

A Complementary Tool for a Sensitive Domain: How Users Seek Health Information Through LLMs

Luana Ponte^{1,2}[0009–0009–9834–6061] and Carla Teixeira
Lopes^{1,2}[0000–0002–4202–791X]

¹ Faculdade de Engenharia, Universidade do Porto, R. Dr. Roberto Frias, 4200-465
Porto, Portugal

² INESC TEC, Faculdade de Engenharia, Universidade do Porto, R. Dr. Roberto
Frias, 4200-465 Porto, Portugal
{up201805965,ctl}@fe.up.pt

Abstract. Conversational AI is increasingly complementing search engines as a tool for information seeking, including in sensitive domains such as healthcare. Yet, how this new information-seeking dynamic is characterised remains poorly understood. This study surveyed 270 participants to examine how users integrate LLMs alongside search engines, their satisfaction with AI-generated answers in general versus health-related searches, the health topics most often searched through LLMs, and how demographic characteristics shape these behaviours. Results show that, although increasingly adopted, LLMs remain secondary to search engines and are used even less frequently for health-related queries, reflecting greater caution with sensitive information. Satisfaction with AI-generated answers is moderate overall but significantly lower in health contexts, and searches concentrate on diseases, exercise, treatments, and mental health rather than institutional aspects of care. Demographic influences are limited overall. These findings position LLMs as a cautious complement rather than a substitute in health information-seeking, underscoring the need for systems that better support reliable, transparent, and context-sensitive use, particularly where validation and trust are critical.

Keywords: Health Information-seeking · Information-seeking behaviour
· LLM · AI-generated information

1 Introduction

The growing availability of health information online has broadened access to knowledge once confined to healthcare professionals, helping to reduce information disparities [17]. In 2025, 60% of individuals in the European Union searched for health information online, a proportion that has risen steadily since 2015 [9,10]. Traditionally, this process relied on keyword-based queries requiring users to evaluate and synthesise information across multiple sources [30], a task

closely tied to how users judge the credibility of what they find [27]. LLMs are reshaping this landscape: rather than returning ranked lists of links, they generate direct, conversational, and contextualised responses [12,13], occupying a role distinct from, yet complementary to, traditional search engines [4,28].

In the health domain, this shift carries particular weight: the sensitive, high-stakes nature of health decisions makes questions of accuracy, overreliance [23], and source credibility especially pressing [21]. Yet little is known about how users actually integrate LLMs into their health information-seeking (HIS) process, and how this compares with their broader search habits. This study addresses that gap by examining how users integrate LLMs into general and health-related information-seeking and which stages of the search process they support, which health topics are most frequently searched through LLMs and how satisfied users are with the responses obtained, and how demographic characteristics — gender, age, education, field of study and work, and chronic illness status — shape these behaviours and perceptions. To address these aims, we surveyed 270 respondents who reported using LLMs for health-related searches.

2 The use of LLMs for Health Information-Seeking

From an information behaviour perspective, HIS involves identifying information needs, selecting appropriate sources, and evaluating retrieved information [14,30]. The growing adoption of Large Language Models (LLMs) introduces new dynamics into this process, particularly regarding the credibility assessment of AI-generated content [20].

Despite the increasing visibility of LLMs, search engines and health-related websites remain the main sources of online health information [31,32]. Rather than replacing existing practices, LLMs appear to play a complementary role [19,22], being used to obtain immediate guidance, clarify complex concepts, and address informational needs outside conventional healthcare hours [8,19], with personal health queries increasing markedly during evening and nighttime hours, when traditional healthcare is least accessible [5]. Users particularly value their conversational interfaces and ability to translate medical knowledge into accessible language, thereby supporting health literacy [2,3].

Compared with traditional search engines, LLMs are perceived as faster and more practical for tasks requiring synthesis, interpretation, and advice [4,28]. However, evidence suggests that users integrate these tools into broader search routines rather than relying on them in isolation [26]. The choice between search engines and LLMs depends on contextual factors such as the seriousness of the health issue and prior experience [26]. Cross-checking behaviours remain limited, however: only a minority of users verify information obtained from LLM-based chatbots, with most preferring to consult healthcare professionals over AI tools when validating health information [32].

The growing use of LLMs in health contexts also raises concerns. Their tendency to generate plausible but inaccurate information [1], combined with users' difficulties in assessing the credibility of AI-generated content [15,25], may en-

courage overreliance and uncritical adoption. Although demographic differences in LLM use have been observed [4], their influence on HIS behaviours remains insufficiently understood.

