Exploring Caregivers' Acceptance of Conversational AI in Pediatric Cancer Caregiving: A Mixed-Methods Study

Mansi Sugan
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
msugan@iu.edu

Salonee Nadkarni
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
sanadka@iu.edu

Samaneh Sanaeipoor
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
sasana@iu.edu

Shreyas Patel
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
sp101@iu.edu

Sharmitha Yazhini Natarajamoorthy
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
shanata@iu.edu

Elmira Rashidi
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
erashidi@iu.edu

Andrew D Miller
Human-Centered Computing, Luddy School of Informatics, Computing, and Engineering
Indiana University Indianapolis
Indianapolis, Indiana, USA
andrewm@iu.edu

Abstract

Caring for a child with cancer involves navigating through complex medical information, often delivered through lengthy handbooks and consultations with healthcare providers. Overnight, parents are expected to become an expert on a domain which they knew nothing about. Conversational UIs, powered by Large Language Models (LLMs) and validated information sources, could play a key role in supporting caregivers. In this paper, we investigate the usability, acceptance, and perceived utility of an LLM-based conversational AI tool for pediatric cancer caregiving, grounded in the Children's Oncology Group Family Handbook–the leading resource in pediatric oncology care. We employed a mixed-methods approach, interviewing and surveying 12 caregivers as they engaged with a functional prototype. We offer insights into caregiver's needs and expectations from AI-driven tools, and design guidelines for developing safer, more personalized, and supportive AI interventions for pediatric cancer care.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3721270

CCS Concepts

• **Human-centered computing → Empirical studies in collaborative and social computing**.

ACM Reference Format:
Mansi Sugan, Salonee Nadkarni, Samaneh Sanaeipoor, Shreyas Patel, Sharmitha Yazhini Natarajamoorthy, Elmira Rashidi, and Andrew D Miller. 2025. Exploring Caregivers' Acceptance of Conversational AI in Pediatric Cancer Caregiving: A Mixed-Methods Study. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3721270

1 Introduction

The diagnosis of pediatric cancer profoundly alters the lives of families, thrusting parents into roles as caregivers tasked with managing complex medical jargon, making critical decisions, and providing emotional support. To support these overwhelmed caregivers, the Children's Oncology Group (COG) developed expert consensus recommendations emphasizing standardized yet accessible information delivery[28]. The COG family handbook implements these recommendations by providing carefully structured information for parents of children diagnosed with cancer, balancing the need for consistent content delivery while accommodating different parental learning styles when sharing details about treatment options, support services, and follow-up care[12]. While comprehensive, the COG Family Handbook and similar dense resources can overwhelm caregivers who are already dealing with significant emotional strain. Furthermore, caregivers may have questions and seek education beyond the hospital, using their own devices and on their own terms.

LLM powered conversational AI, presents a significant opportunity to address caregiver challenges by seamlessly integrating with healthcare systems and evolving alongside caregiver's needs. Unlike static resources, conversational AI tools can dynamically adapt to the shifting priorities of caregivers, from providing precise medical guidance during the initial diagnosis and treatment phases to offering long-term support for emotional well-being and follow-up care. These tools can transform the caregiving experience by delivering timely, tailored, and easily digestible information, reducing the burden of information overload, and enhancing caregiver's ability to support their child effectively. Advancements in conversational AI, particularly those driven by LLMs, offer a new way to deliver timely, contextualized information to caregivers. Retrieval-Augmented Generation (RAG) frameworks, for instance, ensure that conversational agents provide responses grounded in reliable, domain-specific sources, mitigating common issues like hallucination and misinformation[38].

In this paper we ask: **How do caregivers of children with cancer perceive the usability, acceptability, and effectiveness of a conversational AI support tool in meeting their caregiving needs?** To answer this question, we developed a prototype conversational agent, and asked caregivers of children with cancer to interact with it during a live interview. Through a mixed-methods approach that combines surveys and in-depth interviews, we examine how conversational AI can adapt to a caregiver's evolving needs and enhance their caregiving experience across all phases of the cancer journey.

2 Related Work

2.1 Large Language Models in Health and HCI

Recent advances in large language models (LLMs) have positioned them as revolutionary technologies impacting multiple sectors [36], with their potential applications in healthcare standing out as especially significant [15, 27]. While various promising applications and potential use cases have been identified [14, 45], including clinical workflow [11, 17, 18, 37, 42]and health management [2, 9, 24], the rapid adoption of these technologies, particularly conversational AI tools in patient communication and education [29, 41, 47, 51], has begun to reshape various aspects of healthcare delivery and management. In particular, LLMs can be organized into interactive chains to facilitate complex patient interactions [1], where conversational agents equipped with advanced LLM functionalities are increasingly being deployed to enhance patient interactions, enabling more effective information gathering through context-aware questioning and providing timely responses to medical inquiries, ultimately contributing to improved patient outcomes and healthcare accessibility[47]. Despite GPT's demonstrated capabilities in medical contexts, its fundamental limitations in clinical reasoning, the tendency to provide incorrect care recommendations, and uncertain reliability due to limited medical training data have been highlighted[6, 13]. Likewise, the integration challenges identified by Jiang (2024), particularly around trustworthiness, interpretability, and the potential for false outputs in healthcare LLMs, underscore the critical importance of understanding how patients interact with and develop trust in these AI systems [22]. While general-purpose LLMs have shown promise as information-seeking tools for pediatric cancer, these approaches lack personalization and verified data sources [40]. As a result, researchers have developed the Retrieval Augmented Generation (RAG) method, which enhances LLM accuracy by combining generative capabilities with precise information retrieval from verified databases[33]. As demonstrated in studies by Kamalloo et al. (2024), this approach strengthens Question-Answering (QA) systems by grounding responses in authoritative sources rather than relying solely on pretrained knowledge, effectively reducing hallucinations and improving factual precision [25], particularly crucial in healthcare applications where accuracy of information is vital [23].

