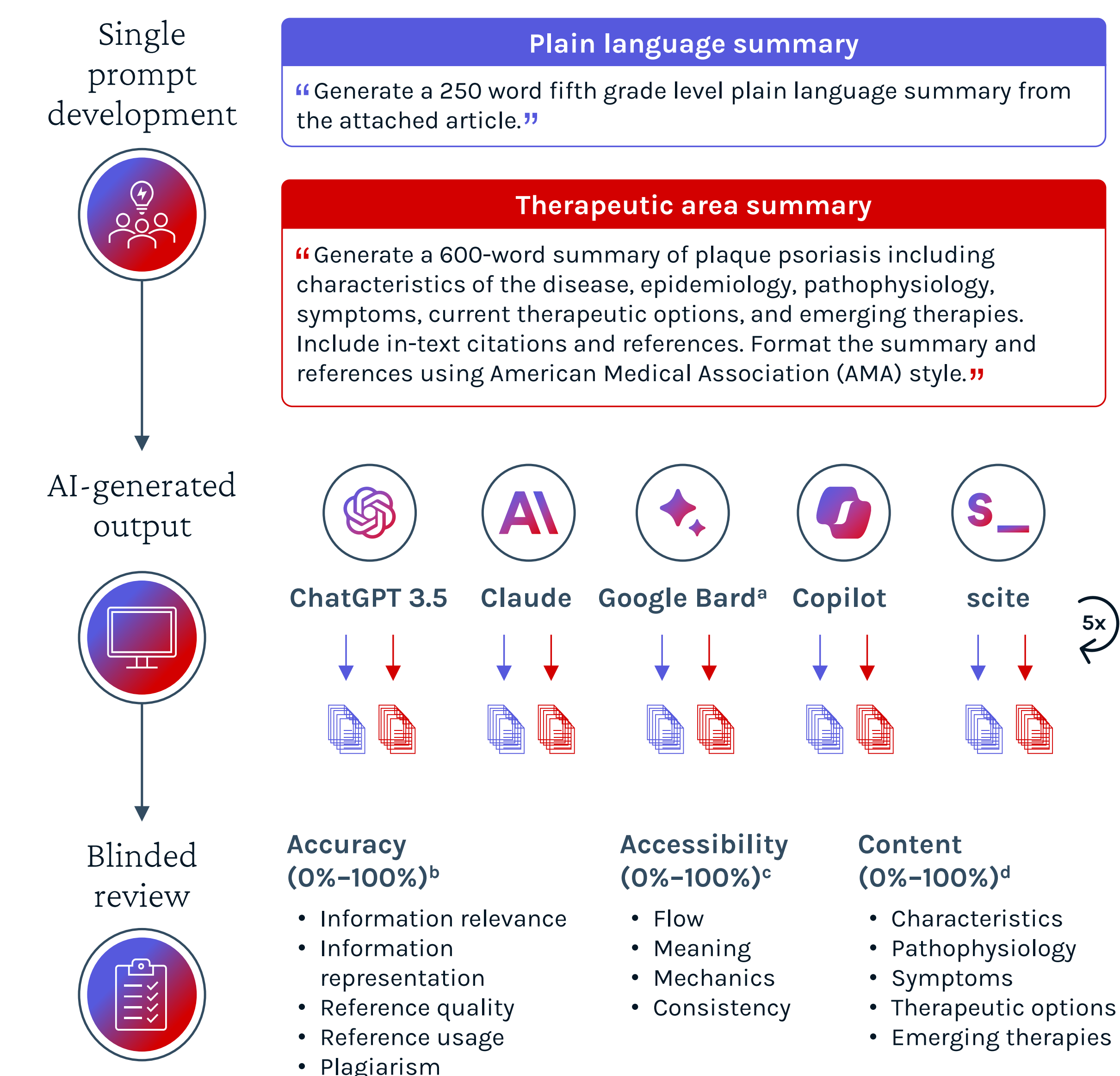
- Generative AI platforms may streamline multiple facets of medical communications, including literature searches, content generation, and summarization tasks¹⁻³
- The International Society for Medical Publication Professionals (ISMPP), the International Committee of Medical Journal Editors (ICMJE), and the Medical Affairs Professional Society (MAPS) recommend transparent use of generative AI⁴⁻⁶
 - ISMPP, ICMJE, and MAPS insist that human accountability and oversight are necessary to ensure accuracy
 - ISMPP and MAPS recommend using AI to enhance accessibility for patients and to bridge language barriers but provide warnings to correct for AI bias

