A Survey of LLM Technology in Medical Training and Human-Computer Interaction

www.surveyx.cn

Abstract

The integration of Large Language Models (LLMs) in medical training and human-computer interaction (HCI) represents a significant interdisciplinary advancement, enhancing educational methodologies and user experiences. This survey explores the transformative potential of LLMs, highlighting their role in improving diagnostic accuracy, personalizing learning environments, and fostering interactive educational experiences through advanced technologies like augmented reality (AR) and virtual reality (VR). The survey underscores the importance of aligning LLM advancements with human values and ethical considerations, advocating for a dynamic alignment process to balance technological progress with the preservation of essential human values. It identifies key challenges, including technological resource demands and the need for standardized evaluation practices, while emphasizing the necessity of human-centered and algorithm-centered evaluations to achieve a holistic understanding of interactive machine learning systems. The survey also highlights the impact of user expectations on HCI and the importance of inclusive design practices to support diverse user needs. In conclusion, the survey calls for a reevaluation of usability practices to prioritize user empowerment and engagement, advocating for continued research and interdisciplinary collaboration to fully realize the potential of LLMs in advancing medical education and HCI, ultimately contributing to improved user experiences and patient care.

1 Introduction

1.1 Interdisciplinary Significance

The integration of Large Language Models (LLMs) with medical training and human-computer interaction (HCI) signifies a pivotal interdisciplinary advancement, fostering innovative educational paradigms and enhancing technological interfaces. This convergence addresses complex challenges in medical education and user interaction design, utilizing LLMs to automate and enrich knowledge acquisition processes [1]. In medical training, LLMs support robust decision-making, facilitating interactive learning environments tailored to healthcare education [2]. Furthermore, the incorporation of augmented reality (AR) technology enhances procedural skill training, demonstrating the synergistic potential of AR and LLMs in improving healthcare education [3].

In HCI, merging LLMs with user interface design optimizes multimodal interactions, thereby enhancing user engagement and experience [4]. This interdisciplinary approach is crucial for transcending existing HCI limitations, promoting innovations that improve user-friendliness and efficiency [5]. Integrating cognitive science and analogical reasoning into HCI is essential for refining user experiences and mitigating misinformation [6]. The emphasis on user-centered design within intelligent human-computer interaction (iHCI) underscores the necessity for a design philosophy accommodating the complexities introduced by AI technologies [7].

Moreover, interdisciplinary collaboration extends to developing adaptive interfaces that utilize computational methods for dynamic, user-centered systems [8]. Addressing emotional and cognitive