Within the field of Human-Computer Interaction (HCI), researchers actively explore how specific conversational capabilities, such as natural language understanding, contextual memory, and adaptive response generation, enable AI-driven agents to foster trust and engagement through personalized interactions and consistent dialogue patterns [4]. Indeed, the empirical methods of HCI research create a synergistic relationship where HCI principles guide the development of more intuitive and user-friendly LLM interfaces, while LLMs enhance HCI by enabling more natural and adaptive human-computer dialogues. Empirical evidence demonstrates that innovations in adaptive interface design and natural language processing have led to measurable improvements in both user engagement metrics and contextual appropriateness of system responses [8, 44], ultimately enhancing human-AI interaction. Likewise, Xu et al. (2024) reveal that LLM-based conversational agents not only demonstrate higher user satisfaction and lower cognitive load compared to traditional messaging apps with human consultants but also facilitate more nuanced interactions by understanding subtle contextual cues, ultimately leading to a more seamless and engaging user experience [50]. Calisto et al. showed that assertiveness may play a role in user acceptance for agent-based communications [7]. Moreover, studies consistently demonstrate that user acceptance and engagement significantly improve when systems provide personalized responses tailored to individual user characteristics and preferences, with personalized interfaces showing a marked positive impact on both user satisfaction metrics and long-term system adoption rates[5, 46, 52]. The recent work by Luetke Lanfer et al. further explores this dynamic, specifically examining how digital clinical empathy can be effectively implemented in healthcare chatbots, emphasizing the need for balanced, authentic, and timely responses while maintaining professional standards [31]. However, the relationship between user expectations and actual system capabilities remains a critical area of focus, as significant disparities between user expectations and system performance, particularly regarding machine intelligence and capability [32] - a gap in understanding that has important implications for system design and implementation, especially in healthcare settings where trust and reliability are essential.

2.2 Cancer Care and LLMs: Challenges and Current Limitations

LLMs have shown potential in cancer care[10, 16, 26]. However, significant challenges persist in applying LLMs to cancer care, particularly in specialized domains like pediatric oncology [40]. Current implementations face notable limitations in domain-specific oncological knowledge and require substantial computational resources, which hinders their practical deployment in clinical settings[30]. The current research domain shows insufficient studies specifically focused on LLMs in oncology, with an even greater gap in childhood cancer applications where treatment protocols are distinctly different from adult care[30, 48]. Moreover, technical hurdles persist in seamlessly incorporating these systems into existing healthcare infrastructure, including issues with interoperability, data standardization, real-time decision support capabilities, data obsolescence[20]in rapidly advancing fields like cancer research, lack of access to training data for validation [3], significant error rates in medical recommendations, and the need for continuous monitoring of model behavior changes over time[21]. Indeed, the intricate nature of cancer care requires LLMs to process and interpret vast amounts of medical literature, clinical trials, and patient data while maintaining accuracy and reliability in their recommendations, especially when dealing with the nuanced requirements of pediatric cases.

In HCI, the potential of LLMs for cancer caregiving has not yet been sufficiently explored explored interactive tools for pediatric oncology caregivers, demonstrating technology's potential to address current challenges in information delivery and emotional support provision[35, 39, 43]. Their work highlighted how caregivers often struggle with accessing and processing complex medical information at different treatment stages, necessitating targeted assessment of information needs. This aligns with our findings on conversational AI's multifaceted role in supporting caregivers, particularly in providing phase-appropriate emotional and informational support. In the pediatric oncology context, Hile et al. (2014) demonstrated that parental stress significantly predicts functional outcomes in pediatric cancer survivors [19], emphasizing the critical need for comprehensive caregiver support systems that address both informational and emotional requirements throughout the care trajectory.

3 Methods

In this study, we followed a mixed-methods approach, incorporating both qualitative interviews and surveys to gather insights into caregiver experiences and their interactions with the conversational AI support agent. We interviewed a total of 12 caregivers of children with cancer using a functional prototype of the conversational agent designed to surface content drawn from the COG Family Handbook. Study sessions took the form of an interview, during which participants interacted with the prototype. Participants began by sharing their experiences with their child's diagnosis and treatment, as well as their familiarity with the COG Family Handbook. Interviews were conducted remotely using platforms like Zoom and Microsoft Teams. The moderator facilitated each session, encouraging participants to think aloud as they interacted with the AI tool. Our study was approved by the Indiana University Institutional Review Board.

In the interviews, participants interacted with the AI tool by asking questions relevant to different stages of the caregiving journey—diagnosis, hospitalization, and post-hospitalization—to explore and address the varying needs of caregivers across cancer journey stages [34]. Before using the prototype, participants completed a pre-interview survey that evaluated caregiver's perception of AI tools, specifically regarding their potential to provide informational support during caregiving. The survey items assessed factors such as perceived usefulness, trust, and ease of use, helping to gauge caregiver's readiness for adopting such tools and their initial attitudes toward AI in healthcare.