3 Methodology

The study addresses the following research questions:

- RQ1: How do users integrate LLMs into general information-seeking practices?*
RQ2: How do users integrate LLMs into health information-seeking?
RQ3: How do demographic characteristics — gender, age, education level, field of study and work, and the existence of a diagnosed chronic illness — influence LLM adoption and behaviours in general and health information-seeking?

3.1 Questionnaire Design and Dissemination

Data were collected through an online questionnaire, developed iteratively and refined through cognitive debriefing [29]. The study was approved by the University of Porto’s Data Protection Unit.

The questionnaire consisted predominantly of closed-ended questions using Likert-type scales, complemented by a final open-ended question on participants’ experiences with LLMs in HIS. It comprised four sections covering informed consent, demographic information, general information-seeking practices, and health-related information-seeking. The full questionnaire is available in the research data repository [18].

Conditional logic ensured that only relevant questions were displayed, with participants who reported never using LLMs for general or health-related searches being directed to the end of the questionnaire.

The categories used in the study and work field questions were based on the Classification of Fields of Science and Technology (FOS) [7] and the National Classification of Education and Training Areas (CNAEF) [6], respectively. Health topics were adapted from earlier HIS research [11].

Before dissemination, the questionnaire underwent cognitive debriefing with four participants aged between 28 and 38 years, leading to minor revisions, including the addition of an “*Arts and Design*” option in the study field question.

The questionnaire was available in Portuguese and English between September 16 and 30, 2025, and was disseminated through social media, institutional email, and QR-code leaflets, with the institutional (university) channel yielding the largest share of responses.

3.2 Data Analysis

Data analysis combined data preparation, quantitative statistical analysis, and qualitative content analysis. A total of 408 responses were collected, of which

three were excluded due to the absence of informed consent. As the study compares general and health-related LLM use, analyses were restricted to participants who reported using LLMs for HIS. Two additional responses were excluded due to inconsistencies in self-reported health-related LLM use, resulting in a final sample of 270 participants.

Open-ended responses provided under the “Other” option in the study field, work field, and LLM use questions were normalised and reclassified whenever they corresponded to existing categories.

Quantitative analysis began with descriptive statistics to characterise the sample and patterns of LLM use. Cross-tabulations examined associations between demographic variables (gender, age, education, study field, and chronic illness) and the main study variables, including frequency of use, satisfaction with AI-generated responses, integration of LLMs across search stages, health topics searched, and reported experiences. Work field was excluded from these cross-tabulations, as 58.52% of participants selected “N/A” for this item – reflecting that most respondents were university students not yet in active employment – which precluded meaningful comparisons across professional categories.

Associations between categorical variables were assessed using the Chi-square test, with Cramer’s V as a measure of effect size. Statistical significance was set at $p < .05$, without correction for multiple comparisons, given the exploratory nature of the analysis. Differences between general and health-related searches in frequency of use and satisfaction were examined using the Wilcoxon signed-rank test.

Participants were also invited to describe a positive or negative experience using LLMs for HIS, yielding 38 short textual responses. These were analysed through qualitative content analysis [24], informed by sentiment analysis principles [16]. An inductive coding process yielded four categories (positive, negative, mixed, and neutral), which were manually assigned according to explicit criteria based on participants’ evaluations of usefulness, trustworthiness, and accuracy. Given the reduced sample size ($n = 38$), demographic cross-tabulations of these categories were not conducted to avoid overinterpreting potentially unstable associations.

4 Results

This section characterises the sample and reports usage patterns, satisfaction with AI-generated responses, and the influence of demographic characteristics on LLM use for HIS.

4.1 Participants’ Characterisation

The sample was predominantly female (66.7%) and young, with 67.78% of participants aged 18–24. Most respondents had completed higher education (62.3%) and reported no diagnosed chronic illness (78.9%). Participants represented diverse academic backgrounds, particularly Health Sciences (17.4%) and Social

Sciences (16.3%). More than half (58.5%) reported no work field, reflecting the high proportion of university students in the sample. Full demographic data are available in the research data repository [18].

4.2 LLMs in Users’ Information-Seeking

The following sections compare LLM use in general and health-related searches, focusing on frequency of use, integration into the search process, and satisfaction with AI-generated responses.

Frequency of LLM Use Across General and Health Contexts To contextualise LLM use in HIS, general search habits were first examined. Search engines are used more frequently, with almost all participants (91.9%) reporting using them three or more times per week, compared with 51.5% for LLMs. Although widely adopted, LLMs are therefore not yet used as intensively as traditional search engines.