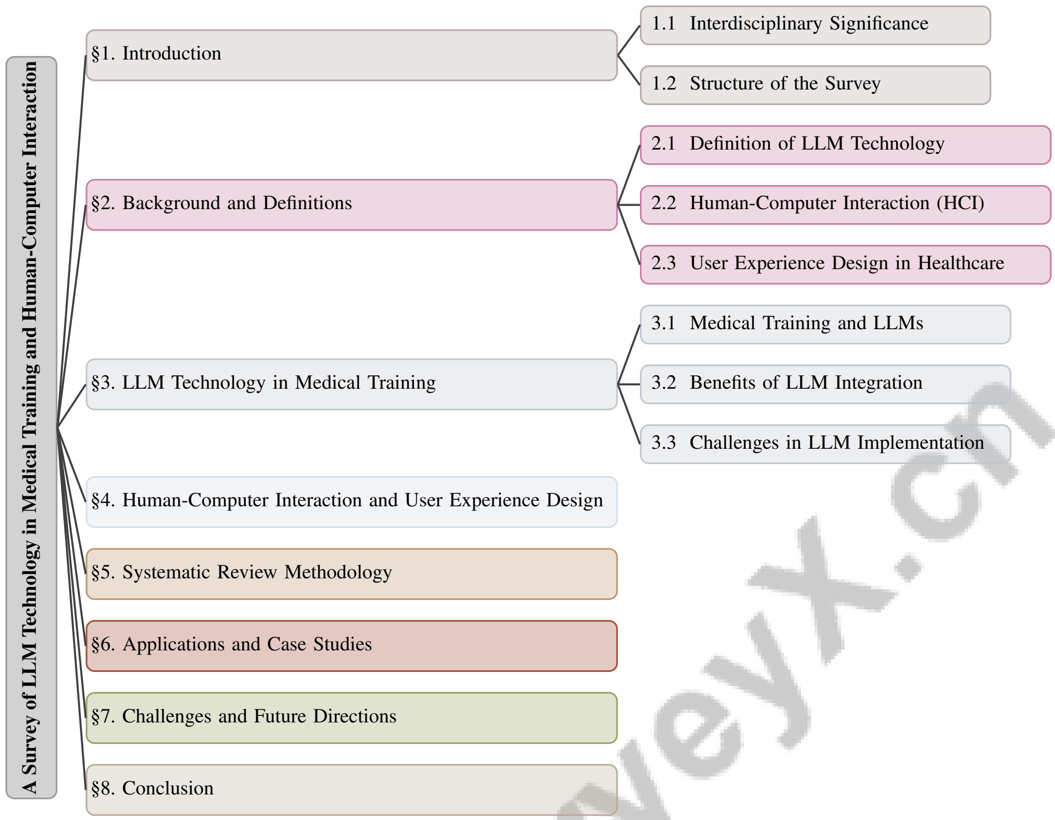

Figure 1: chapter structure

dimensions in LLMs is vital, as these factors significantly enhance human-computer interactions [9]. The survey also highlights the importance of involving older adults in user interface design, underscoring the interdisciplinary significance of technology in healthcare and education [10].

These elements collectively illustrate the transformative potential of integrating LLM technology with medical training and HCI, paving the way for innovative solutions and improved user experiences. This integration not only confronts complex global challenges but also fosters user engagement through a holistic approach, advancing both medical education and human-computer interaction. The survey further investigates how HCI researchers incorporate LLMs into their practices and the associated ethical concerns, emphasizing the broader implications of this interdisciplinary integration [11].

1.2 Structure of the Survey

This survey is meticulously structured to comprehensively explore the integration of Large Language Models (LLMs) within medical training and human-computer interaction (HCI), highlighting the interdisciplinary nature of these domains. It begins with an introduction that underscores the critical importance of integrating LLM technology with medical training and HCI practices, establishing a foundational framework for subsequent discussions on ethical implications, practical applications, and the transformative potential of this convergence in enhancing healthcare information processes and research methodologies [12, 11, 13]. Following the introduction, the survey provides background and definitions, offering a detailed overview of core concepts such as LLM technology, HCI, systematic review methodologies, healthcare education, and user experience design.

The core of the survey is divided into sections addressing specific research aspects. The section on LLM Technology in Medical Training examines how LLMs enhance medical education, discussing their applications, benefits, and implementation challenges [14]. This is followed by an exploration of Human-Computer Interaction and User Experience Design, analyzing user-centered design principles and multimodal interaction technologies to improve healthcare user experiences [15].

A dedicated section on Systematic Review Methodology outlines the methodological framework for synthesizing research findings, encompassing study selection criteria, data extraction processes, and analysis methods, which are crucial for understanding the methodologies, challenges, and implications of current HCI research practices [16].

The survey then presents Applications and Case Studies, showcasing real-world implementations of LLMs in medical training environments, thereby demonstrating their practical impact and potential. The subsequent section on Challenges and Future Directions identifies key technological and methodological hurdles while exploring potential innovations and interdisciplinary collaborations that could drive advancements in these fields.

Finally, the survey concludes with a summary of key findings, emphasizing LLM technology's transformative potential and the importance of ongoing research and collaboration across disciplines. This structured approach ensures a thorough examination of the topics, providing valuable insights into the applications, benefits, challenges, and policy implications of integrating LLM technology in multidisciplinary educational contexts [14].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definition of LLM Technology

Large Language Model (LLM) technology, which integrates information retrieval with generative capabilities, significantly enhances text processing and generation in healthcare [2]. It advances medical natural language processing (NLP) by analyzing unstructured clinical notes and supports clinical decision-making, including mental health referral triaging [2]. LLMs also model cognitive processes in medical imaging, such as predicting gaze behavior on radiographic images, thereby improving diagnostic accuracy [17].