The post-interview survey evaluated participant satisfaction with the AI agent and its usefulness in their caregiving context. Drawing from factors identified in Wutz et al.'s (2023) comprehensive thematic framework[49], participants answered the survey with 16 Likert-scale statements evaluating the tool's key acceptance factors, including usability, trustworthiness, emotional support, perceived empathy, reliability, and relevance to caregiving tasks. They were also asked about the tool's potential impact on their caregiving routine and how it could be improved. Questions addressed the AI's perceived empathy, reliability, and relevance to caregiving tasks.

We analyzed the findings through both qualitative approaches and quantitative metrics. The interview questions were developed based on key acceptance factors including ease of use, trust, emotional impact, and social influence[49]. The qualitative data from the interviews were transcribed and analyzed thematically. This analysis focused on identifying recurring patterns in caregiver's feedback about the conversational AI tool and understanding their perspectives on its utility at different stages of caregiving. For the quantitative data, the responses from the pre and post-interview surveys were analyzed using descriptive statistics.

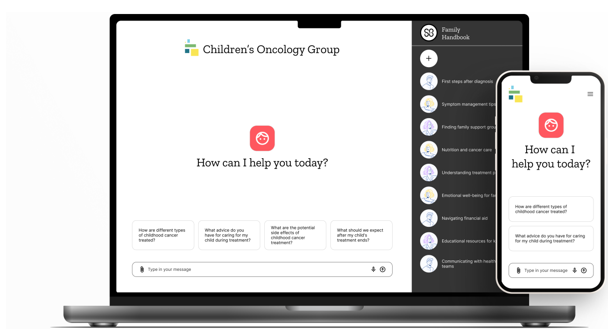

Figure 1: The conversational agent

3.1 System Design

We designed an LLM-supported conversational agent integrated with the COG Family Handbook. The agent uses GPT4.0 soft-prompted with training instances within system instructions. The

Table 1: Demographics of study participants

ID	Age	Race/Ethnicity	Relationship to Child	Child Age	Diagnosis
P1	40	White	Father	1	Pleuropulmonary blastoma
P2	47	White	Mother	4	ALL
P3	36	African American	Mother	1	Neuroblastoma
P4	54	African American	Mother	13	Non-Hodgkin lymphoma
P5	45	African American	Mother	15	Hodgkin lymphoma
P6	19	African American	Legal Guardian	1	Central Nervous System
P7	45	White	Father	3	ALL
P8	51	White	Father	8	ALL
P9	37	Spanish	Mother	7	ALL
P10	33	African American	Mother	8	Other
P11	45	White	Mother	18	Non-Hodgkin lymphoma
P12	50	White	Mother	7	ALL

Note: ALL = Acute Lymphoblastic Leukemia

Children's Oncology Group Family Handbook was first preprocessed by extracting text and its contextual metadata (page headings, links). These entries were then converted into embeddings via OpenAI's API and indexed by Chroma, an open-source AI application database. The vector database was configured to retrieve 20 chunks of the most relevant text and use the 5 most useful chunks measured by a Maximum Marginal Relevance metric. This MMR metric is optimized to return similar documents which are also diverse from each other to ensure the retrieval is comprehensive enough to answer the user query and diverse enough to reduce redundancy or repeated information. The agent used a LangChain component to orchestrate the retrieval and generation process, with real-time streaming to deliver empathetic, context-rich responses that reflect the content of the handbook. The agent was given an instruction prompt to act as a cancer caregiver assistant and answer user questions from the Oncology Group Handbook.

An initial prototype was tested on a small group of internal users using the cognitive walk-through method. The users were given the background of caregiver with a child undergoing cancer treatment and were provided a task list to ask questions regarding the treatment of their child, evaluate usability, trust and accuracy of answers retrieved. The main feedback received was regarding the answer structure, the validity of the answers provided and tone of the agent. The enhancements made included changing the system instruction to provide clear, point-wise responses, ensuring users could quickly identify critical information. A citation referencing mechanism was incorporated in the prompt to instruct the agent to incorporate links to the original sources including page number, heading, and text. An internal usability check was tested using a chain of verification method. Lastly the agent was instructed to be more empathetic and cite when it didn't have an answer present in the handbook.

4 Findings

Overall, participants trusted the conversational AI tool but expressed concerns about its limitations, particularly in personalization, integration with healthcare systems, and privacy safeguards.

4.1 Trust and Credibility

Participants widely trusted the conversational agent, attributing its reliability to the use of validated sources like the Children's Oncology Group Handbook. P2 noted, "No, I mean not the way this is written, 'cause this... seems to be just taken directly out of a handbook that's already been approved. So I wouldn't feel nervous about using this." This sentiment was echoed by P8, who described the agent as having a natural flow akin to interacting with a nurse: "It felt trustworthy because it had a good flow to it, a natural flow. Not like you were talking on a computer." P2 also affirmed trust in the agent's ability to handle sensitive information, expressing confidence in its security measures to upload her child's medical records. Survey data reinforced this perception of trustworthiness. In the pre-survey, 58% of participants strongly agreed that they trusted the conversational agent to provide accurate and reliable health information, and 67% expressed confidence in the technology behind the agent. After using the tool, trust increased, with 67% of participants strongly agreeing that the agent was accurate and reliable. The rise in confidence from pre- to post-survey indicates that direct interaction with the tool reinforced its credibility.

