informatics

Review

Dr. Google vs. Dr. ChatGPT in Online Health Self-Consultation: A Scoping Review of Accuracy, Bias, and Actionability (2023–2025)

Magdalena Trillo-Domínguez [1,*], Juan Ignacio Martin-Neira [2] and María Dolores Olvera-Lobo [1]

[1] Department of Information and Communication Sciences, University of Granada, 18071 Granada, Spain; molvera@ugr.es
[2] Faculty of Communication, Universidad de los Andes, Santiago 755000, Chile; jimartinn@uandes.cl
* Correspondence: mtrillo@ugr.es

Abstract

The rapid adoption of generative artificial intelligence (AI) systems has transformed health information seeking, raising questions about their role as intermediaries in non-professional health self-consultation. This study compares Google Search and ChatGPT as paradigmatic models of algorithmic mediation of health information, focusing on accuracy, biases, information quality and potential harms. A scoping review was conducted following the PRISMA-ScR framework. Empirical studies published between 2023 and 2025 were retrieved from PubMed/MEDLINE, Web of Science (WoS) and Scopus. After screening and eligibility assessment, 63 original empirical studies were included. The results indicate that ChatGPT consistently outperforms Google Search in terms of factual accuracy and information quality, achieving moderate to high DISCERN scores (4–5 out of 5) and showing moderate to strong correlations with expert clinical evaluations. Users also tend to value ChatGPT responses positively due to their clarity, coherence and perceived empathy. However, these advantages coexist with significant structural limitations. Hallucinations are reported in an estimated 31–45% of references, source provenance remains opaque, linguistic complexity is high, and actionability is limited, with only around 40% of responses providing clearly actionable guidance. In contrast, Google Search offers greater source traceability and verifiability, but at the cost of fragmented information and higher exposure to commercial content. The review identifies critical research gaps related to behavioural impacts, critical health literacy, equity of access, professional integration and vulnerable contexts. Overall, the findings highlight the need for hybrid human–AI models, professional mediation and critical AI literacy to ensure safe, equitable and trustworthy use of generative AI in public health communication.

Keywords: health self-consultation; generative artificial intelligence; ChatGPT; google search; health information seeking; algorithmic intermediation; misinformation and bias; actionability

Academic Editor: Antony Bryant

Received: 9 January 2026
Revised: 17 February 2026
Accepted: 3 March 2026
Published: 5 March 2026

1. Introduction and Research Context

From "Dr Google" to "Dr ChatGPT". In these turbulent times of intense digitalisation, paradigm shifts and technological innovations that characterise the third millennium, health-related content has become an exceptional barometer of the population's information consumption preferences and habits and, more specifically, of the ways in which medical knowledge is sought, interpreted and used.

 https://doi.org/10.3390/informatics13030041

In Europe, online searching for health information has become an established practice. More than half of the population aged between 16 and 74 were already consulting health-related content in 2020, a figure that rose to 63% of internet users in 2024, placing this activity among the most frequent digital uses, alongside news consumption and interpersonal communication [1]. The global COVID-19 pandemic intensified this trend and brought to the fore the risks of an infodemic, understood as an "overabundance of information (accurate or not) that makes it difficult to access reliable sources and clear guidance" [2] by exposing citizens to an unprecedented volume of content ranging from information produced by official bodies to inaccurate, manipulated or outright false materials [3].

In this scenario, the algorithmic systems that mediate access to information acquire critical relevance, in a context marked by the erosion of trust in traditional institutions—governments, companies, media and NGOs—to meet the population's health needs [4], and by the challenges that the automation of information flows poses for verification, transparency and public trust [5]. These tensions are particularly salient in the health domain, where the reliability of information affects not only the cognitive sphere but may also have direct consequences for individual well-being, patient safety and trust in health systems.

Over recent decades, health self-consultation has been articulated primarily through search engines such as Google Search, which direct users to institutional websites, commercial content or materials generated by other patients. This model fragments information across multiple links and delegates to users the task of comparing sources, assessing credibility and constructing their own interpretations, with associated risks such as exposure to misinformation. The recent emergence of large language models (LLMs) and generative artificial intelligence systems, such as ChatGPT, introduces a structural change: rather than offering lists of results, these systems generate direct, synthesised and seemingly personalised responses based on large volumes of text. This transition reconfigures the search experience and raises questions about how traditional search engines and generative AI tools compare, in terms of accuracy, biases, information quality and potential risks, when used for non-professional health self-consultation.

1.1. Algorithmic Intermediaries

Within the processes of information mediation that characterise digital environments, search engines such as Google have historically functioned as preferred intermediaries between users and a heterogeneous ecosystem of health-related information sources. In this model, responsibility for assessing the credibility, relevance and applicability of information rests largely with the user, while ranking algorithms prioritise criteria such as popularity, search engine optimisation or interaction over scientific rigour, in line with the dynamics described in the context of eHealth and Medicine 2.0 [6].

This form of algorithmic intermediation may give rise to phenomena such as cyberchondria, understood as the unfounded escalation of concerns about common symptoms resulting from progressive exposure to content about severe or rare diseases, with persistent effects on users' anxiety [7]. From the perspective of digital health literacy (eHealth literacy), the search for health information has been conceptualised as a practice that requires comparing sources, recognising institutional authority and managing informational uncertainty—competences that are central to mitigating the risks derived from exposure to fragmented information or content of uneven quality.

The emergence of large language models and generative artificial intelligence systems introduces a significant shift in these processes of intermediation. Conversational tools such as ChatGPT represent a transition from a logic of link retrieval to the provision of direct, synthesised and seemingly personalised responses, partially displacing the burden

https://doi.org/10.3390/informatics13030041

of selection and synthesis from the user to the system. In the health domain, this transformation raises questions regarding the factual accuracy of generated responses, the extent to which they reproduce or amplify biases present in training data, and the implications of algorithmic opacity and the lack of source traceability in contexts of health self-consultation.

As Hersh [8] notes, generative models are not designed to respond adequately to all types of health-related information needs, particularly those that require authority, source traceability and constant updating. In contexts of health self-consultation, these limitations are far from negligible: users do not merely seek an answer, but also the ability to evaluate who is making a claim, on what evidence and in which context. However, according to Hersh, systems based on large language models carry the risk of offering closed responses, with incomplete or non-existent references, which may hinder critical evaluation on the part of patients or citizens.

1.2. The "Synthetic Tsunami" Unleashed by Generative AI

The accelerated adoption of generative artificial intelligence systems has contributed to what some authors have described as a "synthetic tsunami", characterised by the massive proliferation of algorithmically generated content across the digital ecosystem [9]. However, this expansion has not been accompanied by a solid scientific or regulatory consensus regarding the reliability, biases and potential risks associated with the use of these systems as sources of medical information by individuals without specialised training [10].

As outlined at the outset of this article, public and media debate has tended to frame a simplified dichotomy between "Dr Google" and "Dr ChatGPT". Existing research remains fragmented across experimental studies on diagnostic accuracy, analyses of specific cases, user perception surveys, and technical or institutional assessments, dispersed across disciplines and often located outside traditional academic publishing channels. To date, there are still relatively few studies that critically compare both models of access to health information from a comprehensive perspective, which justifies the need to systematically map the current state of knowledge on accuracy, biases, information quality and potential harms associated with the use of traditional search engines and generative AI systems in non-professional health self-consultation contexts.

