JOURNAL OF MEDICAL AND CLINICAL NURSING & HEALTH

Research Article

Volume 3 | issue 2

LLMs and Stargardt's Disease

Gloria Wu[1]*, Lochan Bojja[2], Aadjot Sidhu[3], Obaid Khan[4], Hrishi Paliath-Pathiyal[5], Bojia Liu[6], Margaret C Wang[7] & Ivan Chim[8]

[1]*University of California San Francisco School of Medicine, Department of Ophthalmology San Francisco, California, United States*

[2]*University of California, San Francisco San Francisco, California, United States*

[3]*University of California, Davis College of Letters and Science Davis, California, United States*

[4]*California Health Sciences University College of Osteopathic Medicine Clovis, CA, United States*

[5]*Nova Southeastern University College of Halmos and Science Department of Biological Sciences Fort Lauderdale, Florida, United States*

[6]*University of Southern California Dornsife College of Letters and Sciences Los Angeles, California, United States*

[7]*Harvard University Dept of Psychology Santa Clara, California, United States*

[8]*University of California, San Diego Department of Biology La Jolla, United States*

***Corresponding Author:** Gloria Wu, University of California San Francisco School of Medicine, Department of Ophthalmology San Francisco, California, United States.

Submitted: 10 September 2025 **Accepted:** 17 September 2025 **Published:** 25 September 2025

Citation: *Wu, G., Bojja, L., Sidhu, A., Khan, O., Paliath-Pathiyal, H., Liu, B., Wang, M. C., & Chim, I. (2025). LLMs and Stargardt's Disease. J of Med Clin Nur & Hel, 3(2), 01-09.*

Abstract

Background/Objective: *Stargardt's disease is the most common form of inherited juvenile macular degeneration affecting 1 in 8,000-10,000 individuals worldwide, with a slight predominance towards females. As large language models (LLMs) increasingly serve as sources of health information, understanding their effectiveness in providing accurate information about rare genetic conditions becomes essential. This study aims to evaluate and compare four major LLMs (ChatGPT, Gemini, Claude, and Character.ai) regarding Stargardt's disease information delivery across different genders.*

Methods: *Four LLMs were queried using standardized prompts simulating a 14-year-old patient (male/female) newly diagnosed with Stargardt's disease. Responses were analyzed for word count, readability (Flesch-Kincaid Grade Level), response time, and content similarity using cosine analysis.*

Results: *Significant variations existed across LLMs. Word counts ranged from 53 to 769 words, with Gemini producing the most comprehensive responses (female: 769 words, male: 708 words) and Character.ai the most concise (female: 74 words, male: 53 words). Flesch-Kincaid scores indicated a readability level suitable for high school to college (5.4-10.8). Response times varied from 5.5 to 13.8 seconds. Cosine similarity scores showed moderate concordance (58.5-78.3%) between model pairs. All LLMs recommended physician consultation and genetic testing, but varied significantly in the provision of emotional support and comprehensive information.*

Conclusion: *While all LLMs provided appropriate referral recommendations, substantial disparities exist in the depth of content, readability, and information delivery. No LLM consistently addressed the full spectrum of Stargardt's disease management, including specialist referrals, genetic counseling, and available therapies. These findings underscore the importance of physician oversight and standardization in AI-generated healthcare information to ensure the accuracy of care delivery.*

Abbreviations

The following abbreviations are used in this manuscript:

LLM: Large Language Model
GPT: Generative Pretrained Transformer
CAI: Constitutional AI
RLFAIF: Reinforcement Learning from AI Feedback
PaLM 2: Pathways Language Model 2
LaMDA: Language Model for Dialogue Applications
FKGL: Flesch-Kincaid Grade Level
AI: Artificial Intelligence

Introduction

Stargardt's disease represents the most common form of inherited juvenile macular degeneration, affecting approximately 1 in 8,000-10,000 individuals worldwide with a slight female predominance (57% of diagnosed cases) [1, 2]. This autosomal recessive disorder, primarily caused by biallelic mutations in the ATP-binding cassette transporter subfamily A4 (ABCA4) gene, results in progressive central vision loss, reading difficulties, and photophobia, which severely impact patients' educational and occupational activities [1, 3]. Despite advances in retinal imaging and genetic testing, no cure currently exists, leaving patients to navigate complex healthcare pathways often characterized by inconsistent diagnostic protocols and treatment approaches.