4.2 Empathy vs Practical Approach

Caregivers expressed mixed opinions on the balance between empathy and practicality in the agent's responses. Some, like P10, appreciated the agent's neutrality, noting that it "feels like a neutral person" and "doesn't make you feel stupid like some doctors do." She suggested that adding empathy—such as saying, "I'm sorry your child has this diagnosis"—could help certain users feel more supported. However, few participants found empathetic tones less meaningful, P9 stated, "How do you really make an AI empathetic, though? You can't. You can say 'I'm sorry to hear that,' but you get told that enough... I would rather it just be factual." For her, straightforward, actionable advice was more valuable than emotional reassurance. In contrast, P5 praised the agent's warmth, describing the responses as "very soft, kind, warm, almost apologetic, and extremely empathetic... It was nothing but human qualities throughout the entire exchange that went on in our conversations" The survey results reflected this split perspective. 50% of the participants expressed a strong preference for AI agents that exhibit human-like behaviors, such as empathy or humor, indicating that these qualities could enhance their likelihood of using the agent. However, a quarter of the respondents were neutral or disagreed, underscoring that emotional relatability is not universally desired. This divide was further evident in the interview data, where 7/12 caregivers valued actionable, factual advice, while others found emotional reassurance through empathy more beneficial.

4.3 Personalization Needs

Participants consistently emphasized the importance of personalization in a conversational agent designed for caregiving. While customization options like tone, gender, or voice were seen as beneficial, they were considered secondary to the need for personalization. Caregivers wanted responses tailored to specific details such as the child's age, type of cancer, and treatment stage. P7 noted, "If it could tailor responses based on the specific type of cancer or patient details, that would improve its usefulness" (50:16). Similarly, P5 shared, "I would like for this agent to be able to speak specific to specific types of cancer instead of just giving the overviews for cancer." This reflects a shared desire for nuanced and context-aware guidance rather than generic advice. However, caregivers like P1 expressed frustration with the lack of depth, comparing the experience to a basic online search: "I was just basically doing a little bit more fancier Google search and then at that point I probably just go to Google and type my question."

Furthermore, pre-survey findings showed strong interest in personalized support, with 8/12 participants strongly agreeing and 4/12 somewhat agreeing that they would use the agent for health-related inquiries specific to their family's situation. Addressing this need for personalization is essential to enhance the tool's relevance and usability.

4.4 Role of the tool- Support agent

The healthcare conversational agent was majorly regarded as a supplementary resource for low-stakes caregiving tasks, particularly during moments when caregivers lacked immediate access to medical professionals. Participants emphasized its role as a collaborator in decision-making, helping organize and clarify information rather than replacing human experts. For instance, P10 shared, "It feels like a friend explaining things to me... It could be a very big partner, especially at the beginning and throughout phase one and two. It gives parents and caregivers the opportunity to ask questions when they don't get enough time with doctors." Similarly, P5 noted, "I really felt like I had a nurse or a doctor or a medical specialist sitting on the other end of the computer answering my questions." Despite its collaborative nature, many participants highlighted the need for deeper integration with healthcare systems to provide specific recommendations and ensure it becomes a trusted partner throughout the caregiving journey. The survey findings revealed that 9/12 participants trusted the tool to enhance their caregiving ability, but privacy concerns and perceived limitations in its capabilities during post-treatment phases persisted.

4.5 Ease of Use

Participants found the conversational agent easy to use, with a simple and intuitive interface that catered to a wide range of technical proficiencies. Survey results indicated that 12/12 participants strongly agreed the tool was straightforward to navigate and required minimal effort to learn. P10 elaborated, "It's really, really easy to type a question in and hit enter and wait a second for it to pop up on the screen... It's so simple that anybody can use it. Even children with older siblings who have questions could use it." While the tool's simplicity was universally appreciated, participants suggested additional features like voice-based interaction and bulleted responses to further streamline information delivery and enhance accessibility. These recommendations underscore the participants' desire for tools that minimize cognitive load during already overwhelming caregiving tasks.

5 Discussion

From the early moments of diagnosis to the complex terrain of treatment, participants viewed the conversational AI as a **tool or a collaborator** rather than a substitute for human expertise. It was often described as a "friend" or "partner," capable of organizing and clarifying information during moments of overwhelming uncertainty. One participant likened it to a trusted guide during the bewildering early days of diagnosis, when every question feels like a matter of life and death. Yet, the tool's limitations were equally apparent. Caregivers emphasized that it could never replace the nuanced, context-sensitive insights of healthcare professionals. Rather than positioning it as an oracle of caregiving wisdom, they saw it as a supplementary resource—one that could offer guidance in low-stakes scenarios or fill the gaps when medical professionals were unavailable. This reinforces the need to design AI tools with a clear understanding of their supporting role in the caregiving process.