A different pattern emerges for health-related searches, which are concentrated at lower frequencies, with most participants reporting monthly use and only a small proportion indicating frequent engagement. This difference is statistically significant: the Wilcoxon Signed-Rank test shows that LLMs are used less frequently for health searches than for general searches ($Z = -12.51$, $p < .001$).

Cross-tabulation analyses indicate that LLM use frequency remains broadly consistent across demographic groups. Although descriptive trends suggest slightly higher use among younger participants, individuals with higher education levels, and those in Engineering and Medical Sciences, these differences are not statistically significant, as indicated by the Chi-square tests and low Cramer’s V values.

LLMs Used for Information-Seeking Participants reported using different LLMs for general information-seeking, with *ChatGPT*-only use being the most common configuration (28.1%). The remaining participants typically combined multiple systems ($M = 2.21$, $SD = 1.13$), most frequently *ChatGPT* with *Google Gemini* (14.8%), *Microsoft Copilot* (7.8%), or both (10.7%), while *ChatGPT* and *DeepSeek* accounted for 7.0% of responses. Less frequent configurations (31.6%) were grouped under “Other combinations” to ensure adequate expected cell frequencies in the association tests. No significant associations were found between LLM combinations and demographic variables ($p > .05$).

When Are LLMs Used for Information-Seeking? Figure 1 shows the relative frequency of LLM use at each search stage. Median usage was 3 (“*Sometimes*”) for all four stages, but mean usage differed: clarifying or deepening information retrieved through the search engine showed the highest average use ($M = 3.25$, $SD = 1.11$), followed by use as a fallback when the search engine proved insufficient ($M = 3.15$, $SD = 1.15$) and exploratory use before starting

the search ($M = 3.01$, $SD = 1.15$). Use to compare or validate information found through the search engine was the least frequent ($M = 2.75$, $SD = 1.22$), with the highest proportion of “Never” responses among the four stages (18.5%).

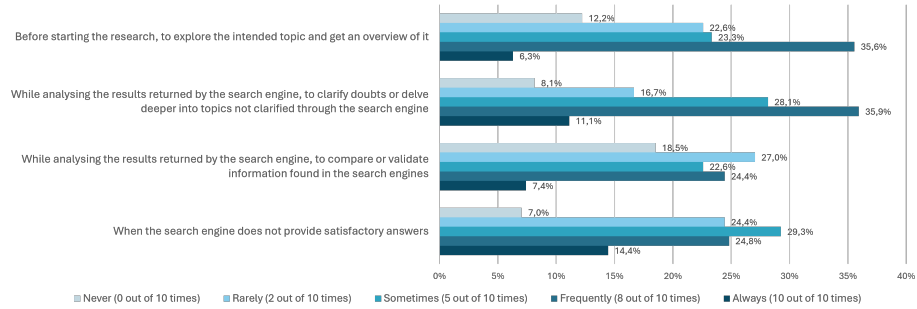


Fig. 1. Frequency (%) of LLM use across information-seeking stages in general searches (N=270)

Age and education were associated with exploratory use, with participants aged 25–44 and those with higher education reporting greater engagement. Validation behaviour varied according to gender and chronic illness status, while the fallback stage showed some variation by age and education. No significant associations were found for study field.

Health Topics Searched in LLMs Table 1 presents the statistically significant associations between demographic characteristics and health topics searched through LLMs. Although several associations were identified, effect sizes were generally small to moderate (Cramer’s $V = 0.16$ – 0.26).

Most health topics were searched infrequently, with responses concentrated in the categories “*Never*” and “*Once a month or less*”. Searches were more common for specific diseases or medical conditions, exercise, medical treatments, and mental health issues, whereas long-term care, end-of-life decisions, health insurance, and pregnancy were rarely explored through LLMs.

Age was associated with searches related to hospitals, environmental risks, memory loss, medical test results, and chronic pain, with younger participants generally reporting lower engagement. Higher education and health-related study fields were associated with more frequent searches concerning medications and medical test results.

Participants with chronic illnesses reported higher engagement with searches concerning specific conditions, medications, exercise, and experimental treatments, while no significant associations were found for gender.