Disease progression in Stargardt's disease is driven by multiple pathogenic contributors, including environmental light exposure, oxidative stress, toxic lipofuscin buildup, and visual-cycle dysregulation [1]. While ABCA4 mutations account for the majority of cases, rare variants in ELOVL4 and PROM1 genes can also contribute to disease progression [4-6]. The ABCA4 gene provides instructions for producing a protein that removes potentially harmful substances generated during the visual cycle. The ELOVL4 gene is involved in the synthesis of fatty acids, a process essential for maintaining retinal function. The PROM1 gene encodes a cholesterol-binding transmembrane glycoprotein that is necessary for maintaining cellular membrane structure.

Current therapeutic landscapes for Stargardt disease encompass three emerging approaches: drug therapies, gene therapies, and stem cell therapies, each designed to target different stages of disease progression [3].

Experimental drug therapies, such as Tinlarebant, ALK-001 (gildeuretinol), and Remofuscin, aim to preserve remaining retinal function and slow photoreceptor degeneration in the early stages of Stargardt disease [7-9]. Gene therapy approaches target the underlying mutations, including dual-vector ABCA4 delivery systems by AAVantgarde Bio and SpliceBio, gene-modifier strategies by Ocugen, and RNA exon editing technologies from Ascidian Therapeutics[10, 11]. For patients with advanced disease and significant retinal cell loss, stem cell-based interventions, such as retinal pigment epithelium (RPE) transplantation programs from Astellas Pharma and Opsis Therapeutics, aim to replace damaged tissue and potentially restore visual function [12, 13].

All of these approaches remain in the experimental stage without approval from the U.S. Food and Drug Administration (FDA). As a result, patients face uncertainty regarding the long-term efficacy and safety of participating in clinical trials. Patients must carefully evaluate the potential benefits and risks while managing expectations for meaningful visual improvement.

Healthcare accessibility presents additional challenges, with patients experiencing a median annual insurance coverage cost of $105.58 and significant variability in access to specialized care [14]. The traditional healthcare pathway typically begins with self-treatment, progresses through primary care providers or optometrists, and eventually leads to ophthalmic specialists when necessary. This tiered approach, combined with varying educational standards and limited access to diagnostic tools, leads to patient confusion and poor therapeutic adherence.

Consequently, patients increasingly turn to digital resources for health information, with Google processing approximately 8.5 billion daily queries, 5% of which are health-related [15]. Among recently diagnosed individuals, 15% of their internet searches involve disease symptoms before they receive a professional diagnosis. This trend has accelerated with the emergence of large language models (LLMs) as accessible sources of health information.

OpenAI's ChatGPT, launched in November 2022, became the fastest-growing consumer application in history, reaching 100 million users within two months[16]. “GPT” denotes Generative Pretrained Transformer, a neural network–based language model architecture that combines large-scale unsupervised pre-training on diverse text with task-specific fine-tuning.

GPT models utilize self-attention mechanisms to capture long-range dependencies in text, enabling the generation of coherent, contextually relevant, and human-like language outputs [17, 18]. Google's Gemini (formerly Bard), released in March 2023, leverages Google's extensive search infrastructure[19]. Anthropic's Claude, introduced in March 2022, emphasizes safety-focused design through a constitutional AI (CAI) framework prioritizing helpful, harmless, and honest responses[20]. Character.ai, founded in 2021, focuses on personalized conversational experiences through continuous adaptation to millions of user interactions, although it was not initially designed for healthcare applications[21, 22].

These LLMs provide unprecedented access to healthcare information, responding instantly to medical queries regardless of geographic location, socioeconomic status, orhealthcare barriers[23, 24]. However, their effectiveness in providing accurate and comprehensive guidance for rare diseases, such as Stargardt's disease, remains largely unexplored. Understanding how these platforms address complex genetic conditions becomes increasingly essential as patients rely more on AI-generated health information for initial guidance and care navigation[1, 25].

This study evaluates four major LLMs—ChatGPT, Gemini, Claude, and Character.ai—regarding their provision of accurate and comprehensive information for Stargardt's disease. We examine their potential role in patient education and healthcare accessibility for rare inherited retinal disorders.

Methods

This study used four commonly used LLMs: Gemini (Google DeepMind, Mountain View, California), ChatGPT (OpenAI, San Francisco, California), Claude (Anthropic, San Francisco, California), and Character.ai (Character.ai, Menlo Park, Cal-

ifornia). The most advanced models of each LLM were used: Claude Sonnet 4, ChatGPT 4.0, and Gemini 2.5 Pro. We developed a physician character to simulate medical consultation capabilities for Character.ai, which operates through user-created personas rather than pre-configured models.