The utility of the conversational agent varied dramatically across the caregiving timeline. During the chaotic early diagnosis phase, the tool's ability to provide quick, reliable information was seen as invaluable. For caregivers, this phase is marked by a desperate search for answers amidst a torrent of fear and confusion. Participants spoke poignantly of how an AI-driven tool could have offered clarity and calm in those critical first weeks. During treatment, the focus shifted from information overload to logistical challenges. Caregivers highlighted the tool's potential to help manage appointments, medication schedules, and procedural reminders. One participant described how it could have been a "lifesaver" by simply organizing caregiving tasks in a single accessible place. However, post-treatment, the tool's relevance waned. Experienced caregivers, having developed their own trusted networks and routines, found the tool less impactful. This highlights a design challenge: **how can such tools adapt to remain relevant during the long, evolving trajectory of caregiving**? Incorporating features tailored to survivorship care or long-term follow-up could sustain their value beyond the acute phases of treatment.

Caregiver's perceptions of the tool's tone revealed a fascinating duality. Some found its neutral and factual responses refreshing, especially when compared to the emotionally charged interactions they often experienced in person. As one participant put it, "It doesn't make you feel stupid like some doctors do." For these users, clear, actionable advice took precedence over emotional support. Others, however, longed for a touch of empathy. They wanted the tool to acknowledge the emotional gravity of their situation with phrases like, "I'm sorry your child has this diagnosis." This divergence underscores the **importance of customizable interaction styles**. A conversational agent that can toggle between empathetic and pragmatic tones might bridge this divide, catering to caregiver's varying preferences in different moments.

Finally, **trust** emerged as both a strength and a vulnerability for the tool. On the one hand, its reliance on validated sources like the Children's Oncology Group Handbook lent it an air of credibility. Many participants described the tool as trustworthy because it drew directly from vetted materials. One caregiver noted, "I wouldn't feel nervous about using this—it's taken straight from an approved handbook." Yet, this trust was not unconditional. Concerns about accuracy, particularly for rare or complex cases, lingered in the background. Participants stressed the importance of clear limitations and professional validation, underscoring that transparency is key to fostering long-term trust in AI-driven healthcare tools.

6 Limitations and Future Directions

While this study provides valuable insights into the use of conversational AI tools for pediatric cancer caregiving, several limitations must be acknowledged. First, the participant sample, though diverse in caregiving roles and cancer types, was relatively small (N=12). While this sample size aligns with qualitative research norms and reached thematic saturation, future work should include larger, more diverse cohorts to strengthen generalizability Second, the study relied heavily on self-reported data from interviews and surveys, which may introduce response bias. Another limitation stems from the controlled, simulated nature of the caregiving scenarios presented during the study. Future studies could address these limitations by expanding the participant pool to include a broader demographic and conducting longitudinal research to observe real-world usage.

Finally, while the tool showed promise in addressing low-stakes informational needs, its effectiveness in real-world, high-stakes situations, such as emergencies, remains unexplored. As we and others move to studying coversational tools for real-time information-seeking during health crises, future work will need to address ethical concerns head-on. Even for a tool such as ours which draws from validated sources, we believe further safeguards are needed to prevent the risk of misinformation, especially in complex or evolving medical cases. Future research should explore strategies for integrating AI-driven recommendations with human oversight, ensuring caregivers can make informed decisions without over-relying on automated responses.

7 Conclusion

This study highlights the importance of trust, usability, and personalization in using conversational AI tools for caregiving. Tools that provide tailored responses based on a child's diagnosis and treatment phase, while relying on trusted resources like the Children's Oncology Group Handbook, can serve as valuable supports during critical caregiving moments. Rather than replacing healthcare professionals, these tools can complement traditional care by helping caregivers navigate complex tasks. Key areas for improvement include better personalization, stronger data privacy measures, and integration with healthcare systems.

Acknowledgments

We would like to thank our participants for their generosity in sharing their stories and their time with us, and our collaborators for sharing their expertise and advice. This study was funded by the National Science Foundation under Grant No. IIS-2047432.

References

[1] Mahyar Abbasian, Iman Azimi, Amir M. Rahmani, and Ramesh Jain. 2024. Conversational Health Agents: A Personalized LLM-Powered Agent Framework. https://doi.org/10.48550/arXiv.2310.02374 arXiv:2310.02374 [cs].

[2] Tarik Abuain and Waleed AbdelKarim Abuain. 2024. Impact of Large Language Models in the Healthcare Industry: SOAR Analyses. In *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*. 1–7. https://doi.org/10.1109/ICCIT62134.2024.10701228

[3] Jacob Beattie, Dylan Owens, Ann Marie Navar, Luiza Giuliani Schmitt, Kimberly Taing, Sarah Neufeld, Daniel Yang, Christian Chukwuma, Ahmed Gul, Dong Soo Lee, Neil Desai, Dominic Moon, Jing Wang, Steve Jiang, and Michael Dohopolski. 2024. Large Language Model Augmented Clinical Trial Screening. https://doi.org/10.1101/2024.08.27.24312646

[4] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (June 2005), 293–327. https://doi.org/10.1145/1067860.1067867

[5] Joëlle Bink. [n. d.]. Personalized Response with Generative AI: Improving Customer Interaction with Zero-Shot Learning LLM Chatbots. ([n. d.]).

[6] Felix Busch, Lena Hoffmann, Christopher Rueger, Elon HC van Dijk, Rawen Kader, Esteban Ortiz-Prado, Marcus R. Makowski, Luca Saba, Martin Hadamitzky, Jakob Nikolas Kather, Daniel Truhn, Renato Cuocolo, Lisa C. Adams, and Keno K. Bressem. 2024. Systematic Review of Large Language Models for Patient Care: Current Applications and Challenges. https://doi.org/10.1101/2024.03.04.24303733 Pages: 2024.03.04.24303733.