Table 1. Statistically significant associations between demographic variables and frequency of health information-seeking via LLMs (N = 270)

	Health topic	χ^2 (df)	p-value	Cramer's V
Age group	Hospitals	31.61 (16)	0.01	0.17
	Environmental risks	26.44 (16)	0.05	0.16
	Memory loss	31.52 (16)	0.01	0.17
	Medical results	49.63 (16)	< 0.001	0.21
	Chronic pain	26.94 (16)	0.04	0.16
Education	Medications	23.66 (12)	0.02	0.17
	Medical results	34.45 (12)	< 0.001	0.21
Study field	Specific conditions	44.64 (28)	0.02	0.20
	Medications	45.28 (28)	0.02	0.21
	Medical results	44.82 (28)	0.02	0.20
Chronic illness	Specific conditions	12.88 (4)	0.01	0.22
	Exercise	11.21 (4)	0.02	0.20
	Medications	14.66 (4)	0.001	0.23
	Exp. treatments	11.18 (4)	0.03	0.20
	Pregnancy	17.79 (4)	< 0.01	0.26

4.3 Satisfaction with AI-generated Answers

Figure 2 shows that satisfaction with LLM responses is moderate in both general and health contexts, with responses concentrated in the “*Sometimes*” category. Satisfaction is, however, lower for health-related searches, which show a greater proportion of “*Rarely*” responses. This difference is statistically significant ($Z = -3.84, p < .001$).

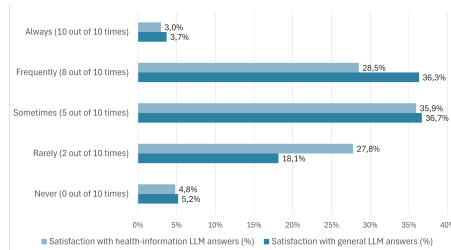
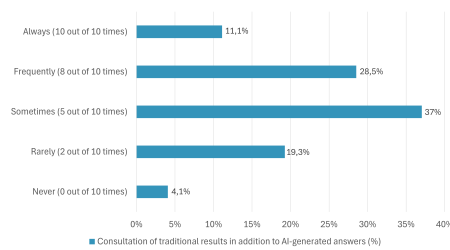
**Fig. 2.** Frequency of satisfaction with AI-generated answers in general and health-related LLM searches (N=270)**Fig. 3.** Frequency of consulting traditional web results after AI overviews in general searches (N=270)

Figure 3 presents users’ behaviour after consulting AI-generated overviews in search engines. Although responses are similarly concentrated in the “*Sometimes*” category, participants frequently consult traditional web results, suggesting that AI-generated summaries do not always fully satisfy users’ information needs.

Few demographic differences were observed. Significant associations emerged only for age in general LLM searches ($\chi^2 = 27.02$, $df = 16$, $p = .041$, $V = .16$), with younger participants reporting higher satisfaction. No significant associations were found for health-related searches or AI-generated overviews.

Participants’ Reported Experiences with LLMs for HIS Thirty-eight participants described positive or negative experiences using LLMs for HIS. Reported experiences concerned the interpretation of medical information, decision support, healthcare guidance, emotional reassurance, and health-related planning. Responses were predominantly positive (57.9%), followed by neutral (26.3%), while negative (10.5%) and mixed (5.3%) experiences were less frequent.

5 Discussion

LLMs are increasingly incorporated into users’ information-seeking practices but remain secondary and complementary to traditional search engines, particularly for health-related searches, where use is less frequent – suggesting greater caution with sensitive information (RQ1, RQ2). Across both contexts, LLMs are used mainly for exploratory, interpretative, and fallback purposes – initial overviews, clarifying complex results, or compensating for insufficient search engine results – rather than for validating information already found, with validation consistently the least common use across the search process, consistent with concerns in the literature about users’ difficulty critically assessing AI-generated content [15,25]. The prevalence of multi-platform use may itself function as an informal validation strategy, with users cross-checking outputs across systems rather than relying on a single model – a behaviour not previously captured in the literature reviewed.

Satisfaction with AI-generated answers is moderate overall but significantly lower for health-related searches, consistent with prior work on the risks of AI-generated information in sensitive domains [1,2,3,15]. A similar pattern emerges for AI-generated overviews in search engines: participants frequently revert to traditional web results afterwards, suggesting these summaries act as an initial layer requiring further validation rather than a substitute for standard search.