Objective

- To evaluate the accuracy and accessibility of AI-generated plain language summaries and therapeutic area summaries

Research Design and Methods

- The free versions of 5 generative AI platforms (ChatGPT 3.5, Claude, Google Bard, Copilot, and scite) were prompted to generate plain language summaries of a clinical study⁷ and therapeutic area summaries about plaque psoriasis



^aRe-branded as Google Gemini in February 2024.

^bAccuracy was scored by a Senior Scientific Technical Editor.

^cAccessibility was scored by a Senior Medical Editor. References were excluded from evaluation.

^dContent thoroughness for each therapeutic area summary was scored by 2 Senior Medical Writers who had previous experience in plaque psoriasis.

For standardization between evaluators, 3 review articles⁸⁻¹⁰ were chosen a priori to compare content with each therapeutic area summary. AI, artificial intelligence.

Conclusions



ChatGPT 3.5

Best readability
Highest content score
References were fabricated



Claude

Good readability
Good content score
Good reference quality
Incorrect citation style



Google Bard

Good readability
Good content score
References were fabricated



Copilot

Worst readability
Worst content score
Low reference quality
Highest plagiarism



scite

Poor readability
Highest reference quality
High plagiarism

Plain language summaries

- Good accuracy across platforms
- Frequent errors in flow and meaning
- Occasional misinterpretation of key messages

Therapeutic area summaries

- Varied accuracy across platforms
- Fabricated and low-quality references and/or plagiarism from several platforms
- All platforms struggled with in-text citations and style guidance