Although studies such as that by Choi and Ahn [11] emphasise the cognitive and contextual challenges faced by health information seekers on the web, they have been developed within the context of traditional search engine results. The emergence of generative models introduces new dynamics of interaction and trust that require these conclusions to be re-evaluated from the perspectives of intelligibility, attribution of authority and the management of uncertainty in automated responses.

In this context, the present study undertakes a scoping review of the scientific literature published between 2023 and 2025, focusing on empirical studies that analyse Google Search and ChatGPT as health information intermediaries. The aim is to synthesise the available evidence on their accuracy, biases, information quality and potential risks, as well as to identify research gaps and emerging challenges for public health communication in an environment increasingly mediated by generative AI systems.

1.3. Research Questions and Study Objectives

In light of the challenges identified in the algorithmic intermediation of health information, this study is structured around two main research questions:

RQ1. What empirical evidence (2023–2025) documents the differences between Google Search and ChatGPT in non-professional health self-consultation contexts with regard to accuracy/information quality, biases/hallucinations, and readability/actionability?

RQ2. What research gaps and opportunities emerge from the comparative analysis of both systems from a public health communication perspective?

Accordingly, the primary objective of this study is to systematically map the recent empirical literature (2023–2025) comparing Google Search and ChatGPT as health information intermediaries, synthesising evidence on their accuracy, biases, information quality and potential risks in non-professional self-consultation. In this vein, the specific objectives are to:

SO1 characterise the methodological approaches, clinical specialties and analytical dimensions of the included studies

SO2 identify patterns in accuracy, biases and readability, as well as their impact on user trust and decision-making

SO3 detect research gaps and propose priority directions for future research in AI-mediated health communication.

2. Methodology

2.1. Selection of Analysed Tools and Consulted Databases

Although a variety of generative artificial intelligence models and systems are currently available—such as Gemini, Perplexity, Claude or DeepSeek—the available empirical scientific literature has overwhelmingly focused on the analysis of ChatGPT. This is largely due to its early availability, ease of access and widespread adoption by the general population since its launch [12–15]. In parallel, Google maintains a clearly dominant position within the information search ecosystem, operating de facto as the hegemonic search engine at a global scale.

In line with this context, the present review is limited to studies that compare both systems within the domain of health information searching, with the aim of ensuring empirical relevance, comparability of results and social applicability. With regard to data sources, PubMed/MEDLINE was prioritised due to its focus on public health, health literacy and patient behaviour, alongside Scopus and Web of Science (WoS) for its interdisciplinary coverage across the fields of technology, health and communication.

2.2. Inclusion and Exclusion Criteria and Search Strategy

The temporal delimitation of the corpus was strictly restricted to scientific output published between 2023 and 2025, a period that coincides with the widespread public access to ChatGPT following its official launch [16]. This temporal boundary ensures a focus on studies conducted in a context of effective co-existence between traditional search engines and conversational generative AI systems, while excluding earlier research that does not address this specific interaction.

The final sample comprised exclusively original empirical studies published in the standard IMRaD format—Introduction, Methods, Results and Discussion—[17], systematically excluding other academic and journalistic formats such as doctoral theses, expert commentaries, editorials, narrative reviews, institutional reports and conference proceedings. Duplicate records were also removed, as were those whose research objectives did not align directly with the comparative analysis of Google Search and ChatGPT in the health information domain (Table 1).

https://doi.org/10.3390/informatics13030041

Table 1. Inclusion and exclusion criteria.

Criterion	Inclusion Criteria	Exclusion Criteria
Publication period	Articles published between 2023 and 2025	Publications prior to 2023
Document type	Original empirical articles in IMRaD format	Commentaries, editorials, theses, narrative reviews, conference proceedings
Language	English, Spanish, Portuguese	Other languages
Central topic	Explicit comparison between Google Search and ChatGPT in health information seeking	Studies focusing on a single system; non-health contexts (education, business, etc.)
Methodological design	Experimental, observational and comparative studies, and surveys with original data	Theoretical articles or opinion pieces without empirical data
Population or context	Non-professional health self-consultation (patients and general public)	Exclusively professional use or assisted clinical contexts

The definition and iterative refinement of the search equations were carried out between 4 and 6 December 2025 and 12 February 2026, during which multiple combinations of terms were tested in order to optimise sensitivity and specificity. Once the final search strings had been consolidated, the systematic search was conducted between 6 and 9 December 2025 in both Scopus and PubMed/MEDLINE, and between 13 and 15 February at WoS, with the aim of ensuring comprehensive retrieval of the relevant literature available at that time. The full electronic search strategy for Scopus/WoS is provided in Supplementary File S1.

2.3. Study Selection Procedure

The review initially identified 72 records that met the preliminary characteristics for inclusion in the corpus. Following the screening of titles, abstracts and keywords, the exclusion criteria were applied, resulting in a final sample of 63 documents for analysis. Figure 1 presents the study selection process developed in accordance with the PRISMA-ScR methodology, as well as the elements considered in the analytical process.

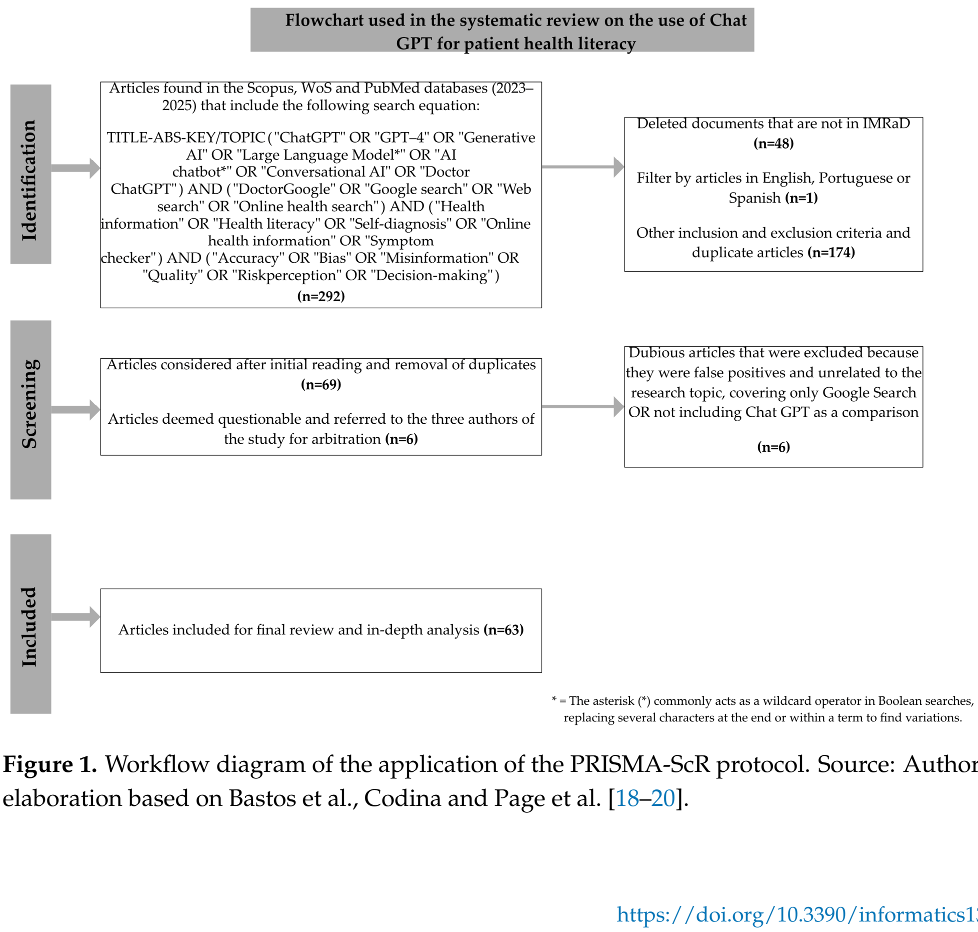

Figure 1. Workflow diagram of the application of the PRISMA-ScR protocol. Source: Authors' own elaboration based on Bastos et al., Codina and Page et al. [18–20].