Health-related searches through LLMs concentrate on diseases, exercise, treatments, and mental health, while institutional topics (hospitals, insurance, end-of-life decisions) are rarely explored – indicating that LLMs are used mainly to interpret personal health situations rather than to navigate healthcare systems, in line with [5]. This is reinforced by the qualitative findings: reported experiences are predominantly positive and align with the quantitative satisfaction results. Demographic factors (RQ3) show a generally limited effect, with most associations non-significant or of small effect size; the exceptions – age,

education, study field, and chronic-illness status – align with expected patterns (e.g., health-related respondents or chronically ill participants engaging more with specific topics). Given the sample’s concentration in younger, highly educated participants, and the likely confounding between age and education level – since younger participants are more often still completing their studies – these associations should be interpreted with caution rather than generalised to the wider population. Overall, findings suggest LLMs are embedded in HIS practices but used cautiously and as a complement, shaped by topic sensitivity, perceived reliability, and the persistent need for external validation.

6 Conclusions

This study examined how users integrate LLMs into general and health-related information-seeking. LLMs are increasingly adopted but remain complementary to search engines, particularly in health contexts, where use is more cautious, exploratory, and focused on diseases, exercise, treatments, and mental health, rather than navigating healthcare systems. Satisfaction is moderate and significantly lower for health searches, reinforced by users’ frequent reliance on traditional web results after AI overviews.

These patterns point to clear limits on the current utility of LLMs in the health domain. Institutional or systemic aspects of care – hospitals, insurance, long-term care, end-of-life decisions – are rarely explored, which suggests LLMs currently support personal sense-making of health information rather than more consequential, system-level health decisions. The lower satisfaction observed in health searches, together with users’ frequent fallback on traditional sources after general LLM use, further suggests that users do not yet regard LLM-generated content, health-related or otherwise, as sufficiently reliable to forgo external validation.

The influence of demographic characteristics is limited overall, though age, education, study field, and chronic-illness status show some associations. These findings should nonetheless be interpreted with caution: although dissemination combined social media, physical leaflets, and institutional channels, the survey’s reach was disproportionately driven by university-affiliated channels, which likely explains the sample’s skew toward young, highly educated participants with a low prevalence of chronic illness. This convenience-sampling effect constrains the generalisability of the reported associations to the wider population.

Future work will explore users’ motivations, prompt (re)formulation strategies, and validation/trust practices in more depth, ideally through behavioural or log-based data. Replicating this study with a more demographically balanced – particularly age-stratified – sample would help assess the generalisability of these findings, and longitudinal designs could track how usage evolves as LLM technologies advance.

Acknowledgments. This work was supported by FCT - Fundação para a Ciência e Tecnologia, I.P. by project reference 2024.06479.BD and DOI identifier doi.org/10.54499/2024.06479.BD.

References

1. Alkaissi, H., McFarlane, S.I.: Artificial hallucinations in chatgpt: implications in scientific writing. *Cureus* **15**(2) (2023). <https://doi.org/10.7759/cureus.35179>
2. Ayo-Ajibola, O., Davis, R.J., Lin, M.E., Riddell, J., Kravitz, R.L.: Characterizing the adoption and experiences of users of artificial intelligence-generated health information in the united states: cross-sectional questionnaire study. *Journal of medical Internet research* **26**, e55138 (2024). <https://doi.org/10.2196/55138>
3. Caramancion, K.M.: Large language models vs. search engines: evaluating user preferences across varied information retrieval scenarios (2024). <https://doi.org/10.48550/arXiv.2401.05761>
4. Chatterji, A., Cunningham, T., Deming, D.J., Hitzig, Z., Ong, C., Shan, C.Y., Wadman, K.: How people use chatgpt (2025). <https://doi.org/10.3386/w34255>
5. Costa-Gomes, B., Tolmachev, P., Taysom, E., Sounderajah, V., Richardson, H., Schoenegger, P., Liu, X., Nour, M.M., Spielman, S., Way, S.F., Shah, Y., Bhaskar, M., Nori, H., Kelly, C., Hames, P., Gross, B., Suleyman, M., King, D.: Public use of a generalist LLM chatbot for health queries. *Nature Health* (2026). <https://doi.org/10.1038/s44360-026-00117-x>
6. DGEEC: CNAEF, <https://cnaef.dgeec.medu.pt/>
7. DGEEC: Classificação de Domínios Científicos e Tecnológicos (2007)
8. Eliasson, G., Gullström, T.: User perceptions of trust in generative ai for healthcare advice (2025), <https://lup.lub.lu.se/student-papers/search/publication/9202761>
9. Eurostat: Digital economy and society statistics - households and individuals (2025), https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals
10. Eurostat: Individuals - internet activities (2025). https://doi.org/10.2908/ISOC_CI_AC_I
11. Fox, S.: Health Topics (May 2011), <https://www.pewresearch.org/internet/2011/05/12/health-topics/>
12. Liu, L., Meng, J., Yang, Y.: Llm technologies and information search. *Journal of Economy and Technology* **2**, 269–277 (2024). <https://doi.org/10.1016/j.ject.2024.08.007>
13. Mannuru, N.R., Mannuru, A., Lund, B.: Large language models (llms) as a tool to facilitate information seeking behavior. *InfoScience Trends* **1**(2), 34–42 (2024). <https://doi.org/10.61186/ist.202401.01.15>
14. Marchionini, G.: Information seeking in electronic environments. No. 9, Cambridge university press (1995)
15. Nov, O., Singh, N., Mann, D.: Putting chatgpt’s medical advice to the (turing) test: survey study. *JMIR Medical Education* **9**, e46939 (2023). <https://doi.org/10.2196/46939>
16. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Now Publishers Inc (2008)
17. Percheski, C., Hargittai, E.: Health information-seeking in the digital age. *Journal of American College Health* **59**(5), 379–386 (2011). <https://doi.org/10.1080/07448481.2010.513406>
18. Ponte, L., Teixeira Lopes, C.: Survey dataset on LLM adoption and health information seeking: User behaviour, satisfaction, and demographic influences dataset (version 1) (2026). <https://doi.org/10.25747/svyn-rj84>