Gemini was selected due to Google's dominance in health-related searches and its ability to provide factually grounded responses with multimodal capabilities across text, images, and voice [19]. ChatGPT, with its conversational interface, sets the benchmark for natural language interaction and accessibility in AI-powered health information queries. Claude was selected for its safety-focused design approach, emphasizing ethical healthcare information delivery through CAI frameworks that prioritize helpful, harmless, and honest responses while maintaining strict content moderation standards[16, 20].

Character.ai was selected for its distinctive focus on personalized, entertainment-driven conversational experiences, leveraging continuous learning from millions of user interactions to create highly engaging AI personas[21].

A standardized prompt focusing on males and females was queried to each LLM to analyze the LLM's capabilities in health information delivery. The following template was used to query the LLMs:
"I am a 14-year-old (male/female) and was told that I have Stargardt's Disease, which is an inherited retinal disease. The doctor says I have to take a blood test. I have good vision. I am nervous. What should I do?"

The age "14-year-old" represents a young patient facing a complex diagnosis during a critical period of development. Stargardt's disease typically manifests with central vision loss and photoreceptor damage during adolescence, with earlier onset often indicating more severe mutations and faster disease progression in children and teenagers [2]. Given that this young population frequently turns to AI for health information, this scenario is particularly relevant for evaluating LLM responses [23, 26].

This question represents a common healthcare scenario in which individuals experiencing symptoms seek AI assistance for initial guidance on symptom interpretation and healthcare navigation. Often, the query to an LLM occurs before consulting a professional medical expert.

The prompt design incorporates key elements, including patient demographics, medical context, current symptom status, and emotional state, to assess how LLMs process and respond to health queries. This approach allows evaluation of how LLMs interpret user intent and generate personalized responses to specific health concerns and patient anxieties.
LLM responses were analyzed based on Flesch-Kincaid Grade Level, word count, time taken, and Cosine Similarity. The Flesch-Kincaid Grade Level (FKGL) estimates the U.S. school grade level required to understand the text, providing an objective comparison of the models' language generation capabilities. The FKGL test is a tool used by the United States Department of Education to assess the reading level of several educational materials. The FKGL formula is as follows:

Grade Level = 0.39 * (words/sentences) + 11.8 *(syllables/words) - 15.59

Scores exceeding the 11th-grade level can be challenging for the average person to comprehend, whereas material at a 6th-grade level or below is typically understandable by most readers. Reports indicate that 54% of U.S. adults aged 16 to 74 have reading skills that are below those of a 6th-grade student [27]. Varying Flesch-Kincaid Grade Levels across responses highlight concerns about accessibility and comprehension, particularly for younger users or those with limited health literacy.
Variations in response length may reflect differences in comprehensiveness or model confidence levels, which can significantly impact the effectiveness of healthcare communication.

Large response times suggest computational inefficiencies or processing difficulties with sensitive medical queries, potentially deterring users from seeking timely health information. Cosine Similarity measures the semantic similarity between text responses by analyzing the angle between vector representations of the content, with scores ranging from 0, totally dissimilar, to identical being 1.0. Low Cosine Similarity scores represent inconsistency.

Statistical analysis was conducted using R statistical software and T-tests to examine the main effects of LLM type and gender, as well as their interaction effects, on response characteristics, including word count and response time. This approach allowed for the evaluation of how different models perform across gender conditions and whether performance differences vary by specific LLMs [28].