The reading of all documents and their subsequent classification were conducted independently by two of the authors between 10 and 22 December, and between 13 and 16 February. Full-text selection (*n* = 69) was also performed independently by two authors. Discrepant cases (*n* = 6; 8.7%) were resolved through consensus discussion with the third author, ensuring consistency in the application of the inclusion criteria. A detailed description of the texts and their characteristics can be found in the Supplementary Materials.

2.4. Data Extraction, Analysis and Validation

The systematisation of variables across the 63 analysed documents was carried out with technological assistance in order to streamline processing, using the Google NotebookLM platform. This artificial intelligence tool is restricted to the content provided by the researcher and generates responses in accordance with the iteratively refined prompts supplied to it [21,22].

In line with these working procedures, this phase involved the design of precise instructions through successive testing until the outputs were aligned with the objectives of the study. Once the data had been obtained, the information was cross-checked against the original data extraction table using NVivo software version 1.7.2. In this way, the most frequently used terms were identified—allowing for the generation of a word cloud that facilitated the detection of patterns and related concepts—and a technique of idea and element analysis was applied through a simple text query [23] to highlight the main reflections or lines of research of interest. The data extraction table for the included studies is available as Supplementary File S2.

In accordance with international standards [24], the authors ensure that the results presented in this study, although assisted by NotebookLM pro version, were rigorously audited and corrected by the researchers, who fully assumed responsibility for the analytical work and the interpretation of the findings. In addition, ChatGPT was used exclusively for stylistic revision and translation, without contributing to the generation of scientific knowledge.

3. Results

The analysis of the 63 empirical studies included in this scoping review makes it possible to identify consistent patterns in the comparison between Google Search and ChatGPT as health information intermediaries in non-professional self-consultation contexts. The results are organised into two main analytical blocks: (a) the characteristics of the corpus and the overall landscape of findings; and (b) the specific challenges and emerging lines of research identified in the recent literature.

3.1. Characteristics of the Analysed Corpus

The scientific output analysed (Figure 2) is largely concentrated in the most recent years. Of the total number of included studies (=63), 36 were published in 2025 (57%), 24 in 2024 (38%), and only 3 in 2023 (5%), indicating an accelerated growth of academic interest in the analysis of generative artificial intelligence systems and traditional search engines in the health domain.

With regard to medical specialty, the studies are unevenly distributed. Research focusing on orthopaedics and surgery predominates, with 16 studies (25%). This grouping reflects a substantial overlap in clinical focus—for example, in areas such as trauma and joint replacement—and aligns with the way several reviewed articles self-categorised under general surgery while primarily addressing musculoskeletal conditions. Although the disciplinary distinction is acknowledged, this decision facilitates comparative synthesis within the framework of the present review.

https://doi.org/10.3390/informatics13030041

Categorization by health topics	
Primary Area of Specialization	**Number of Studies (n=63)**
Orthopedics	16
Oncology	11
General medicine	9
Ear, Nose, and Throat Medicine	7
Other topics	6
Urology	4
Ophtalmology	3
Plastic/Cosmetic Surgery	3
Mental health	2
Healthcare systems	2

Classification by Year of Research	
Year of Publication	**Number of Studies (n=63)**
2025	36
2024	24
2023	3

Classification by Country of Origin of the First Author	
Country of Affiliation of First Author	**Number of Studies (n=63)**
United States	38
Australia	5
China	4
Turkey	4
Japan	3
United Kingdom	2
Other nations (Canada, Spain, Israel, Italy, Jordan, Malaysia, Singapore)	7

Figure 2. Characteristics of the texts analysed in the scoping review. Source: Authors' own elaboration.

This is followed by oncology with 11 studies (17%) and general medicine with 9 studies (14%). The remainder of the corpus is distributed across specialties such as ophthalmology, urology and health systems studies, among others.

From a geographical perspective, the United States accounts for the largest number of publications, with 38 studies (60%). A further 18 studies (29%) list first authors with institutional affiliations in countries such as China, Australia, Türkiye, Japan or the United Kingdom. The remaining corpus corresponds to studies originating from single countries, including Spain or Italy.

A detailed description of the 63 studies, including authorship, year of publication, medical specialty and main findings, is provided in full in the Supplementary Materials.

3.2. Comparison Between ChatGPT and Google Search as Health Information Intermediaries

3.2.1. Reliability of Responses and Patient Trust

The studies analyzed conclude that tools based on large language models play a complementary role in health communication between professionals and patients [25]. From the users' perspective, responses generated by ChatGPT are positively valued for their conciseness, discursive coherence and clarity of exposition when compared with traditional search engines and are even perceived as "clearer and more precise than speaking with a doctor in person" [26–29].

Several studies indicate that this type of response contributes to generating positive emotional reactions and to reducing the anxiety associated with seeking health information. In this regard, Mendel et al. [30] report that users describe interactions with ChatGPT as "more human than traditional search engines" and "much more personalised". Likewise, some studies suggest that these tools may help to reduce information disparities among patients with low socioeconomic status by offering personalised and easily accessible health guidance [31].

In terms of factual accuracy, several studies consistently report a higher performance of GPT-4 compared with Google Search in responding to specific medical queries [32–35]. Using the DISCERN instrument to assess the quality of health information, responses generated by language models achieved moderate to high quality scores (4–5 out of 5), consistently outperforming Google Search [36–40].

Specific clinical evaluations reinforce these findings. In the field of breast cancer, a moderate to strong correlation has been observed between ChatGPT's assessments and those conducted by human experts [41]. Similarly, 67.5% of consulted orthopaedic specialists reported being "satisfied" or "very satisfied" with the AI-generated descriptions of femoroacetabular impingement syndrome [42]. On the other hand, in scoliosis surgery, 90% of AI responses were "excellent" compared to 0% for Google [43].

3.2.2. Biases, Hallucinations and Source Transparency

Despite the positive results in terms of accuracy and user experience, multiple studies consistently identify hallucinations as one of the main risks associated with the use of ChatGPT in health information contexts. For instance, several studies noted that ChatGPT occasionally produced medically implausible terms, for example, as confusing Pembrolizumab (immunotherapy for cancer) with Palivizumab (an antiviral) which are not recognised in clinical practice [44] or in the study on radon gas in Spain, ChatGPT invented safety level regulations (becquerels) that did not exist in current Spanish legislation [45]. These hallucinations primarily manifest in the generation of fabricated references, in the absence of evidence or citations without a verifiable real origin [30,46–49].

