



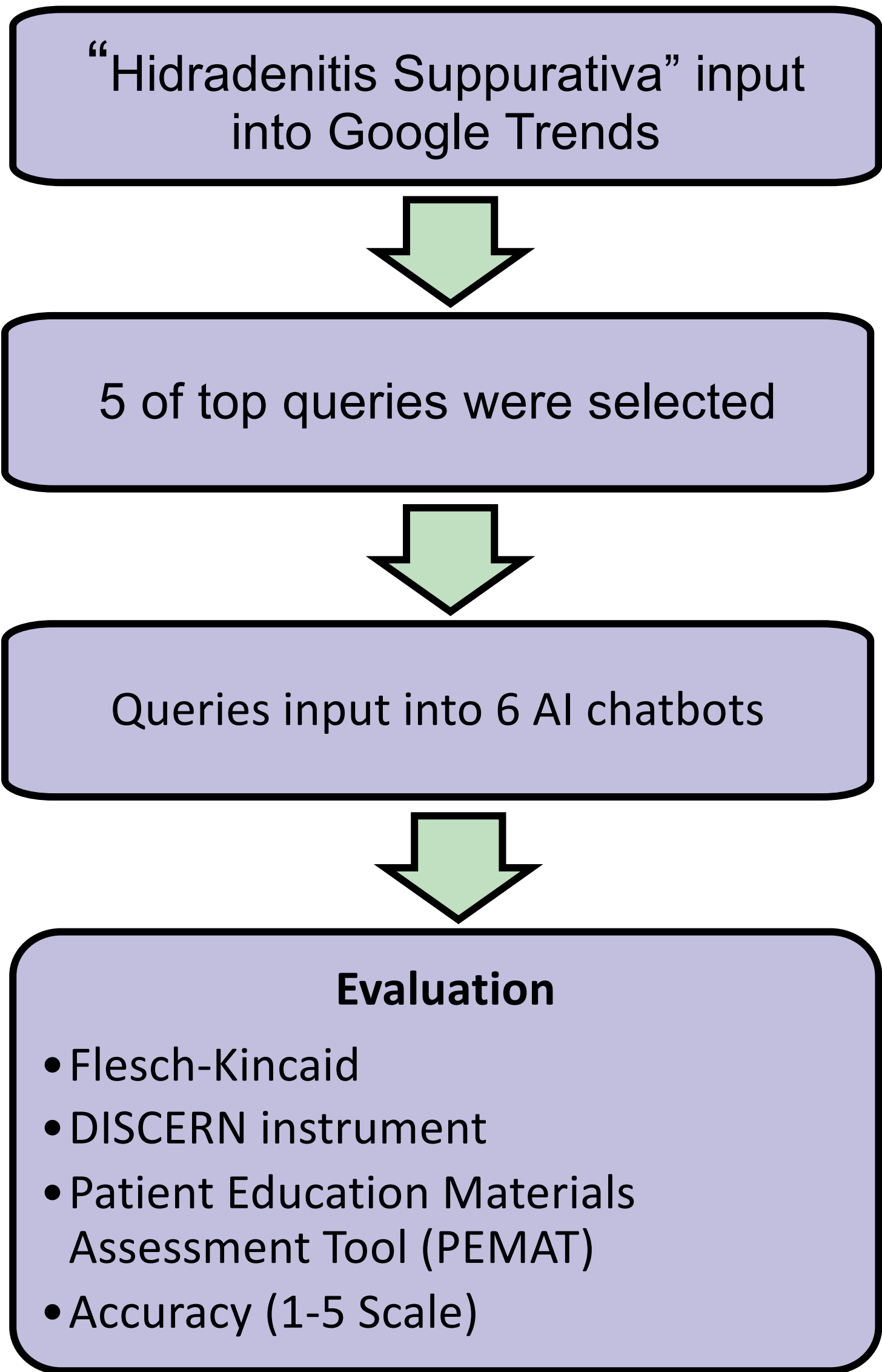
INTRODUCTION

- Patients use online platforms for medical information
- AI chatbots have become increasingly utilized
- Few studies evaluate responses to specific queries, such as hidradenitis suppurativa (HS)
- Understanding the characteristics of various chatbot outputs responses are essential.

AIMS

- Evaluate the quality, accuracy, and readability of AI chatbot responses regarding HS using established assessment instruments .