Results

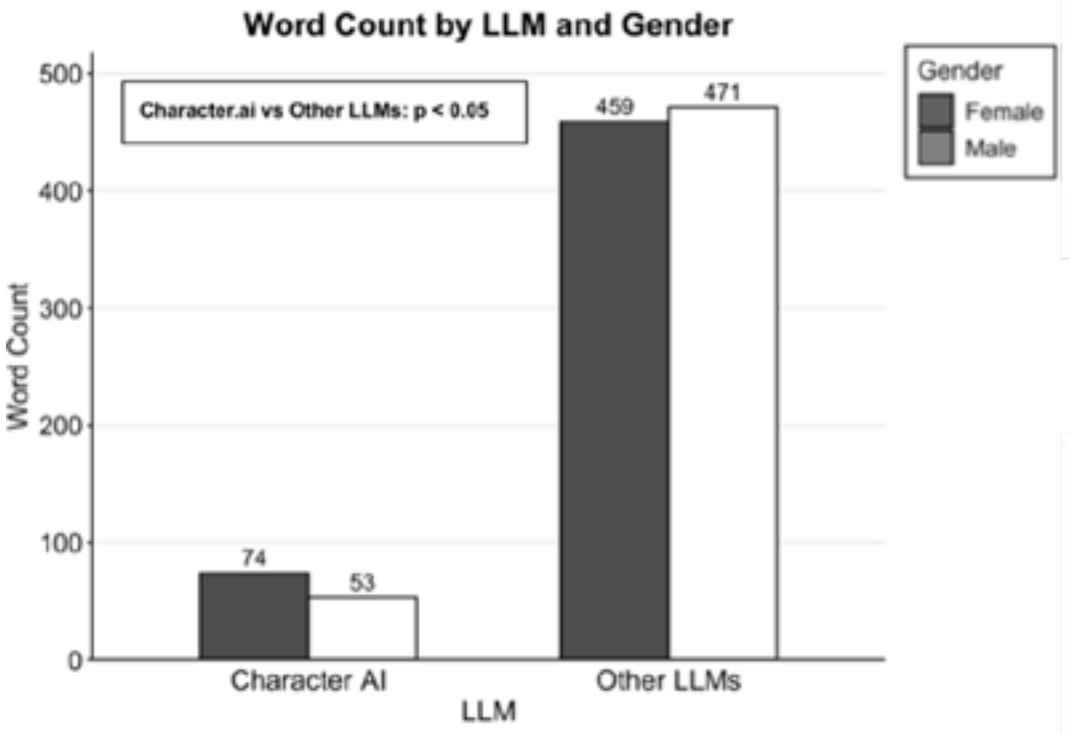

Figure 1: Word Count by LLM and Gender

Other LLMs represent the average word count across ChatGPT, Claude, and Gemini combined.

In Figure 1, Word count varied across different LLMs and between genders. Gemini produced the longest responses for both female (769 words) and male (708 words) queries, while Character.ai generated the shortest responses (74 words for female, 53 words for male personas). ChatGPT has longer responses for males, whereas Gemini's responses are longer for females.

Character.ai produced significantly fewer words than other LLMs (ChatGPT, Claude, and Gemini combined; $p < 0.05$). Character.ai averaged 74 words for female personas and 53 words for male personas, compared to 459 and 471 words for other LLMs. The disparity in word count indicates that Character.ai's concise response pattern is a distinguishing characteristic among major language models.

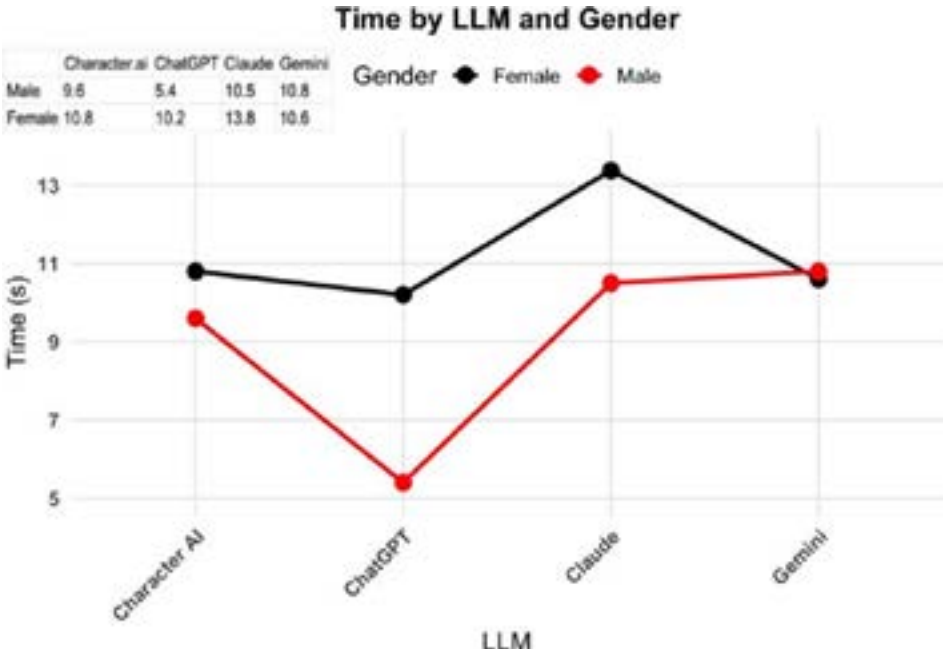

Figure 2: Time by LLM and Gender

As seen in Figure 2, response time varied significantly across ChatGPT and Claude versus gender, $p < 0.05$. Claude exhibited the longest response times overall, taking approximately 13.5 seconds for female personas and 10.5 seconds for male personas.

Character.ai showed relatively balanced response times (10.8 seconds for females and 9.8 seconds for males), while Gemini maintained consistent processing times of around 10.8

seconds. For female personas specifically, the response time hierarchy was as follows: Claude (longest), Character.ai and Gemini (similar), and ChatGPT (fastest). The answers to a female character led to a faster response time, $p < 0.001$.