The reviewed literature attributes this phenomenon to several factors. On the one hand, to the lack of immediate updating of training models, which may lead to the generation of inaccurate or outright false information [50,51]. On the other hand, to cross-linguistic interference between languages—particularly between English and Chinese—which may result in the invention of non-existent words or terms [44]. In addition, several studies document that ChatGPT systematically denies the specific provenance of its sources when explicitly questioned, significantly affecting the transparency and traceability of the information provided [46,52].

With regard to actionability, understood as the capacity to offer clear, specific and actionable instructions for patients, the results indicate a lower performance of ChatGPT compared with Google Search. Approximately only 40% of AI-generated responses include clearly identifiable practical guidance, whereas traditional search engines tend to provide greater operational orientation [53–56].

3.2.3. Readability, Accessibility and Response Variability

Although several studies highlight that ChatGPT prioritises governmental and academic sources over the commercial websites that frequently appear in prominent positions in Google Search [53,57,58], this tendency towards the "academisation" of responses introduces relevant challenges in terms of readability and accessibility.

The reviewed literature consistently documents that responses generated by ChatGPT exhibit high levels of linguistic complexity, typically corresponding to grades 12–14 on the Flesch–Kincaid scale, which hampers comprehension among the general population [59–66]. While some studies show that readability can be improved through specific iterative prompting without loss of accuracy [33,67–69], this process requires advanced technical competences on the part of the user [70].

Moreover, relevant temporal inconsistencies are observed, whereby identical queries may generate different responses on different days, a phenomenon that is particularly problematic in otorhinolaryngological emergency contexts [48]. In parallel, it is noted that paid versions, such as GPT-4, systematically outperform free versions of the models in terms of accuracy [27,69,71].

Overall, access to reliable health-related responses is conditioned by socioeconomic factors such as educational level, income and access to premium models, which influence both the accuracy of the information received and users' perceptions of quality [46].

3.3. Emerging Challenges and Identified Gaps

Despite the broad consensus regarding the complementary potential of generative artificial intelligence in health consultation, the reviewed literature highlights the persistence of significant reservations among both patients and health professionals, particularly in relation to hallucinations and informational biases [45,72].

The main gaps identified in the analysed corpus relate to the following aspects:

1. Actual behavioural impact. There is a lack of longitudinal studies assessing the concrete behavioural effects of ChatGPT use, such as diagnostic delays, inappropriate self-treatment or therapeutic adherence, beyond the textual accuracy of responses.
2. Critical health literacy. None of the reviewed studies analyse educational interventions aimed at fostering critical use of ChatGPT, including the validation of hallucinations, effective prompt iteration or recognition of the system's limitations.
3. Equity of access. Paid versions, such as GPT-4, demonstrate an estimated accuracy advantage of between 15% and 20% over free versions, thereby deepening socioeconomic gaps in access to reliable health information [27,69].
4. Professional integration. There is an absence of operational models that systematically integrate AI-mediated consultation with human medical validation.
5. Vulnerable contexts. Populations with low health literacy and countries with low- and middle-income levels are clearly underrepresented in the existing literature.
6. An emerging challenge is the lack of verification. One study [73] reveal that only 19% of users cross-check the information provided by AI, increasing the risk of accepting subtle errors as truths.

The comparative synthesis presented in Figure 3 summarises the main advantages associated with ChatGPT—particularly in terms of accuracy and discursive clarity—alongside the identified risks, such as hallucinations, source opacity and inequalities in access.

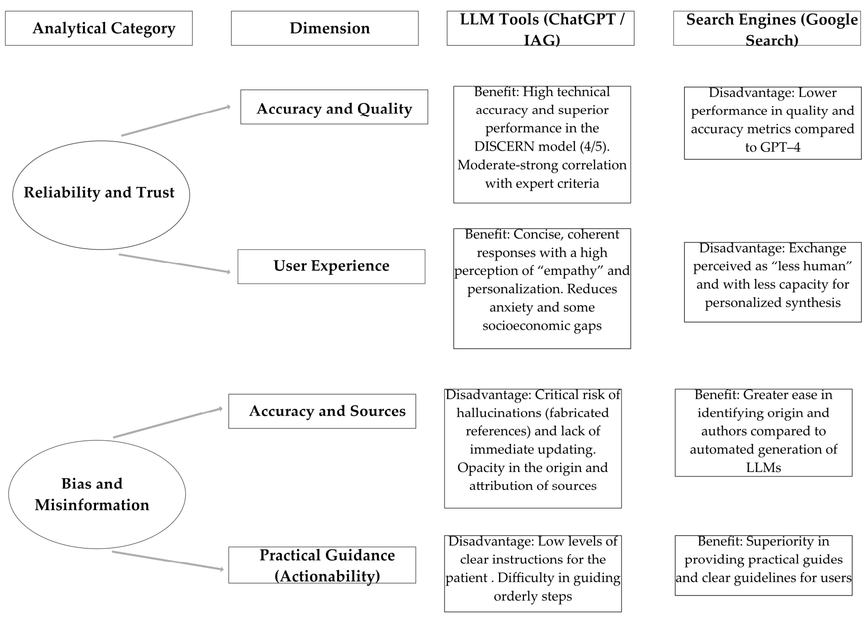

Figure 3. Advantages and disadvantages in the comparison between ChatGPT and Google Search for health consultation. Source: Authors' own elaboration.

4. Discussion

4.1. Main Findings

The 63 empirical studies analysed indicate that ChatGPT consistently demonstrates superior performance compared with Google Search in terms of factual accuracy and information quality in non-professional health self-consultation contexts. In particular, 24 of the reviewed studies report that responses generated by ChatGPT achieve scores of 4–5 out of 5 on the DISCERN scale, as well as moderate to strong correlations with evaluations conducted by clinical experts [36,41].

Nevertheless, this technical superiority coexists with significant structural limitations. Several studies document the presence of hallucinations in an estimated range of 31% to 45% of references, which may be fabricated or non-verifiable [30,46]. In addition, ChatGPT responses tend to exhibit high levels of linguistic complexity, typically ranging between grades 12 and 14 on the Flesch–Kincaid scale, which limits their accessibility for large segments of the population [59]. Finally, the capacity of these responses to guide clinical action is limited, with approximately 40% of instructions being clearly actionable, a performance that is lower than that observed for traditional search engines [53].

4.2. Implications for Health Professionals and Health Communication

The integration of generative language models into health self-consultation processes opens up relevant opportunities to optimise interactions between health professionals and patients, while at the same time requiring a substantial redefinition of the professional role. Although 18 of the analysed studies document a technical superiority of ChatGPT in terms of factual accuracy, this advantage does not necessarily translate into a greater capacity to guide clinical action. In particular, the limited actionability of AI-generated responses—with only 40% of instructions being clearly actionable—highlights the need for human mediation to translate algorithmic information into clinically relevant and contextually appropriate recommendations [39,53–55].