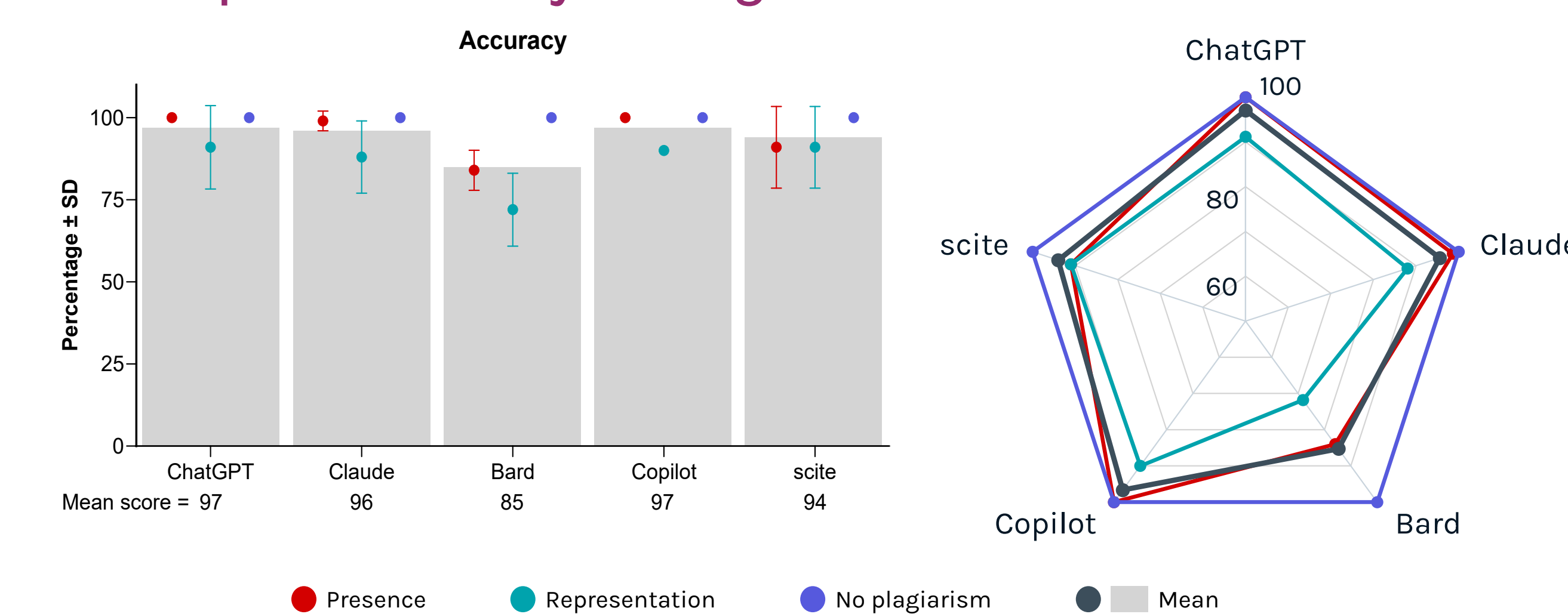
Recommendations

- All platforms failed to meet medical communication standards of accuracy and accessibility
- Outputs require rigorous verification by humans
- AI may be useful as a starting point for generating ideas in the prewriting stage

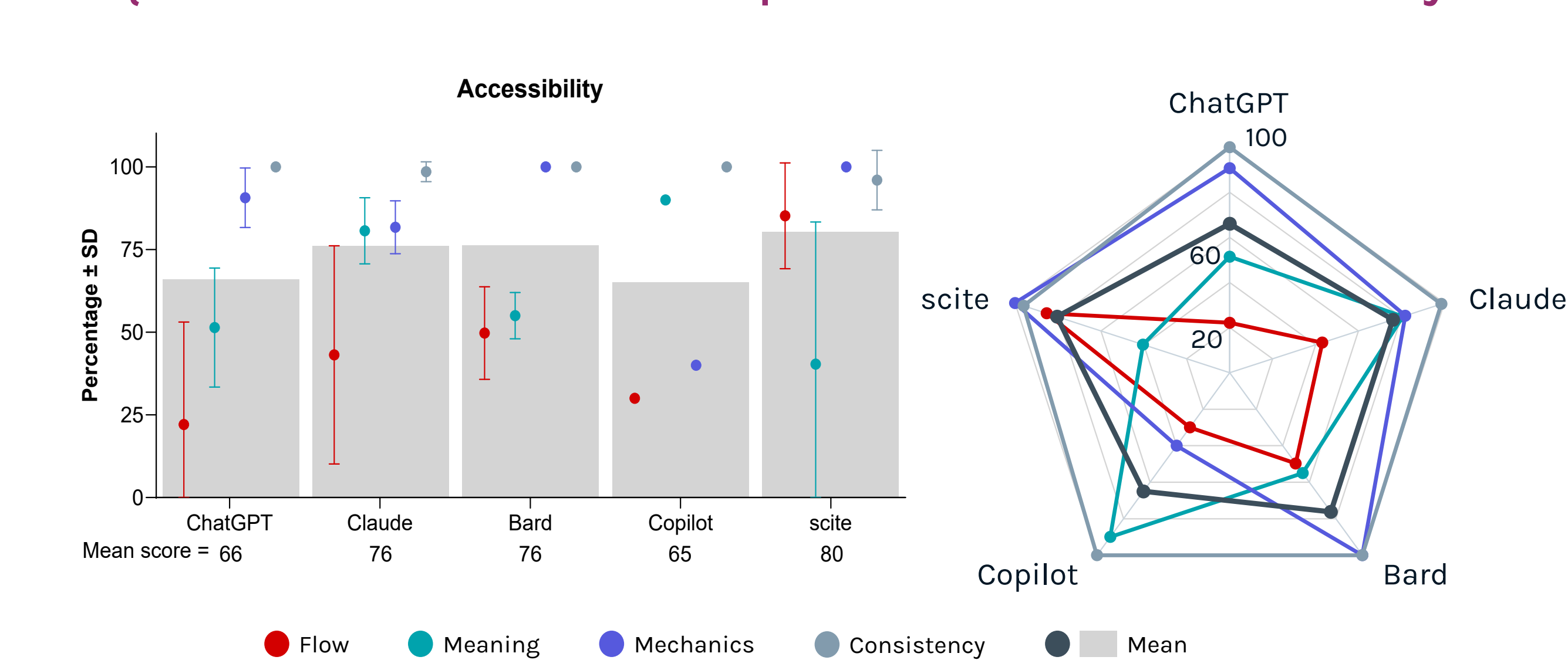
Results

Plain language summaries

High summarization accuracy across platforms with occasional misinterpretation of key messages