[7] Francisco Maria Calisto, João Fernandes, Margarida Morais, Carlos Santiago, João Maria Abrantes, Nuno Nunes, and Jacinto C. Nascimento. 2023. Assertiveness-based Agent Communication for a Personalized Medicine on Medical Imaging Diagnosis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3580682

[8] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2017. Towards a chatbot for digital counselling. BCS Learning & Development. https://doi.org/10.14236/ewic/HCI2017.24

[9] Luis Cardenas, Katherine Parajes, Ming Zhu, and Shengjie Zhai. 2024. AutoHealth: Advanced LLM-Empowered Wearable Personalized Medical Butler for Parkinson's Disease Management. In *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*. 0375–0379. https://doi.org/10.1109/CCWC60891.2024.10427622

[10] Nicolas Carl, Franziska Schramm, Sarah Haggenmüller, Jakob Nikolas Kather, Martin J. Hetz, Christoph Wies, Maurice Stephan Michel, Frederik Wessels, and Titus J. Brinker. 2024. Large language model use in clinical oncology. *npj Precision Oncology* 8, 1 (Oct. 2024), 240. https://doi.org/10.1038/s41698-024-00733-4

[11] Ye Chen, Igor Couto, Wei Cai, Cong Fu, and Bruno Dorneles. 2024. SoftTiger: A Clinical Foundation Model for Healthcare Workflows. https://doi.org/10.48550/arXiv.2403.00868 arXiv:2403.00868 [cs].

[12] Children's Oncology Group. 2024. *The Children's Oncology Group Family Handbook for Children with Cancer.* Patient Resource. Children's Oncology Group. https://childrensoncologygroup.org/cog-family-handbook Available in multiple languages.

[13] Avishek Choudhury and Zaria Chaudhry. 2024. Large Language Models and User Trust: Consequence of Self-Referential Learning Loop and the Deskilling of Healthcare Professionals. *Journal of Medical Internet Research* 26 (April 2024), e56764. https://doi.org/10.2196/56764 arXiv:2403.14691 [cs].

[14] Debadutta Dash, Rahul Thapa, Juan M. Banda, Akshay Swaminathan, Morgan Cheatham, Mehr Kashyap, Nikesh Kotecha, Jonathan H. Chen, Saurabh Gombar, Lance Downing, Rachel Pedreira, Ethan Goh, Angel Arnaout, Garret Kenn Morris, Honor Magon, Matthew P. Lungren, Eric Horvitz, and Nigam H. Shah. 2023. Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in

healthcare delivery. https://doi.org/10.48550/arXiv.2304.13714 arXiv:2304.13714 [cs].

[15] Sagar Goyal, Eti Rastogi, Sree Prasanna Rajagopal, Dong Yuan, Fen Zhao, Jai Chintagunta, Gautam Naik, and Jeff Ward. 2024. HealAI: A Healthcare LLM for Effective Medical Documentation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. ACM, Merida Mexico, 1167–1168. https://doi.org/10.1145/3616855.3635739

[16] Michael Gubanov, Anna Pyayt, and Aleksandra Karolak. 2024. CancerKG.ORG - A Web-scale, Interactive, Verifiable Knowledge Graph-LLM Hybrid for Assisting with Optimal Cancer Treatment and Care. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. ACM, Boise ID USA, 4497–4505. https://doi.org/10.1145/3627673.3680094

[17] Joko Gunawan. 2023. Exploring the future of nursing: Insights from the ChatGPT model. *Belitung Nursing Journal* 9, 1 (Feb. 2023), 1–5. https://doi.org/10.33546/bnj.2551

[18] Jiyeon Han, Jimin Park, Jinyoung Huh, Uran Oh, Jaeyoung Do, and Daehee Kim. 2024. AscleAI: A LLM-based Clinical Note Management System for Enhancing Clinician Productivity. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3613905.3650784

[19] Sarah Hile, Sarah J. Erickson, Brittany Agee, and Robert D. Annett. 2014. Parental stress predicts functional outcome in pediatric cancer survivors. *Psycho-Oncology* 23, 10 (2014), 1157–1164. https://doi.org/10.1002/pon.3543 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pon.3543.

[20] Giovanni Maria Iannantuono, Dara Bracken-Clarke, Charalampos S. Floudas, Mario Roselli, James L. Gulley, and Fatima Karzai. 2023. Applications of large language models in cancer care: current evidence and future perspectives. *Frontiers in Oncology* 13 (Sept. 2023). https://doi.org/10.3389/fonc.2023.1268915 Publisher: Frontiers.