In this scenario, the reviewed literature identifies a set of emerging functions for health professionals in environments increasingly mediated by generative artificial intelligence. First, professionals act as expert curators, responsible for validating, contextualising and adapting AI-generated responses through individualised clinical judgement. This role is particularly relevant in light of evidence showing high levels of specialist satisfaction with the descriptive accuracy of ChatGPT responses—such as the case of the 67.5% of orthopaedic specialists satisfied with the description of femoroacetabular impingement syndrome—but more limited assessments regarding their therapeutic usefulness [42].

In addition, the results highlight a key function of health professionals as mediators of health literacy, aimed at enabling patients to interact critically with these systems. This mediation includes both the teaching of strategies to formulate more effective prompts and the development of competences to identify hallucinations, errors and other limitations inherent to language models [33,67,70].

A third identified function is that of expectation managers, insofar as professionals must explicitly explain the structural limitations of generative artificial intelligence, including opacity regarding source provenance, temporal variability of responses and dependence on the model version used. This task is essential to prevent dynamics of overtrust or the inappropriate substitution of professional medical consultation [48].

Finally, the analysed studies emphasise the need for health professionals to act as a bridge between AI-mediated consultation and health systems, participating in the design of protocols that systematically integrate AI-generated information with processes of professional validation. This integration is particularly critical in the case of vulnerable populations, where inequalities in health literacy and access to resources may amplify the risks associated with the unmediated use of these technologies [31].

From a broader perspective, these findings suggest that public health communication must evolve towards hybrid models that preserve the epistemic authority of health professionals while leveraging the synthetic and communicative capacities of generative artificial intelligence. Recent studies have also begun to explore the use of large language models to analyse patient-generated texts and online experience narratives [74], showing that AI systems can help identify unmet health needs and patterns of patient dissatisfaction beyond traditional evaluation mechanisms [75]. In this sense, generative AI may function not only as an informational intermediary but also as a communicative mediator that structures and articulates health needs expressed through dispersed patient experiences. This emerging line of research reinforces the interpretation of AI-mediated consultation as a process of communication and interpretation, rather than solely one of information retrieval or accuracy.

4.3. Priority Directions for Future Research

The analysis of the corpus allows for the identification of a set of critical gaps that delineate priority directions for future research in the field of algorithmic intermediation of health information. One of the most relevant shortcomings is the absence of longitudinal studies assessing the actual behavioural impact of ChatGPT use, incorporating indicators such as diagnostic delays, inappropriate self-treatment and therapeutic adherence, beyond the evaluation of the textual accuracy of responses.

In addition, the reviewed literature highlights the need to develop educational interventions aimed at critical artificial intelligence literacy applied to health. Such interventions should enable users to validate potential hallucinations, effectively iterate prompts and recognise the structural limitations of algorithmic systems [33]. Closely related to this, it is essential to further examine algorithmic equity, given that paid versions of language models, such as GPT-4, outperform free versions in terms of accuracy by an estimated range of 15–20%, thereby contributing to the widening of socioeconomic gaps in access to reliable health information [27].

Another relevant gap identified in the corpus is the absence of empirically validated human–AI hybrid models that systematically integrate algorithmic consultation with processes of professional validation. This is compounded by the need to broaden the research focus towards currently underrepresented vulnerable contexts, including populations with low health literacy and low- and middle-income countries, where the risks associated with the unmediated use of these technologies may be amplified.

Accordingly, the available evidence points to the priority of developing prospective studies that allow for the measurement of real harms associated with the use of generative artificial intelligence systems, as well as the design of hybrid literacy programmes that coordinately integrate artificial intelligence, health professionals and patients.

Beyond the research gaps, the findings also reveal four critical implications for public health communication. First, the results reinforce the need for indispensable professional mediation, in which generative artificial intelligence acts as a complementary tool rather than a substitute for health consultation, requiring expert curation processes to contextualise and validate algorithmic responses. Second, hybrid literacy in health and artificial intelligence emerges as a priority, particularly aimed at empowering vulnerable populations in the critical use of these systems, including the formulation of effective prompts and the identification of hallucinations or errors. Third, the findings underscore the importance of equitable AI system design that addresses gaps derived from differential access to paid and free versions—with estimated accuracy differences of between 15% and 20%—as well as low levels of health literacy affecting approximately 40% of the population. Finally, the review highlights the need to prioritise longitudinal research that goes beyond the evaluation of textual accuracy and examines real harms associated with the use of these systems, such as diagnostic delays or inappropriate self-treatment.

5. Conclusions

This study confirms that Google Search and ChatGPT embody distinct paradigms of algorithmic intermediation of health information, each associated with specific strengths and structural risks in non-professional self-consultation contexts. While ChatGPT stands out for its higher factual accuracy—with DISCERN scores of 4–5 out of 5—its narrative coherence, and a high user perception of empathy, it also presents relevant limitations, such as the systematic presence of hallucinations (31–45% of responses), opacity regarding source provenance, and a limited capacity to guide clinical action, with only around 40% of responses being clearly actionable. By contrast, Google Search offers greater source traceability and verifiability, albeit at the cost of significant information fragmentation and greater exposure to commercial content.

In a context marked by a persistent infodemic, the responsible integration of generative artificial intelligence into health information seeking requires simultaneous progress in algorithmic governance, critical education and the development of human–AI hybrid models that preserve trust, patient safety and the quality of health communication, without renouncing the opportunities for technological innovation offered by these systems. Within this framework, the present study contributes to empirically delineating the strengths, risks and tensions associated with the use of generative language models as health information intermediaries, providing an analytical basis for the development of policies, professional practices and future research.

Supplementary Materials: The following supporting information can be downloaded from the Zenodo repository at https://zenodo.org/records/18868188, accessed on 12 December 2025, Supplementary File S1: Full electronic search strategy for Scopus/WoS (Word format); Supplementary File S2: Data extraction table for the included studies (Excel format); Supplementary File S3: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist (accessed on 2 March 2026). Reference [76] is cited in Supplementary Material File S3.

Author Contributions: M.T.-D.: Conceptualization; Methodology; Resources; Investigation; Writing—original draft and review; Validation. J.I.M.-N.: Conceptualization; Resources; Software; Data curation; Visualization; Writing—original draft. M.D.O.-L.: Conceptualization; Formal analysis; Methodology; Validation; Project administration; Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the project PID 2022-14015OB-100 funded by MCIN/AEI/10.13039/501100011033 (Ministry of Science and Innovation, State Research Agency, Spain) and by "ERDF A way of making Europe" (European Union).