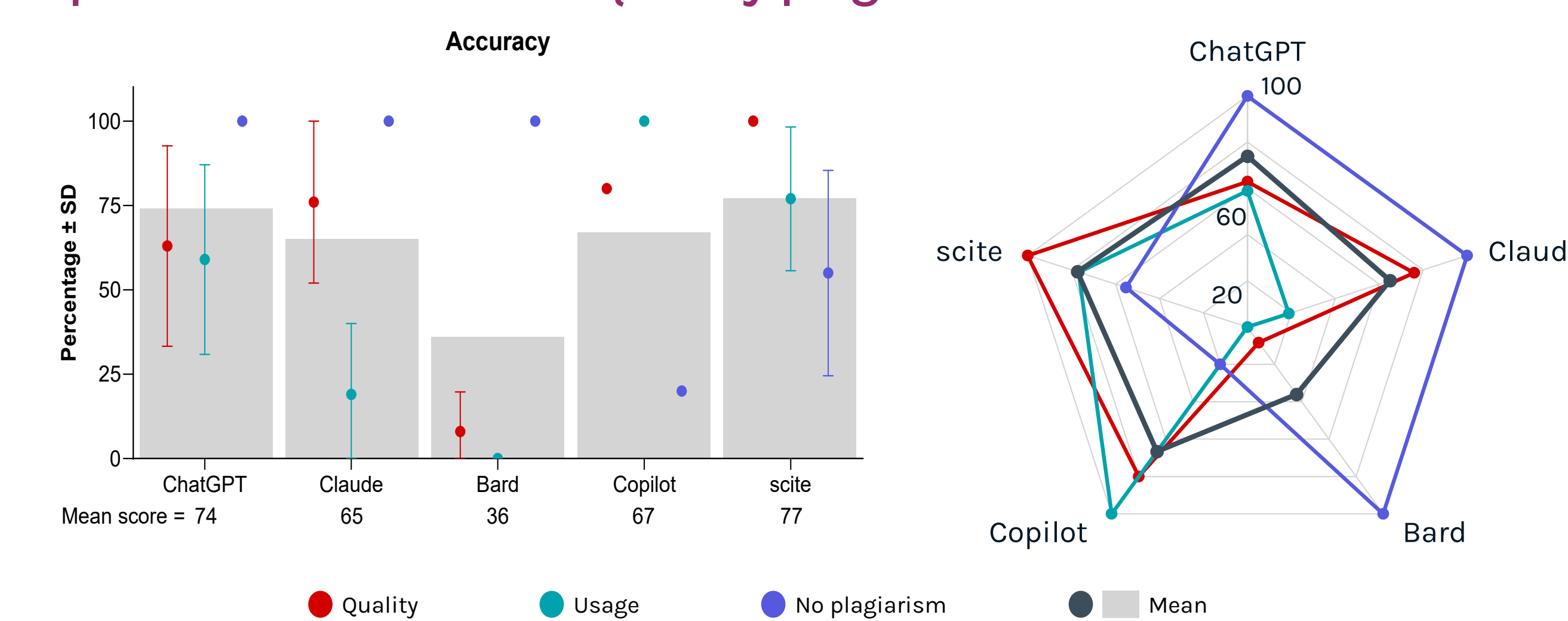


Frequent editorial errors across platforms limited accessibility

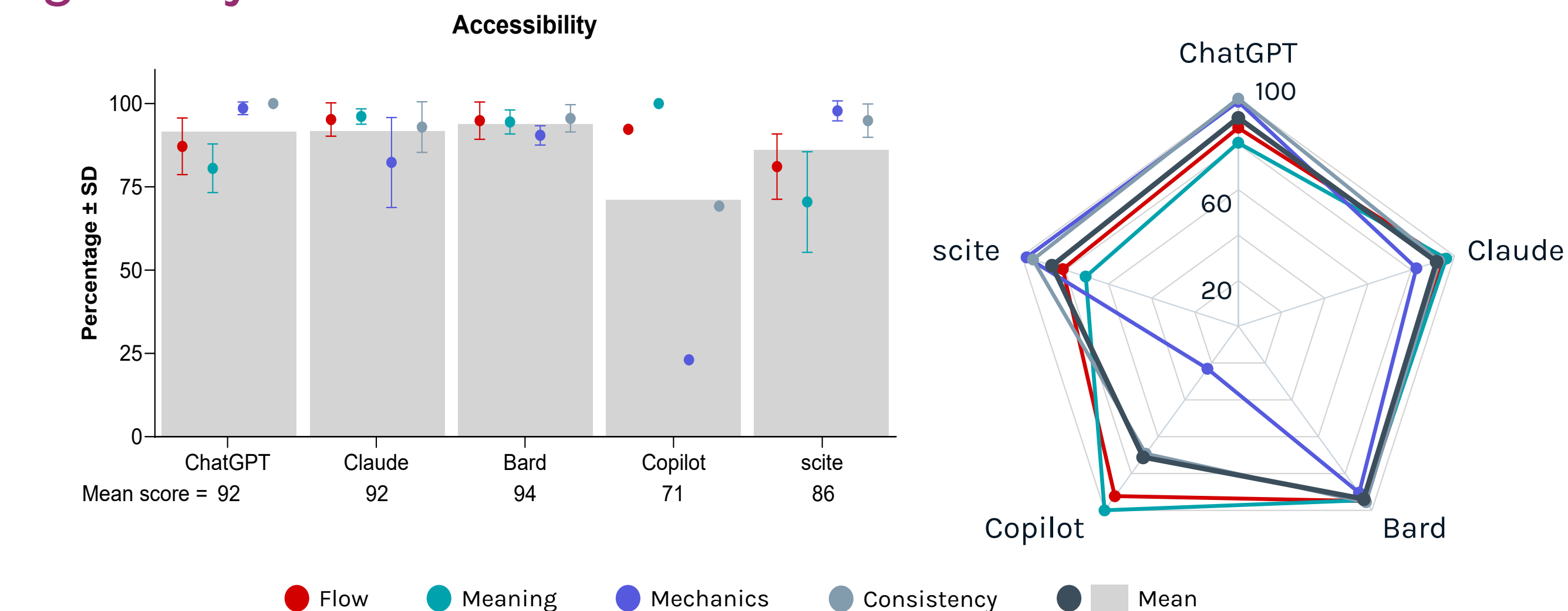


Therapeutic area summaries

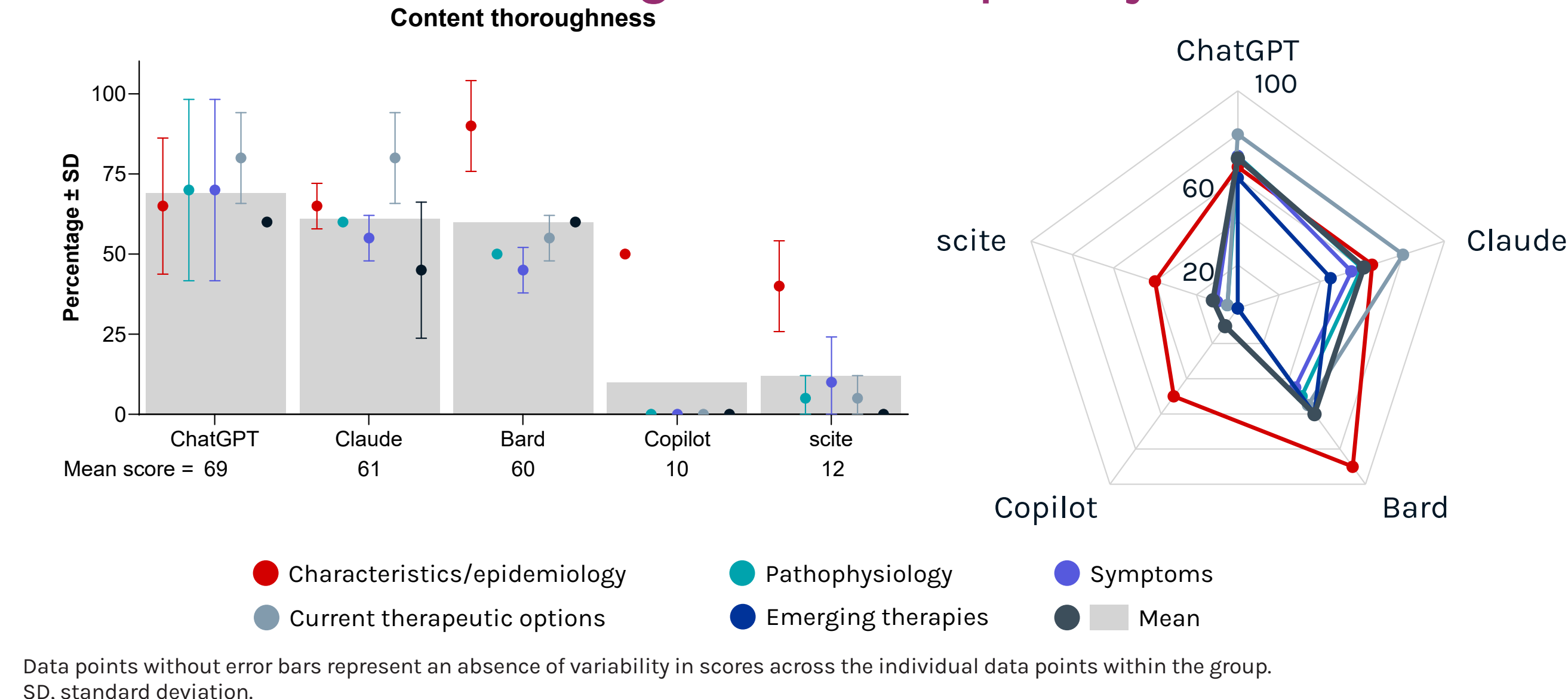
Copilot and scite had the highest usage of real references from reputable sources but frequently plagiarized content



Compared to plain language summary results, editorial errors were generally low



ChatGPT, Claude, and Bard generated comparably mediocre content



Limitations

- Only free versions of the AI platforms were used
- There were no further interaction with the AI platforms beyond the initial prompt

Acknowledgements