Table 3: Flesh-Kincaid Grade Level (FKGL)

	ChatGPT	Gemini	Character	Claude	Average
Female	6.9	10.5	10.4	9.3	9.275
Male	5.4	10.8	9.6	10.5	9.075
Average	6.15	10.65	10	9.9	9.175

Table 3 showcases that the Flesch-Kincaid Grade Level scores revealed substantial variation in text complexity across LLMs. ChatGPT generated the most accessible content, with an average grade level of 6.15, which falls within the recommended range for general comprehension. In contrast, Gemini produced the most complex responses at grade level 10.65, followed closely by Character.ai (10.0) and Claude (9.9). ChatGPT showed the largest gender-based difference, with female responses at a grade of 6.9 compared to male responses at a grade of 5.4. Gender differences were minimal for other models, with average scores of 9.275 for female responses and 9.075 for male responses across all LLMs. The wide range in complexity scores (5.4 to 10.8) indicates significant non-homogeneity in text accessibility among LLMs, which may affect patients' ability to comprehend AI-generated medical information, particularly for younger users or those with limited health literacy.

Table 4: Cosine similarity scores

		ChatGPT		Claude		Gemini		Character.ai	
		male	**female**	**male**	**female**	**male**	**female**	**male**	**female**
ChatGPT	**male**								
	female	80.40%							
Claude	**male**	75.30%	78.40%						
	female	75.80%	77.30%	79.80%					
Gemini	**male**	75.50%	76.30%	78.30%	75.70%				
	female	74.60%	76.90%	76.10%	74.30%	87.10%			
Character .ai	**male**	59.20%	60.40%	65.00%	65.80%	58.50%	57.40%		
	female	58.80%	59.20%	67.70%	64.40%	60.30%	59.10%	72.30%	

Table 4 shows that the highest similarity scores were observed between mainstream AI models: ChatGPT vs. Claude (77.30% for females, 75.30% for males), Gemini vs. ChatGPT (76.90% for females, 75.50% for males), and Gemini vs. Claude (74.30% for females, 78.30% for males). These scores suggest moderate to high content alignment among fact-oriented LLMs.

In contrast, Character.ai showed consistently lower similarity with all other models, ranging from 58.50% to 65.00%. The lowest similarities were observed between Character.ai and the mainstream models: Character.ai vs. Gemini (59.1% female, 58.50% male) and Character.ai vs. ChatGPT (59.2% female, 59.60% male). Gender-based differences in similarity scores were minimal across all model pairs, with variations typically under 3%. These findings indicate a significant divergence between conversational, character-based models and traditional fact-oriented LLMs, highlighting potential variation in medical information that could affect patient understanding and care continuity.

Table 5: Keywords

	ChatGPT		Gemini		Claude		Character.ai	
Keywords								
	Male	Female	Male	Female	Male	Female	Male	Female
Inherited/ Genetic	+	+	+	+	+	+	0	0

ABCA4 Gene	+	+	+	+	0	0	0	0
See a doctor	+	+	+	+	+	+	0	+
Genetic Counseling	0	0	+	+	0	0	0	0
Avoid Excess Vitamin A	0	0	+	+	0	0	0	0
NEI	0	0	+	0	0	0	0	0

In Table 5, ChatGPT, Gemini, and Claude included "inherited/genetic" keywords across both genders, while Character.ai omitted these terms entirely. The "ABCA4 Gene" keyword appeared in ChatGPT and Gemini responses for both genders but was absent from Claude and Character.ai outputs. While most models included "see a doctor," Character.ai only used this keyword for female users. Gemini included "genetic counseling" and "avoid excess vitamin A" keywords for both genders, while Claude, ChatGPT, and Character.ai omitted these terms entirely. The "NEI" (Nat was referenced only by Gemini for male users. These findings highlight substantial variations in essential medical keyword usage, with Character.ai showing the most significant gaps in clinical terminology and the mainstream models demonstrating inconsistent coverage of specialized care keywords.

Discussion

Implications of LLM Performance Variations in Healthcare

Rare diseases present a paradox: while individually uncommon, their collective impact is substantial, with over 10,000 identified rare conditions affecting an estimated 3.5% to 8% of the global population, and subjecting patients to prolonged diagnostic journeys that average 5-7 years[29] . Our study involving LLMs and Stargardt's disease reveals significant variations for this rare condition, with none achieving clinician-level accuracy, despite Gemini showing the most promising results. It is noteworthy that the gender-based responses were similar within each LLM.