Institutional Review Board Statement: Not applicable. This study did not involve human participants or animal subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.18510248.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

1. European Commission. *Digital Economy and Society Statistics—Households and Individuals*; Eurostat: Luxembourg, 2024. Available online: https://acortar.link/lRzA7B (accessed on 12 December 2025).
2. World Health Organization. Managing the COVID-19 Infodemic. 2020. Available online: https://acortar.link/g1V2yF (accessed on 12 December 2025).
3. Martin-Neira, J.I.; Trillo-Domínguez, M.; y Olvera Lobo, M.D. Comunicación científica tras la crisis del COVID-19: Estrategias de publicación en TikTok en el tablero transmedia. *Rev. Lat. Comun. Soc.* **2023**, *81*, 109–132. [CrossRef]
4. Edelman. Edelman Trust Barometer Special Report: Trust and Health. 2025. Available online: https://www.edelman.com/trust/2025/trust-barometer/special-report-health (accessed on 12 December 2025).
5. IBERIFIER. *Analysis of Trends and Innovations in the Media Ecosystem in Spain and Portugal (2025–2030)*; European Digital Media Observatory: Navarra, Spain, 2024. Available online: https://acortar.link/hc5s3r (accessed on 12 December 2025).
6. Eysenbach, G. Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. *J. Med. Internet Res.* **2008**, *10*, e22. [CrossRef]
7. White, R.W.; Horvitz, E. Cyberchondria: Studies of the escalation of medical concerns in web search. *ACM Trans. Inf. Syst.* **2009**, *27*, 1–37. [CrossRef]
8. Hersh, W. Search still matters: Information retrieval in the era of generative AI. *J. Am. Med. Inform. Assoc.* **2024**, *31*, 2159–2161. [CrossRef] [PubMed]
9. Nori, H.; King, N.; McKinney, S.M.; Carignan, D.; Horvitz, E. Capabilities of GPT-4 on medical challenge problems. *arXiv* **2023**. [CrossRef]
10. World Health Organization. Ethics and Governance of Artificial Intelligence for Health. 2023. Available online: https://www.who.int/publications/i/item/9789240029200 (accessed on 14 December 2025).
11. Choi, H.; Ahn, S. Classifications, changes, and challenges of online health information seekers during COVID-19. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9495. [CrossRef]
12. Cruz, A.; Mora, R.; Andonova, V.; Rosales Torres, C.S.; Carrasco, C.; Castillo Leska, A. *fAIr Tech Radar: Exploring the Adoption of Artificial Intelligence in Latin America and the Caribbean*; Inter-American Development Bank: New York, NY, USA, 2025. [CrossRef]
13. Haque, M.A.; Li, S. Exploring ChatGPT and its impact on society. *AI Ethics* **2025**, *5*, 791–803. [CrossRef]
14. Hu, K. *ChatGPT Sets Record for Fastest-Growing User Base*; Reuters: London, UK, 2023. Available online: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (accessed on 14 December 2025).
15. Skjuve, M.; Brandtzaeg, P.B.; Følstad, A. Why do people use ChatGPT? Exploring user motivations for generative conversational AI. *First Monday* **2024**, *29*, 1–23. [CrossRef]
16. OpenAI. Presentamos ChatGPT. 2022. Available online: https://bit.ly/4qpxJCB (accessed on 14 December 2025).
17. Codina, L. El modelo IMRyD de artículos científicos: ¿qué es y cómo se puede aplicar en humanidades y ciencias sociales? *Hipertext.Net* **2022**, *24*, 1–8. [CrossRef]
18. Bastos, S.; Lopezosa, C.; Tous-Rovirosa, A. La evolución y el impacto del SEO en el periodismo en los últimos cinco años: Revisión sistemática. *Estud. Sobre Mensaje Periodís.* **2024**, *30*, 25–34. [CrossRef]
19. Codina, L. Revisiones tradicionales, sistemáticas o de alcance: ¿cómo elegir el tipo de revisión de la literatura que corresponde en cada caso? *Infonomy* **2024**, *2*, e24021. [CrossRef]
20. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ. Clin. Res. Ed* **2021**, *372*, n71. [CrossRef]
21. Codina, L. NotebookLM: Inteligencia Artificial para Usos Académicos, Curación de Contenidos y ¿Revisiones de la Literatura? 2025. Available online: https://bit.ly/4p7YR8b (accessed on 16 December 2025).
22. Shor, R.; Greene, E.A.; Sumberg, L.; Weingrad, A.B. AI Tools in Academia: Evaluating NotebookLM as a Tool for Conducting Literature Reviews. *Psychiatry* **2025**, 1–10. [CrossRef]
23. Trigueros-Cervantes, C.; Rivera-García, E.; Rivera-Trigueros, I. *Técnicas Conversacionales y Narrativas: Investigación Cualitativa Con Software NVivo*; Escuela Andaluza de Salud Pública/Universidad de Granada: Granada, Spain, 2018. Available online: https://bit.ly/3y4O0oh (accessed on 20 December 2025).
24. COPE Council. COPE Position—Authorship and AI—English. 2023. Available online: https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools (accessed on 12 December 2025).
25. Nian, P.P.; Saleet, J.; Magruder, M.; Wellington, I.J.; Choueka, J.; Houten, J.K.; Saleh, A.; Razi, A.E.; Ng, M.K. ChatGPT as a Source of Patient Information for Lumbar Spinal Fusion and Laminectomy: A Comparative Analysis Against Google Web Search. *Clin. Spine Surg.* **2024**, *37*, E394–E403. [CrossRef]
26. Casey, J.C.; Dworkin, M.; Winschel, J.; Molino, J.; Daher, M.; Katarincic, J.A.; Gil, J.A.; Akelman, E. ChatGPT: A concise Google alternative for people seeking accurate and comprehensive carpal tunnel syndrome information. *Hand Surg. Rehabil.* **2024**, *43*, 101757. [CrossRef]