METHODS



RESULTS

Table 1: AI Chatbot Metrics for Top Searched Queries Regarding Hidradenitis Suppurativa

Measure	Chat GPT3	Chat GPT4	Bing AI	Google Bard	Perplexity	ChatSonic	Overall
DISCERN Reliability Score	2.9 [2.9 - 2.9]	2.9 [2.9 - 2.9]	3.2 [3.1 - 3.4]	3.2 [3.1 - 3.4]	3.2 [3.1 - 3.3]	3.1 [3.0 - 3.3]	2.9 [2.9 - 3.1]
DISCERN Quality Score	3.1 [2.9-3.8]	2.7 [2.1 – 4.0]	2.3 [2.0 - 3.1]	2.7 [2.7 - 2.9]	2.7 [2.4 - 3.5]	2.7 [2.4 - 3.0]	2.9 [2.4 - 3.5]
DISCERN Overall Score	3.0 [3.0-3.8]	3.0 [3.0 – 3.0]	3.0 [2.0 – 3.0]	4.0 [4.0 – 4.0]	3.0 [2.0 - 3.8]	3.0 [2.0 – 3.0]	3.0 [2.0 – 4.0]
PEMAT understandability score %	72.7 [63.6 - 72.7]	81.8 [72.7 - 90.9]	72.7 [72.7 - 72.7]	77.3 [72.7 -81.8]	54.6 [54.6 -63.6]	77.3 [72.7 - 81.8]	72.7 [61.4 - 81.8]
PEMAT actionability score %	40.0 [20.0 – 60.0]	40.0 [40.0- 60.0]	40.0 [40.0 – 40.0]	40.0 [40.0 – 60.0]	20.0 [20.0 – 40.0]	40.0 [40.0 – 60.0]	40.0 [20.0 – 45.0]
Accuracy	4.0 [2.0 – 4.0]	3.0 [2.0 – 4.0]	2.0 [2.0 – 3.0]	3.0 [2.0 – 3.0]	2.0 [2.0 – 2.0]	3.0 [2.0 – 3.0]	2.5 [2.0 – 3.0]
Flesch-Kincaid Grade Level	12.2 [11.8 - 15.1]	12.9 [12.7 - 13.6]	11.6 [11.3 - 12.8]	9.4 [8.4 - 9.5]	13.6 [13.6 - 13.7]	11.9 [9.5 - 12.4]	12.3 [10.35 - 13.6]
Word count	408 [371 - 416]	304 [304 - 322]	191 [178 - 211]	353 [335 - 398]	156 [139 - 164]	133 [114 - 148]	243.5 [166 - 348]

Table 2: AI Chatbot Output Citations for Top Searched Queries Regarding Hidradenitis Suppurativa

	Sources Cited (number of citations)
Chat GPT3	None
Chat GPT4	None
Bing AI	Mayo Clinic (4); MSN.com (3); Cleveland Clinic (2); WebMD (2); Healthline.com (2); Canadian Dermatology Association (1); NIH (1); SingleCare.com (1); VerywellHealth.com (1); American Family Physician (1); Health.com (1)
Google Bard	None
Perplexity	American Academy of Dermatology Association (6); Mayo Clinic (4); Cleveland Clinic (4); NHS (4); NIH (3); WebMD (3); MedicalNewsToday.com (1); Hidrawear.com (1); EverydayHealth.com (1); Healthline.com (1)
ChatSonic	American Academy of Dermatology Association (4); Mayo Clinic (3); Cleveland Clinic (2); NHS (2); Healthline.com (1)

CONCLUSIONS

- AI Chatbots show potential as supplemental informational tools
- Accuracy and actionability signal room for enhancement
- Readability surpasses recommendations by institutions like the AMA & NIH
- Frequent citations from reputable sources
- Equitable access to AI technology remains to be addressed
- Future studies should focus on:
 - Optimization of AI tools
 - Exploring the effectiveness of prompting techniques
 - Exploring synergistic effects of AI-provider responses

REFERENCES

1. Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl.* 2004;86(5):334-338.

2. Mendoza CS, Safdar N, Okada K, Myers E, Rogers GF, Linguraru MG. Personalized assessment of craniosynostosis via statistical shape modeling. *Med Image Anal.* 2014;18(4):635-646.

3. Readability Scoring System. <https://readabilityformulas.com/readability-scoring-system.php>. [Accessibility verified December 4, 2023]

4. The Patient Education Materials Assessment Tool (PEMAT) and User’s Guide | Agency for Healthcare Research and Quality. <https://www.ahrq.gov/health-literacy/patient-education/pemat.html>. [Accessibility verified December 4, 2023]

5. DISCERN - The DISCERN Instrument. http://www.discern.org.uk/discern_instrument.php. [Accessibility verified December 4, 2023]