Furthermore, LLMs reshape perceptions of authorship in educational contexts [18]. Despite their potential, challenges in transparency and responsible AI deployment persist, necessitating unified frameworks to enhance problem-solving and address emotional cognition [9]. In this survey, LLMs are pivotal in exploring intersections of medical training and human-computer interaction, automating knowledge acquisition and advancing educational methodologies and user experience design. Ethical and social dimensions are highlighted, emphasizing interdisciplinary research relevance [18].

2.2 Human-Computer Interaction (HCI)

Human-Computer Interaction (HCI) focuses on improving interactions between humans and computers through intuitive, effective, and user-friendly systems [19]. In healthcare, HCI enhances digital health solution accessibility, improving patient outcomes and engagement [6]. It addresses discrepancies between intended usability and actual experiences, which can lead to misinformation [6]. Effective interfaces must consider unconscious cognitive processes, balancing simplicity and functionality to prevent user overwhelm [8, 5]. HCI research explores interactive behavior modeling through user interactions with input devices, revealing methodological limitations [20].

In healthcare, HCI emphasizes user-centered systems that streamline processes for non-specialists, including patients and providers [7]. This approach accommodates diverse user groups, especially older adults who may struggle with abstract concepts [10]. HCI must also address ethical challenges in integrating LLMs, ensuring responsible technology deployment [11]. By tackling interpretability, design coherence, and cultural inclusivity, HCI advances healthcare technology and patient-centered care, engaging diverse user groups and meeting varied needs [21].

2.3 User Experience Design in Healthcare

User experience (UX) design in healthcare focuses on creating intuitive, accessible systems tailored to patients and providers. This approach is crucial for developing functional, user-friendly smart healthcare devices and digital solutions [19]. Evaluating Virtual Learning Environments (VLEs) through metrics like the HCI and Educational Index highlights the importance of integrating educational principles with UX to enhance learning outcomes [22].

Incorporating user-centered design principles ensures healthcare technologies cater to diverse needs, including varying technological proficiency and accessibility requirements. Device usability directly influences patient outcomes and care delivery efficiency, necessitating consideration of cognitive and emotional user interactions for seamless experiences [19]. However, developing effective HCI solutions in healthcare faces challenges such as technological constraints and resource consumption [23]. Addressing these challenges ensures interfaces are innovative, practical, and sustainable. The relevance of multimodal systems, which integrate various input and output modalities, underscores the potential for dynamic healthcare interfaces accommodating broader user interactions [24].

Prioritizing UX design in healthcare technologies enhances acceptance and effectiveness, improving interactions between patients and providers, streamlining information retrieval, and fostering better communication. These advancements contribute to higher patient care and satisfaction levels, offering intuitive, efficient, and personalized solutions. Applying HCI principles can significantly improve training environments and healthcare delivery, supporting an informed and effective healthcare ecosystem [13, 19, 25]. The integration of user-centered design in healthcare exemplifies HCI's evolving nature and its critical role in advancing healthcare technologies.

In recent years, the integration of Large Language Models (LLMs) into various fields has garnered significant attention, particularly in medical training. As illustrated in Figure 2, this figure highlights the multifaceted role of LLMs in simulating complex scenarios and providing educational feedback. It categorizes the numerous benefits associated with LLMs, including personalization, creativity, and the enhancement of diagnostic tools. However, it also addresses the inherent challenges, such as the complexity of evaluation and concerns regarding data quality. By emphasizing these aspects, the figure underscores the transformative potential of LLMs in enriching medical education through immersive learning experiences and a user-centered design approach. This comprehensive view not only elucidates the advantages of LLMs but also prompts critical reflection on the obstacles that must be navigated to fully harness their capabilities in educational contexts.