19. Raina, A., Mishra, P., Kumar, D., et al.: Ai as a medical ally: Evaluating chatgpt's usage and impact in indian healthcare (2024). <https://doi.org/10.48550/arXiv.2401.15605>
20. Rieh, S.Y., Danielson, D.R., et al.: Credibility: A multidisciplinary framework. *Annual review of information science and technology* **41**(1), 307–364 (2007)
21. Sayres, R., Hao, Y., Ward, A., Wang, A., Freeman, B., Zhan, S., Ardila, D., Li, J., Lee, I.C., Iurchenko, A., et al.: Towards better health conversations: The benefits of context-seeking (2026). <https://doi.org/10.1145/3772318.3793199>
22. Sezgin, E., Jackson, D.I., Kocaballi, A.B., Bibart, M., Zupanec, S., Landier, W., Audino, A., Ranalli, M., Skeens, M.: Can large language models aid caregivers of pediatric cancer patients in information seeking? a cross-sectional investigation. *Cancer Medicine* **14**(1), e70554 (2025). <https://doi.org/10.1002/cam4.70554>
23. Spatharioti, S.E., Rothschild, D., Goldstein, D.G., Hofman, J.M.: Effects of llm-based search on decision making: Speed, accuracy, and overreliance. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. pp. 1–15 (2025). <https://doi.org/10.1145/3706598.3714082>
24. Spurgin, K.M., Wildemuth, B.M.: Content analysis. Applications of social research methods to questions in information and library science pp. 297–307 (2009)
25. Sun, X., Ma, R., Zhao, X., Li, Z., Lindqvist, J., Ali, A.E., Bosch, J.A.: Trusting the search: unraveling human trust in health information from google and chatgpt (2024). <https://doi.org/10.48550/arXiv.2403.09987>
26. Wardle, C., Urbani, S., Wang, E.: Evolving health information-seeking behavior in the context of google ai overviews, chatgpt, and alexa: Interview study using the think-aloud protocol. *Journal of Medical Internet Research* **27**, e79961 (2025). <https://doi.org/10.2196/79961>
27. Wathen, C.N., Burkell, J.: Believe it or not: Factors influencing credibility on the web. *Journal of the American society for information science and technology* **53**(2), 134–144 (2002)
28. White, R.W.: Advancing the search frontier with ai agents. *Communications of the ACM* **67**(9), 54–65 (2024). <https://doi.org/10.1145/3655615>
29. Willis, G.B.: *Cognitive interviewing: A tool for improving questionnaire design*. sage publications (2004)
30. Wilson, T.D.: Models in information behaviour research. *Journal of documentation* **55**(3), 249–270 (1999). <https://doi.org/10.1108/EUM0000000007145>
31. Xiao, Y., Zhou, K.Z., Liang, Y., Shu, K.: Understanding the concerns and choices of public when using large language models for healthcare (2024). <https://doi.org/10.48550/arXiv.2401.09090>
32. Yun, H.S., Bickmore, T.: Online health information-seeking in the era of large language models: cross-sectional web-based survey study. *Journal of medical Internet research* **27**, e68560 (2025). <https://doi.org/10.2196/68560>