27. Mastrokostas, P.G.; Mastrokostas, L.E.; Emara, A.K.; Salman, M.; Wellington, I.J.; Ginalis, E.; Dalton, J.; Houten, J.K.; Khalsa, A.S.; Saleh, A.; et al. GPT-4 as a source of patient information for cervical disc arthroplasty: A comparative analysis against Google web search. *J. Spine Surg.* **2025**, *11*, 450–462. [CrossRef]
28. Yoseph, E.T.; Gonzalez-Suarez, A.D.; Lang, S.; Desai, A.; Hu, S.; Zygourakis, C. Patient perspectives on AI: A pilot study comparing large language model and physician-generated responses to routine cervical spine surgery questions. *Artif. Intell. Surg.* **2024**, *4*, 267–277. [CrossRef]
29. Lin, X.; Bai, L.; Zhao, X.; Cai, L.; Yang, J. Online platform vs. doctors: A comparative exploration of congenital cataract patient education from virtual to reality. *Front. Artif. Intell.* **2025**, *8*, 1548385. [CrossRef] [PubMed]
30. Mendel, T.; Singh, N.; Mann, D.M.; Wiesenfeld, B.; Nov, O. Laypeople's Use of and Attitudes Toward Large Language Models and Search Engines for Health Queries: Survey Study. *J. Med. Internet Res.* **2025**, *27*, e64290. [CrossRef] [PubMed]
31. Beohon, B.; Lewis, J.E.; Nguyen, P.; Dao, M.Q.; Ghogomu, M.; El Ayadi, A.; Wolf, S.E.; Song, J. Evaluating ChatGPT's Utility in Addressing Socioeconomic Disparities in Burn Patients: A Comparative Study with Google. *J. Burn. Care Res.* **2025**, *47*, 113–119. [CrossRef]
32. Aydin, M.; Aral, F.; Dasci, M.F.; Surucu, S.; Mahirogullari, M.; Citak, M. A comparative analysis of ChatGPT and Google in providing quality and clinical relevance of responses to patients' frequently asked questions on robotic-assisted total knee arthroplasty. *Arch. Orthop. Trauma Surg.* **2025**, *145*, 477. [CrossRef] [PubMed]
33. Cohen, S.A.; Fisher, A.C.; Xu, B.Y.; Song, B.J. Comparing the Accuracy and Readability of Glaucoma-related Question Responses and Educational Materials by Google and ChatGPT. *J. Curr. Glaucoma Pract.* **2024**, *18*, 110–116. [CrossRef] [PubMed]
34. Li, K.; Peng, Y.; Li, L.; Liu, B.; Huang, Z. Evaluating ChatGPT's Utility in Biologic Therapy for Systemic Lupus Erythematosus: Comparative Study of ChatGPT and Google Web Search. *JMIR Form. Res.* **2025**, *9*, e76458. [CrossRef]
35. Sezgin, E.; Chekeni, F.; Lee, J.; Keim, S. Clinical Accuracy of Large Language Models and Google Search Responses to Postpartum Depression Questions: Cross-Sectional Study. *J. Med. Internet Res.* **2023**, *25*, e49240. [CrossRef] [PubMed]
36. Kolac, U.C.; Karademir, O.M.; Ayik, G.; Kaymakoglu, M.; Familiari, F.; Huri, G. Can popular AI large language models provide reliable answers to frequently asked questions about rotator cuff tears? *JSES Int.* **2025**, *9*, 390–397. [CrossRef]
37. Musheyev, D.; Pan, A.; Loeb, S.; Kabarriti, A.E. How Well Do Artificial Intelligence Chatbots Respond to the Top Search Queries About Urological Malignancies? *Eur. Urol.* **2024**, *85*, 13–16. [CrossRef]
38. Patel, K.; Radcliffe, R. Evaluating the Readability and Quality of Bladder Cancer Information from AI Chatbots: A Comparative Study Between ChatGPT, Google Gemini, Grok, Claude and DeepSeek. *J. Clin. Med.* **2025**, *14*, 7804. [CrossRef]
39. Stapleton, P.; Santucci, J.; Thet, M.; Lawrentschuk, N.; Dodds, L.; Cundy, T.; Sathianathen, N. Quality of information on hypospadias from artificial intelligence chatbots: How safe is AI for patient and family information? *J. Pediatr. Urol.* **2025**, *21*, 1551–1556. [CrossRef]
40. Kianian, R.; Carter, M.; Finkelshtein, I.; Eleswarapu, S.V.; Kachroo, N. Application of Artificial Intelligence to Patient-Targeted Health Information on Kidney Stone Disease. *J. Ren. Nutr. Off. J. Counc. Ren. Nutr. Natl. Kidney Found.* **2024**, *34*, 170–176. [CrossRef]
41. Fushimi, A.; Terada, M.; Tahara, R.; Nakazawa, Y.; Iwase, M.; Shibayama, T.; Kotti, S.; Yamashita, N.; Iesato, A. Assessing the quality of Japanese online breast cancer treatment information using large language models: A comparison of ChatGPT, Claude, and expert evaluations. *Breast Cancer* **2025**, *32*, 960–969. [CrossRef]
42. Chen, Y.; Zhang, S.; Tang, N.; George, D.M.; Huang, T.; Tang, J.P. Using Google web search to analyze and evaluate the application of ChatGPT in femoroacetabular impingement syndrome. *Front. Public Health* **2024**, *12*, 1412063. [CrossRef]
43. Tekin, S.B.; Ince, K.; Tekin, B.G.; Servet, E.; Bozgeyik, B. Evaluation of Google and ChatGPT responses to common patient questions about scoliosis. *Spine Deform.* **2025**, *14*, 3–18. [CrossRef] [PubMed]
44. Menz, B.D.; Modi, N.D.; Abuhelwa, A.Y.; Ruanglertboon, W.; Vitry, A.; Gao, Y.; Li, L.X.; Chhetri, R.; Chu, B.; Bacchi, S.; et al. Generative AI chatbots for reliable cancer information: Evaluating web-search, multilingual, and reference capabilities of emerging large language models. *Eur. J. Cancer* **2025**, *218*, 115274. [CrossRef]
45. Pascual-Presa, N.; Fernández-Pichel, M.; Losada, D.E.; García-Orosa, B. Web Search or Conversation with an Artificial Intelligence? Analysis of Misinformation and Relevance in the Case of Radon Gas. *Prof. Inf.* **2024**, *33*, 1–15. [CrossRef]
46. Burns, C.; Bakaj, A.; Berishaj, A.; Hristidis, V.; Deak, P.; Equils, O. Use of Generative AI for Improving Health Literacy in Reproductive Health: Case Study. *JMIR Form. Res.* **2024**, *8*, e59434. [CrossRef]
47. Shen, O.Y.; Pratap, J.S.; Li, X.; Chen, N.C.; Bhashyam, A.R. How Does ChatGPT Use Source Information Compared with Google? A Text Network Analysis of Online Health Information. *Clin. Orthop. Relat. Res.* **2024**, *482*, 578–588. [CrossRef] [PubMed]
48. Soon, S.; Perry, B. Paging Dr. ChatGPT: Safety, accuracy and readability of ChatGPT in ENT emergencies. *Aust. J. Otolaryngol.* **2025**, *8*, 8. [CrossRef]
49. 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. *J. Med. Internet Res.* **2025**, *27*, e79961. [CrossRef]