[21] Giovanni Maria Iannantuono, Dara Bracken-Clarke, Fatima Karzai, Hyoyoung Choo-Wosoba, James L Gulley, and Charalampos S Floudas. 2024. Comparison of Large Language Models in Answering Immuno-Oncology Questions: A Cross-Sectional Study. *The Oncologist* 29, 5 (May 2024), 407–414. https://doi.org/10.1093/oncolo/oyae009

[22] Zhanxu Jiang. 2024. AI in Personalized Health Management: Practices and Challenges:. In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence*. SCITEPRESS - Science and Technology Publications, Shanghai, China, 46–51. https://doi.org/10.5220/0012897900004508

[23] Zhengbao Jiang, Jun Araki, Haibo Ding, and Graham Neubig. [n. d.]. How Can We Know When Language Models Know? On the Calibration of Language Models for Question Answering. ([n. d.]). https://doi.org/10.1162/tacl_a_00407

[24] Mingyu Jin, Qinkai Yu, Chong Zhang, Dong Shu, Suiyuan Zhu, Mengnan Du, Yongfeng Zhang, and Yanda Meng. 2024. *Health-LLM: Personalized Retrieval-Augmented Disease Prediction Model.* https://doi.org/10.13140/RG.2.2.13783.04000

[25] Ehsan Kamalloo, Shivani Upadhyay, and Jimmy Lin. 2024. Towards Robust QA Evaluation via Open LLMs. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 2811–2816. https://doi.org/10.1145/3626772.3657675

[26] Kyungwon Kim, Yongmoon Lee, Doohyun Park, Taejoon Eo, Daemyung Youn, Hyesang Lee, and Dosik Hwang. 2024. LLM-Guided Multi-modal Multiple Instance Learning for 5-Year Overall Survival Prediction of Lung Cancer. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel (Eds.). Vol. 15003. Springer Nature Switzerland, Cham, 239–249. https://doi.org/10.1007/978-3-031-72384-1_23 Series Title: Lecture Notes in Computer Science.

[27] Tiffany H. Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, and Victor Tseng. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health* 2, 2 (Feb. 2023), e0000198. https://doi.org/10.1371/journal.pdig.0000198 Publisher: Public Library of Science.

[28] Wendy Landier, JoAnn Ahern, Lamia P. Barakat, Smita Bhatia, Kristin M. Bingen, Patricia G. Bondurant, Susan L. Cohn, Sarah K. Dobrozsi, Maureen Haugen, Ruth Anne Herring, Mary C. Hooke, Melissa Martin, Kathryn Murphy, Amy R. Newman, Cheryl C. Rodgers, Kathleen S. Ruccione, Jeneane Sullivan, Marianne Weiss, Janice Withycombe, Lise Yasui, and Marilyn Hockenberry. 2016. Patient/Family Education for Newly Diagnosed Pediatric Oncology Patients. *Journal of Pediatric Oncology Nursing: Official Journal of the Association of Pediatric Oncology Nurses* 33, 6 (2016), 422–431. https://doi.org/10.1177/1043454216655983

[29] Alessandro Giaj Levra, Mauro Gatti, Roberto Mene, Dana Shiffer, Giorgio Costantino, Monica Solbiati, Raffaello Furlan, and Franca Dipaola. 2025. A large language model-based clinical decision support system for syncope recognition in the emergency department: A framework for clinical workflow integration. *European Journal of Internal Medicine* 131 (Jan. 2025), 113–120. https://doi.org/10.1016/j.ejim.2024.09.017

[30] Mingchen Li, Jiatan Huang, Jeremy Yeung, Anne Blaes, Steven Johnson, Hongfang Liu, Hua Xu, and Rui Zhang. 2024. CancerLLM: A Large Language Model in Cancer Domain. https://doi.org/10.48550/arXiv.2406.10459 arXiv:2406.10459 [cs].

[31] Hanna Luetke Lanfer, Doreen Reifegerste, Winja Weber, Paula Memenga, Eva Baumann, Julia Geulen, Stefanie Klein, Anne Müller, Andrea Hahne, and Susanne Weg-Remers. 2024. Digital clinical empathy in a live chat: multiple findings from a formative qualitative study and usability tests. *BMC Health Services Research* 24, 1 (March 2024), 314. https://doi.org/10.1186/s12913-024-10785-8

[32] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[33] Kurnia Muludi, Kaira Milani Fitria, Joko Triloka, and Sutedi. 2024. Retrieval-Augmented Generation Approach: Document Question Answering using Large Language Model. *International Journal of Advanced Computer Science and Applications (IJACSA)* 15, 3 (2024). https://doi.org/10.14569/IJACSA.2024.0150379 Number: 3 Publisher: The Science and Information (SAI) Organization Limited.

[34] Sarah Nikkhah, Akash Uday Rode, Neha Kulkarni, Priyanjali Mittal, Emily Mueller, and Andrew Miller. 2022. Designing for Families as Adaptive Systems: Collaborative Emotional Support and Resilience in the Children's Hospital. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3491101.3519787

[35] Dwi Novrianda, Fitri Haryanti, Eddy Supriyadi, Luthfan Lazuardi, and Elisabeth Siti Herini. 2022. Development and Evaluation of Internet-based Health Technology in Pediatric Oncology: A Scoping Review. *Asian Pacific journal of cancer prevention: APJCP* 23, 4 (April 2022), 1125–1135. https://doi.org/10.31557/APJCP.2022.23.4.1125

[36] Dana-Mihaela Petroșanu, Alexandru Pîrjan, and Alexandru Tăbușcă. 2023. Tracing the Influence of Large Language Models across the Most Impactful Scientific Works. *Electronics* 12, 24 (Jan. 2023), 4957. https://doi.org/10.3390/electronics12244957 Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

[37] Arya Rao, Michael Pang, John Kim, Meghana Kamineni, Winston Lie, Anoop K. Prasad, Adam Landman, Keith Dreyer, and Marc D. Succi. 2023. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. *Journal of Medical Internet Research* 25, 1 (Aug. 2023), e48659. https://doi.org/10.2196/48659 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[38] Zafaryab Rasool, Stefanus Kurniawan, Sherwin Balugo, Scott Barnett, Rajesh Vasa, Courtney Chesser, Benjamin M. Hampstead, Sylvie Belleville, Kon Mouzakis, and Alex Bahar-Fuchs. 2024. Evaluating LLMs on Document-Based QA: Exact Answer Selection and Numerical Extraction using Cogtale dataset. https://doi.org/10.48550/arXiv.2311.07878 arXiv:2311.07878 [cs].

