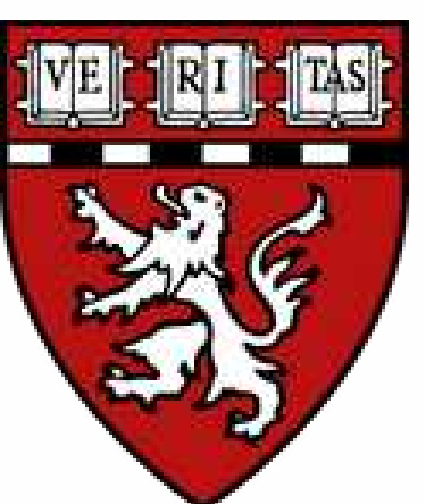


Patient Education For Gender Affirmation Surgery In The Digital Era: A Multilingual Comparison Of Web Search Vs. ChatGPT-4

Raunak Goyal, MBBS¹, Jade Tso, BS², Shivangi Saha, MBBS, MS, MCh³, Anurag Jain, MBBS, MS, MCh⁴, Gabriel Alfonso Del Corral, MD⁵, Maneesh Singhal, MBBS, MS, MCh³, Kristen Rezak, MD⁶, Kavitha Ranganathan, MD¹, Ashit Patel, MBChB⁶

¹Brigham and Women's Hospital, Boston, MA, USA, ²Harvard School Of Medicine, Program In Global Surgery And Social Change, Boston, MA, USA, ³The All India Institute Of Medical Sciences, New Delhi, India, ⁴Deen Dayal Upadhyay Hospital, New Delhi, India, ⁵MedStar Georgetown University Hospital, WA, USA, ⁶Duke University, Durham, NC, USA

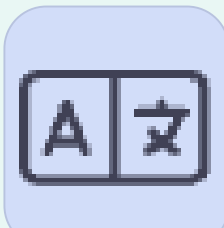


Background



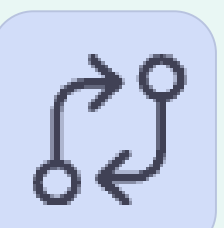
Traditional Materials

- Variable accuracy and clarity
- Variable emotional sensitivity



ChatGPT Advantages

- More accurate and comprehensive
- Clearer and better patient utility



Comparison Needed

Google search vs. ChatGPT-4 effectiveness unknown for GAS

Marginalized Communities

Improved Access



Language Barriers
Health Literacy Challenges

Underserved Areas

Reaching populations with limited resources

Non-English Speakers

Breaking down language barriers

Healthcare Systems

Improving educational material delivery

Fig1. Addressing Marginalization in Healthcare

Fig 2. Impact on Healthcare Delivery

Materials and Methods



Web Search Analysis

Top 5 web pages in English, Spanish, and Hindi



ChatGPT-4 Queries

12 frequently asked gender affirming surgery questions



Expert Evaluation

Seven board-certified plastic surgeons rated content



Statistical Analysis

Descriptive statistics and two-tailed t-tests

Results

Table 1. Multimetric, Multilingual Comparison of Web Search and ChatGPT performance

Language	Accuracy			Safety			Emotional Sensitivity			Patient Utility			Readability		
	Google	ChatGPT	P value	Google	ChatGPT	P value	Google	ChatGPT	P value	Google	ChatGPT	P value	Google	ChatGPT	P value
English	3.80 (0.06)	3.93 (0.41)	0.305	3.85 (0.65)	3.95 (0.39)	0.338	3.85 (0.37)	4.12 (0.41)	0.113	2.95 (0.64)	3.83 (0.49)	0.004	14	16	0.063
Spanish	2.60 (0.89)	3.50 (0.90)	0.040	2.40 (0.89)	3.41 (0.79)	0.003	2.40 (0.89)	2.66 (0.65)	0.250	2.20 (0.44)	3.00 (0.95)	0.048	17	21	0.022
Hindi	3.10 (0.96)	3.66 (0.55)	0.068	2.80 (0.75)	3.40 (0.55)	0.040	3.60 (0.65)	4.20 (0.54)	0.032	2.90 (0.96)	3.45 (0.54)	0.072	2	3	0.204

ChatGPT Superiority

Higher scores for accuracy, safety, emotional sensitivity, and patient utility

Language Gaps

AI bridges language barriers effectively

ChatGPT content more complex across languages

Readability Challenge

AI shows greater sensitivity to gender diverse context

Emotional Intelligence

Fig 1. Key Findings

Table Description. Accuracy, Safety, Emotional Sensitivity, and Patient Utility are measured on a Likert scale ranging from 1-5. Readability by corresponding school grade level.

Conclusions

Bridging Language Gaps

Cultural Competence

Improved Accessibility

Better Health Outcomes

Health Equity

Fig3. Potential of LLMs for Patient Education