This study was sponsored by Red Nucleus.

All authors have no conflicts of interest to disclose.

Corresponding author

Danielle L Rubin-Shepherd (DShepherd@rednucleus.com)

References

- LeCun Y, et al. Nature. 2015;521(7553):436-44.
- Wang H, et al. Nature. 2023;620(7972):47-60.
- Martin K. Medical Writing. 2023;32(3):22-7.
- International Society for Medical Publication Professionals (ISMPP). Curr Med Res Opin. 2024;40(1):9-10.
- International Committee of Medical Journal Editors (ICMJE). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated January 2024. Accessed January 11, 2024. <https://www.icmje.org/icmje-recommendations.pdf>.
- Medical Affairs Professional Society (MAPS). Vision for generative artificial intelligence in medical affairs. Accessed January 11, 2024. <https://medicalaffairs.org/wp-content/uploads/2023/09/MAPS-WP-GenAI-Final.pdf>.
- Perren TJ, et al. N Engl J Med. 2011;365(26):2484-96.
- Griffiths CEM, et al. Lancet. 2021;397(10281):1301-15.
- Armstrong AW, et al. JAMA. 2020;323(19):1945-60.
- Drakos A, et al. Dermatol Ther (Heidelb). 2022;12(12):2715-30.



A copy of this poster and supplemental materials obtained through QR (Quick Response) code are for personal use only and may not be reproduced, retransmitted, or further distributed in any form without permission from the corresponding author.