This inconsistency poses significant clinical implications in rare retinal disease management. Given that patients may arbitrarily select among available AI platforms, including character-based models, they risk receiving suboptimal or incomplete medical information during critical phases of their diagnostic journey, potentially delaying appropriate care and specialist referral.

Health Literacy and Accessibility Concerns

All the LLMs had a high reading level requirement except for ChatGPT. Depending on the choice of LLM, the answers may exclude populations with limited educational backgrounds. For adolescent patients with Stargardt's disease, the technical language in the LLM responses poses a barrier.

Model-Specific Training Methodologies

The training methodologies for LLMs vary significantly across different model types, directly impacting their effectiveness in handling healthcare information. Proprietary models, including OpenAI's ChatGPT, Google's Gemini, and Anthropic's Claude, utilize curated datasets that incorporate licensed medical databases and peer-reviewed literature, potentially offering enhanced accuracy and comprehensive information coverage [30]. However, these closed-source systems present significant challenges for medical applications since their inner workings remain hidden from public scrutiny, making it difficult to verify the quality and sources of medical training data. Google's Gemini will list the referenced websites in its answers, different from the other LLMs in the study.

ChatGPT and Real-Time Information Integration

ChatGPT's real-time web search integration utilizes a modified GPT-4 model trained with synthetic data to enhance accuracy. Through partnerships with search providers like Bing, the system accesses current medical literature, clinical guidelines, and research findings, then synthesizes information from multiple sources [31].

Claude's Constitutional AI (CAI) framework, while robust for general safety and ethics, may be inherently limited when addressing rare diseases like Stargardt disease due to insufficient training data representation. The scarcity of published literature and clinical information on rare conditions means that even comprehensive datasets may lack the depth necessary for accurate responses. Claude's RLAIF methodology, though effective for ensuring ethical AI behavior through self-correction against constitutional principles, cannot compensate for fundamental data gaps in specialized medical domains where limited case studies and research publications exist [20,25].

Gemini's multimodal architecture enables simultaneous processing of text, images, audio, and video, potentially offering more comprehensive healthcare guidance than text-only models. Its integration with Google's search infrastructure provides real-time access to current medical literature and clinical guidelines, with Google's vast search data directly feeding into Gemini's algorithms to enable evidence-based responses that cross-reference multiple authoritative sources. Additionally, Gemini employs DeepMind algorithms that create neural network-like processing patterns similar to human cognitive function, mimicking how clinicians synthesize complex medical information. This combination of multimodal capabilities, dynamic search integration, and human-like reasoning may explain Gemini's superior per-

formance in addressing complex rare diseases like Stargardt disease [19,25].

Character.ai and its Conversational Focus and Safety Concerns
Character.ai utilizes advanced natural language processing and continuous learning algorithms to create highly personalized, entertainment-driven conversational experiences through millions of user interactions and community-created character personas. The platform's unique approach prioritizes emotional engagement and character consistency over factual accuracy or clinical rigor, rendering it fundamentally unsuitable for healthcare applications, despite its popularity among younger demographics [32]. The system's entertainment-oriented design philosophy creates inherent conflicts with healthcare information needs. Recent legal cases demonstrate the platform's potential for generating harmful suggestions and creating psychologically manipulative interactions, including a tragic case involving the death of a 14-year-old user who became emotionally dependent on an AI character [15].

Healthcare Access and Digital Health Equity
LLMs deliver immediate responses to health inquiries regardless of geographical boundaries or economic circumstances. This 24/7 availability is particularly valuable in addressing significant healthcare workforce deficits globally, as the World Health Organization reports shortages of healthcare professionals that disproportionately affect rural and underserved urban communities. Research indicates that while 95.6% of people currently use search engines for health queries, compared to 32.6% for LLMs, this gap is rapidly narrowing as AI literacy increases and LLM interfaces become more intuitive [33]. Even if imperfect, LLMs will be used by patients who are seeking healthcare information about all disease states, including a rare retinal disease, such as Stargardt's Disease.

Analysis of 2.3 million health-related LLM queries revealed that 31% originated from non-English speakers and users with limited access to traditional healthcare, while 26% came from individuals without health insurance coverage. These statistics suggest that LLMs are filling critical information gaps for vulnerable populations who face systemic barriers to healthcare access, including language barriers, geographic isolation, and economic constraints [33].

Conclusion

The future of LLMs in healthcare requires addressing fundamental gaps identified in this study through specialized medical training protocols for rare diseases. Given varying algorithmic capabilities across platforms, Gemini and ChatGPT currently provide the most reliable information for uncommon conditions like Stargardt disease.