50. Hristidis, V.; Ruggiano, N.; Brown, E.L.; Ganta, S.R.; Stewart, S. ChatGPT vs. Google for Queries Related to Dementia and Other Cognitive Decline: Comparison of Results. *J. Med. Internet Res.* **2023**, *25*, e48966. [CrossRef]
51. Tan, S.Y.T.; Sng, G.G.R.; Lee, P.C. Accuracy of Large Language Model Responses Versus Internet Searches for Common Questions About Glucagon-Like Peptide-1 Receptor Agonist Therapy: Exploratory Simulation Study. *JMIR Form. Res.* **2025**, *9*, e78289. [CrossRef]
52. Hunter, N.; Allen, D.; Xiao, D.; Cox, M.; Jain, K. Patient education resources for oral mucositis: A google search and ChatGPT analysis. *Eur. Arch. Oto-Rhino-Laryngol.* **2025**, *282*, 1609–1618. [CrossRef] [PubMed]
53. Dubin, J.A.; Bains, S.S.; Chen, Z.; Hameed, D.; Nace, J.; Mont, M.A.; Delanois, R.E. Using a Google Web Search Analysis to Assess the Utility of ChatGPT in Total Joint Arthroplasty. *J. Arthroplast.* **2023**, *38*, 1195–1202. [CrossRef] [PubMed]
54. Taimeh, D.; Atef, A.; Shawish, H.A.; Atrash, M.; Qaralleh, M.; Flaifl, Y. Evaluating online health information on PIFP: A cross-platform analysis of content quality and readability. *Sci. Rep.* **2025**, *15*, 40578. [CrossRef]
55. Türe, N.; Tahir, E.; Enver, N. Readability, understandability, and quality of online education materials and large language models for retrograde cricopharyngeal muscle dysfunction. *Eur. Arch. Oto-Rhino-Laryngol.* **2025**, *282*, 4711–4720. [CrossRef] [PubMed]
56. Ito, S.; Furukawa, E.; Okuhara, T.; Okada, H.; Kiuchi, T. Comparison of Japanese Mpox (Monkeypox) Health Education Materials and Texts Created by Artificial Intelligence: Cross-Sectional Quantitative Content Analysis Study. *JMIR AI* **2025**, *4*, e70604. [CrossRef]
57. Tharakan, S.; Klein, B.; Bartlett, L.; Atlas, A.; Parada, S.A.; Cohn, R.M. Do ChatGPT and Google differ in answers to commonly asked patient questions regarding total shoulder and total elbow arthroplasty? *J. Shoulder Elb. Surg.* **2024**, *33*, e429–e437. [CrossRef]
58. Yang, F.; Ma, J.; Liu, M.; Du, Z. The Impact of Google Search versus ChatGPT on Patient Understanding and Potential Adherence in PICC Line Care: A Comparative Analysis. *Patient Prefer. Adherence* **2025**, *19*, 3275–3284. [CrossRef]
59. Alamleh, S.; Mavedatnia, D.; Francis, G.; Le, T.; Davies, J.; Lin, V.; Lee, J.J.W. Readability, Reliability, and Quality Analysis of Internet-Based Patient Education Materials and Large Language Models on Meniere's Disease. *J. Otolaryngol. Head Neck Surg.* **2025**, *54*. [CrossRef]
60. Benaim, E.H.; O'Rourke, S.P.; Dillon, M.T. What Do People Want to Know About Cochlear Implants: A Google Analytic Study. *Laryngoscope* **2025**, *135*, 840–847. [CrossRef]
61. Dihan, Q.; Chauhan, M.Z.; Eleiwa, T.K.; Hassan, A.K.; Sallam, A.B.; Khouri, A.S.; Chang, T.C.; Elhusseiny, A.M. Using Large Language Models to Generate Educational Materials on Childhood Glaucoma. *Am. J. Ophthalmol.* **2024**, *265*, 28–38. [CrossRef]
62. Hlavinka, W.J.; Sontam, T.R.; Gupta, A.; Croen, B.J.; Abdullah, M.S.; Humbyrd, C.J. Are large language models a useful resource to address common patient concerns on hallux valgus? A readability analysis. *Foot Ankle Surg. Off. J. Eur. Soc. Foot Ankle Surg.* **2025**, *31*, 15–19. [CrossRef]
63. Mastrokostas, P.G.; Mastrokostas, L.E.; Emara, A.K.; Wellington, I.J.; Ginalis, E.; Houten, J.K.; Khalsa, A.S.; Saleh, A.; Razi, A.E.; Ng, M.K. GPT-4 as a Source of Patient Information for Anterior Cervical Discectomy and Fusion: A Comparative Analysis Against Google Web Search. *Glob. Spine J.* **2024**, *14*, 2389–2398. [CrossRef] [PubMed]
64. Raquepo, T.M.; Tobin, M.J.; Gettings, M.; Yamin, M.; Lee, B.T.; Cauley, R.P. A multimetric health literacy analysis of phalloplasty techniques: Comparing artificial intelligence and online resources. *J. Plast. Reconstr. Aesthetic Surg.* **2025**, *100*, 166–169. [CrossRef]
65. Cohen, S.A.; Brant, A.; Fisher, A.C.; Pershing, S.; Do, D.; Pan, C. Dr. Google vs. Dr. ChatGPT: Exploring the Use of Artificial Intelligence in Ophthalmology by Comparing the Accuracy, Safety, and Readability of Responses to Frequently Asked Patient Questions Regarding Cataracts and Cataract Surgery. *Semin. Ophthalmol.* **2024**, *39*, 472–479. [CrossRef] [PubMed]
66. Mastrokostas, P.G.; Lavi, A.B.; Zhang, B.B.; Mastrokostas, L.E.; Liu, S.; Connors, K.M.; Hashem, J. GPT-4 as a Source of Patient Information for Carpal Tunnel Surgery: A Comparative Analysis Against Google Web Search. *J. Am. Acad. Orthop. Surg.* **2025**, *33*, e1429–e1439. [CrossRef]
67. Cohen, S.A.; Yadlapalli, N.; Tijerina, J.D.; Alabiad, C.R.; Chang, J.R.; Kinde, B.; Mahoney, N.R.; Roelofs, K.A.; Woodward, J.A.; Kossler, A.L. Comparing the Ability of Google and ChatGPT to Accurately Respond to Oculoplastics-Related Patient Questions and Generate Customized Oculoplastics Patient Education Materials. *Clin. Ophthalmol.* **2024**, *18*, 2647–2655. [CrossRef] [PubMed]
68. Shen, S.A.; Perez-Heydrich, C.A.; Xie, D.X.; Nellis, J.C. ChatGPT vs. web search for patient questions: What does ChatGPT do better? *Eur. Arch. Oto-Rhino-Laryngol.* **2024**, *281*, 3219–3225. [CrossRef]
69. Simmich, J.; Ross, M.H.; Russell, T.G. Assessing the Capability of Large Language Models for Navigation of the Australian Health Care System: Comparative Study. *JMIR AI* **2025**, *4*, e76203. [CrossRef]
70. Fanning, J.E.; Escobar-Domingo, M.J.; Foppiani, J.; Lee, D.; Miller, A.S.; Janis, J.E.; Lee, B.T. Improving Readability and Automating Content Analysis of Plastic Surgery Webpages with ChatGPT. *J. Surg. Res.* **2024**, *299*, 103–111. [CrossRef]
71. Zakirah, A.Z.; Saniasiaya, J. Parental Perception on Usage of AI Chatbot to Understand Paediatric Otorhinolaryngology Condition: A Survey. *Indian J. Otolaryngol. Head Neck Surg. Off. Publ. Assoc. Otolaryngol. India* **2025**, *77*, 2078–2087. [CrossRef]

72. Chavda, H.; Sontam, T.R.; Skinner, W.C.; Ingall, E.M.; Zide, J.R. Comparison of Responses from ChatGPT-4, Google Gemini, and Google Search to Common Patient Questions About Ankle Sprains: A Readability Analysis. *J. Am. Acad. Orthop. Surg.* **2025**, *33*, 924–930. [CrossRef] [PubMed]
73. Yun, H.S.; Bickmore, T. Online Health Information-Seeking in the Era of Large Language Models: Cross-Sectional Web-Based Survey Study. *J. Med. Internet Res.* **2025**, *27*, e68560. [CrossRef] [PubMed]
74. Li, J.; Yang, Y.; Chen, R.; Zheng, D.; Pang, P.C.-I.; Lam, C.K.; Wong, D.; Wang, Y. Identifying healthcare needs with patient experience reviews using ChatGPT. *PLoS ONE* **2025**, *20*, e0313442. [CrossRef]
75. Li, J.; Yang, Y.; Mao, C.; Pang, P.C.I.; Zhu, Q.; Xu, D.; Wang, Y. Revealing patient dissatisfaction with health care resource allocation in multiple dimensions using large language models and the International Classification of Diseases 11th Revision: Aspect-based sentiment analysis. *J. Med. Internet Res.* **2025**, *27*, e66344. [CrossRef]
76. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [CrossRef] [PubMed]