[39] Cheryl C. Rodgers, Catherine M. Laing, Ruth Anne Herring, Nancy Tena, Adrianne Leonardelli, Marilyn Hockenberry, and Verna Hendricks-Ferguson. 2016. Understanding Effective Delivery of Patient and Family Education in Pediatric Oncology: A Systematic Review From the Children's Oncology Group. *Journal of Pediatric Oncology Nursing* 33, 6 (Nov. 2016), 432–446. https://doi.org/10.1177/1043454216659449 Publisher: SAGE Publications Inc STM.

[40] Emre Sezgin, Daniel I. Jackson, A. Baki Kocaballi, Mindy Bibart, Sue Zupanec, Wendy Landier, Anthony Audino, Mark Ranalli, and Micah Skeens. 2025. Can Large Language Models Aid Caregivers of Pediatric Cancer Patients in Information Seeking? A Cross-Sectional Investigation. *Cancer Medicine* 14, 1 (Jan. 2025), e70554. https://doi.org/10.1002/cam4.70554

[41] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R. Pfohl, Heather Cole-Lewis, Darlene Neal, Qazi Mamunur Rashid, Mike Schaekermann, Amy Wang, Dev Dash, Jonathan H. Chen, Nigam H. Shah, Sami Lachgar, Philip Andrew Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Agüera y Arcas, Nenad Tomašev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle K. Barral, Dale R. Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. 2025. Toward expert-level medical question answering with large language models. *Nature Medicine* (Jan. 2025), 1–8. https://doi.org/10.1038/s41591-024-03423-7 Publisher: Nature Publishing Group.

[42] Satvik Tripathi, Rithvik Sukumaran, and Tessa S Cook. [n. d.]. Efficient healthcare with large language models: optimizing clinical workflow and enhancing patient care. ([n. d.]). https://doi.org/10.1093/jamia/ocad258

[43] Daniele Vieira, Anniely Soares, Anna Guedes, Luciano Marques dos Santos, Beatriz Toso, Elenice Maria Vaz, Neusa Collet, and Altamira Reichert. 2023. Educational intervention with nurses regarding childcare consultation: a mixed method study. *Texto & Contexto - Enfermagem* 32 (Dec. 2023). https://doi.org/10.1590/1980-265x-tce-2023-0132en

[44] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling Conversational Interaction with Mobile UI using Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. https://doi.org/10.1145/3544548.3580895

[45] Dandan Wang and Shiqing Zhang. 2024. Large language models in medical and healthcare fields: applications, advances, and challenges. *Artificial Intelligence Review* 57 (Sept. 2024). https://doi.org/10.1007/s10462-024-10921-0

[46] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. 2024. Understanding User Experience in Large Language Model Interactions. https://doi.org/10.48550/arXiv.2401.08329 arXiv:2401.08329 [cs].

[47] Ziyu Wang, Hao Li, Di Huang, and Amir M. Rahmani. 2024. HealthQ: Unveiling Questioning Capabilities of LLM Chains in Healthcare Conversations. https://doi.org/10.48550/arXiv.2409.19487 arXiv:2409.19487 [cs].

[48] Dominik Wawrzuta, Aleksandra Napieralska, Katarzyna Ludwikowska, Laimonas Jaruševičius, Anastasija Trofimoviča-Krasnorucka, Gints Rausis, Agata Szulc, Katarzyna Pędziwiatr, Kateřina Poláchová, Justyna Klejdysz, and Marzanna Chojnacka. 2025. Large language models for pretreatment education in pediatric radiation oncology: A comparative evaluation study. *Clinical and Translational Radiation Oncology* 51 (March 2025), 100914. https://doi.org/10.1016/j.ctro.2025.100914

[49] Maximilian Wutz, Marius Hermes, Vera Winter, and Juliane Köberlein-Neu. 2023. Factors Influencing the Acceptability, Acceptance, and Adoption of Conversational Agents in Health Care: Integrative Review. *Journal of Medical Internet Research* 25 (Sept. 2023), e46548. https://doi.org/10.2196/46548

[50] Yuanyuan Xu, Weiting Gao, Yining Wang, Xinyang Shan, and Yin-Shan Lin. 2024. Enhancing user experience and trust in advanced LLM-based conversational agents. *Computing and Artificial Intelligence* 2, 2 (Aug. 2024), 1467–1467. https://doi.org/10.59400/cai.v2i2.1467 Number: 2.

[51] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (May 2024), 1–35. https://doi.org/10.1145/3659625

[52] Hang Zeng, Chaoyue Niu, Fan Wu, Chengfei Lv, and Guihai Chen. 2024. Personalized LLM for Generating Customized Responses to the Same Query from Different Users. https://doi.org/10.48550/arXiv.2412.11736 arXiv:2412.11736 [cs].

Received 23 January 2025