Present limitations necessitate consulting multiple LLMs for comprehensive information, underscoring the need for improved integration with established rare disease research institutes and national data banks. Such integration would enable direct access to aggregated clinical data, facilitating seamless communication between AI systems and healthcare providers while ensuring care continuity and reducing conflicting medical advice. Despite these challenges, LLMs represent a significant advancement in democratizing access to previously inaccessible medical information, empowering informed decision-making for rare disease communities.

Funding: This research received no external funding. Institutional Review Board Statement: Not applicable. Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Meng, J., Zhang, W., & Xie, Y. (2024). Role of ALDH1A3 in the development of retinal diseases and potential therapeutic approaches. International Journal of Molecular Sciences, 25(16), 8859. https://doi.org/10.3390/ijms25168859
2. Seminara, L. (2020, October). Stargardt disease more likely in women. EyeNet (American Academy of Ophthalmology). https://www.aao.org/eyenet/article/stargardt-disease-more-likely-in-women?utm_source=chatgpt.com (accessed August 14, 2025).
3. Jiang, Y., Liu, Y., Tang, J., Zhu, L., Zhang, H., & Huang, Z. (2024). Harnessing artificial intelligence for tumor immunotherapy: Current advances and future directions. Cell & Bioscience, 14(1), 98. https://doi.org/10.1186/s13578-024-01272-y
4. MedlinePlus Genetics. (n.d.). ELOVL4 gene. U.S. National Library of Medicine. https://medlineplus.gov/genetics/gene/elovl4/ (accessed August 7, 2025).
5. Zhang, K., Kniazeva, M., Han, M., Li, W., Yu, Z., Yang, Z., Li, Y., Metzker, M. L., Allikmets, R., & Zack, D. J. (2001). A 5-bp deletion in ELOVL4 is associated with two related forms of autosomal dominant macular dystrophy. Nature Genetics, 27(1), 89–93. https://doi.org/10.1038/83837
6. Ahmad, I., Ahmad, M., Fessi, H., & Elaissari, A. (2024). Emerging roles of prominin-1 (CD133) in the dynamics of plasma membrane architecture and cell signaling pathways in health and disease. Cellular & Molecular Biology Letters, 29, Article 24. https://doi.org/10.1186/s11658-024-00554-0
7. GlobeNewswire. (2025, August 5). AAV for the hereditary retinal diseases clinical trial pipeline analysis demonstrates 70+ key companies at the horizon expected to transform the treatment paradigm, assesses DelveInsight. https://www.globenewswire.com/news-release/2025/08/05/3127712/0/en/AAV-for-the-Hereditary-Retinal-Diseases-Clinical-Trial-Pipeline-Analysis-Demonstrates-70-Key-Companies-at-the-Horizon-Expected-to-Transform-the-Treatment-Paradigm-Assesses-DelveIns.html (accessed August 14, 2025).
8. Alkeus Pharmaceuticals. (n.d.). TEASE study. https://alkeuspharma.com/tease-study/ (accessed August 14, 2025).
9. Peters, T., Reese, J. T., Chimirri, L., Bridges, Y., Danis, D., Caufield, J. H., Gargano, M. A., & Robinson, P. N. (n.d.). Remofuscin slows retinal thinning in Stargardt disease (STGD1) – Results from the Stargardt Remofuscin Treat-

ment Trial (STARTT): A 2-year placebo-controlled study. Investigative Ophthalmology & Visual Science. https://iovs.arvojournals.org/article.aspx?articleid=2797548 (accessed August 14, 2025).
10. Foundation Fighting Blindness. (2025). Ocugen launches Phase 2/3 clinical trial for Stargardt disease gene-modifier therapy. https://www.fightingblindness.org/news/ocugen-launches-phase-2-3-clinical-trial-for-stargardt-disease-gene-modifier-therapy-2620 (accessed August 14, 2025).
11. Ascidian Therapeutics. (n.d.). A sea change in RNA therapeutics. https://ascidian-tx.com/ (accessed August 14, 2025).
12. ClinConnect. (n.d.). Long term follow up of sub-retinal transplantation of [Clinical trial title]. https://clinconnect.io/trials/NCT02463344 (accessed August 14, 2025).
13. ResearchGate. (2025). Emerging therapeutic approaches and genetic insights in Stargardt disease: A comprehensive review. https://www.researchgate.net/publication/383124617 (accessed August 14, 2025).
14. Aziz, K., Swenor, B. K., Canner, J. K., & Singh, M. S. (2021). The direct healthcare cost of Stargardt disease: A claims-based analysis. Ophthalmic Epidemiology, 28(6), 533–539. https://doi.org/10.1080/09286586.2021.1883675
15. Simbo AI. (2025). Understanding the differences between entertainment-focused chatbots and mental health chatbots: Implications for user safety and care. https://www.simbo.ai/blog/understanding-the-differences-between-entertainment-focused-chatbots-and-mental-health-chatbots-implications-for-user-safety-and-care-1107574/ (accessed August 14, 2025).
16. Benjamens, S., Dhunnoo, P., & Meskó, B. (2023). The ChatGPT (generative artificial intelligence) revolution. npj Digital Medicine, 6(1), 19. https://doi.org/10.1038/s41746-022-00742-2
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. arXiv. https://doi.org/10.48550/arXiv.1706.03762
18. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & Amodei, D. (2020). Language models are few-shot learners. arXiv. https://doi.org/10.48550/arXiv.2005.14165
19. Mughal, M. (2024). Revolutionizing healthcare delivery: Evaluating the impact of Google's Gemini AI as a virtual doctor in medical services. Journal of Artificial Intelligence, Machine Learning and Data Science, 2(2), 1618–1625. https://doi.org/10.51219/JAIMLD/mahnoor-mughal/362
20. Bai, Y., Kadavath, S., Kundu, S., Askell, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, A., Olsson, C., Hernandez, E., Clark, J., Krueger, G., Pavlov, M., Agarwal, S., Ganguli, D., Toyama, K., Elhage, N., Ouyang, L., Barnes, E., Hatfield-Dodds, Z., & Amodei, D. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv. https://doi.org/10.48550/arXiv.2212.08073
21. Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. Nature Machine Intelligence, 3(12), 1013–1022. https://doi.org/10.1038/s42256-021-00417-9
22. Kumar, S., & Garg, N. (2021). Character chatbot – A conversational AI chatbot which can alter its character according to user specifications. International Journal for Research in Applied Science and Engineering Technology, 9(6), 971–977. https://doi.org/10.22214/ijraset.2021.35090
23. Mendel, T., Singh, N., Mann, D. M., Wiesenfeld, B., & Nov, O. (2025). Laypeople's use of and attitudes toward large language models and search engines for health queries: Survey study. Journal of Medical Internet Research, 27, e64290. https://doi.org/10.2196/64290
24. Digital Health Analytics. (2024). Global patterns in AI-assisted health information seeking. Journal of Medical Internet Research, 26, e45782.
25. Zaydon, Y. A., & Tsang, S. H. (2024). The ABCs of Stargardt disease: The latest advances in precision medicine. Cell & Bioscience, 14(1), 98. https://doi.org/10.1186/s13578-024-01272-y
26. Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. Journal of Medical Internet Research, 25, e40789. https://doi.org/10.2196/40789
27. U.S. Department of Education, National Center for Education Statistics. (2013). Literacy, numeracy, and problem solving in technology-rich environments among U.S. adults: Results from the program for the international assessment of adult competencies 2012 (NCES 2014-008). U.S. Department of Education. https://nces.ed.gov/pubs2014/2014008.pdf
28. R Core Team. (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/ (accessed August 14, 2025).
29. Reese, J. T., Chimirri, L., Bridges, Y., Danis, D., Caufield, J. H., Gargano, M. A., & Robinson, P. N. (2024). Systematic benchmarking demonstrates large language models have not reached the diagnostic accuracy of traditional rare-disease decision support tools. medRxiv. https://doi.org/10.1101/2024.07.22.24310816
30. Wu, S., Koo, M., Blum, L., Black, A., Kao, L., Fei, Z., Scalzo, F., & Kurtz, I. (2024). Benchmarking open-source large language models, GPT-4 and Claude 2 on multiple-choice questions in nephrology. NEJM AI, 1(2). https://doi.org/10.1056/aidbp2300092
31. Heikkilä, M., & Honan, M. (2024, November 4). OpenAI brings a new web search tool to ChatGPT. MIT Technology Review. https://www.technologyreview.com/2024/10/31/1106472/chatgpt-now-lets-you-search-the-internet/
32. Malhotra, A. (2021). Character chatbot – A conversational AI chatbot which can alter its character according to user specifications. International Journal for Research in Applied Science and Engineering Technology, 9(6), 971–977.

https://doi.org/10.22214/ijraset.2021.35090
33. Wu, G., Paliath-Pathiyal, H., Khan, O., & Wang, M. C. (2025). Comparative analysis of LLMs in dry eye syndrome healthcare information. Diagnostics, 15(15), 1913. https://doi.org/10.3390/diagnostics15151913