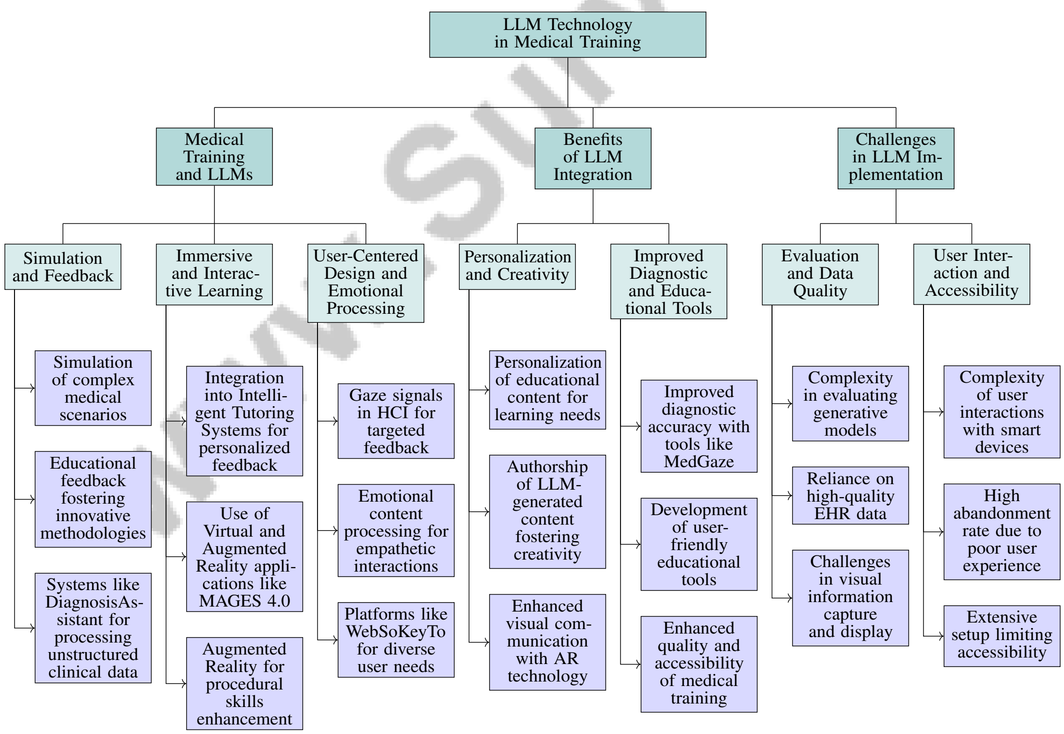

Figure 2: This figure illustrates the integration of Large Language Models (LLMs) in medical training, highlighting their role in simulating complex scenarios and providing educational feedback. It categorizes the benefits of LLMs, such as personalization and creativity, and improved diagnostic tools, while also addressing challenges like evaluation complexity and data quality. The figure underscores LLMs' transformative potential in enhancing medical education through immersive learning experiences and user-centered design.

3 LLM Technology in Medical Training

3.1 Medical Training and LLMs

Large Language Models (LLMs) have revolutionized medical training by simulating complex medical scenarios and providing educational feedback, thus fostering innovative educational methodologies. Systems like DiagnosisAssistant exemplify LLMs' capability to process unstructured clinical data from electronic health records (EHRs), enhancing decision-making and team assembly in clinical settings [2]. Benchmark evaluations demonstrate LLMs' proficiency across various internal medicine subspecialties, aiding trainees in mastering diverse medical knowledge [26].

LLMs integrated into Intelligent Tutoring Systems (ITS) offer personalized feedback, increasing interactivity and engagement among medical trainees, especially under resource constraints [27, 28]. Virtual and augmented reality applications, such as MAGES 4.0, leverage LLMs to create immersive educational experiences, facilitating AI-driven discussions [29]. Augmented Reality (AR) further enriches training by simulating procedures like central venous catheter placement, thereby enhancing procedural skills [3].

Incorporating gaze signals into human-computer interaction (HCI) enriches simulations by identifying focal areas in medical images, providing targeted feedback. MedGaze, for instance, predicts radiologists' scan paths on X-ray images, aiding novice radiologists [30, 17]. Additionally, advancements in processing emotional content have improved LLMs' empathetic interactions, crucial for training tools responsive to learners' emotional and cognitive needs [31].

User-centered design methodologies, exemplified by platforms like WebSoKeyTo, emphasize creating scenarios that accommodate diverse user needs, including those of individuals with disabilities [4]. Participatory design principles for older adults further showcase LLMs' adaptability in varied educational contexts [10].

As illustrated in Figure 3, the integration of LLMs in medical training highlights key applications, enhanced training methods, and design methodologies, while also addressing the ethical challenges involved. By enhancing scenario simulations and feedback, LLMs are essential in preparing healthcare professionals for modern medical complexities. Interdisciplinary collaboration continues to address ethical challenges, ensuring responsible LLM integration in medical education [32].

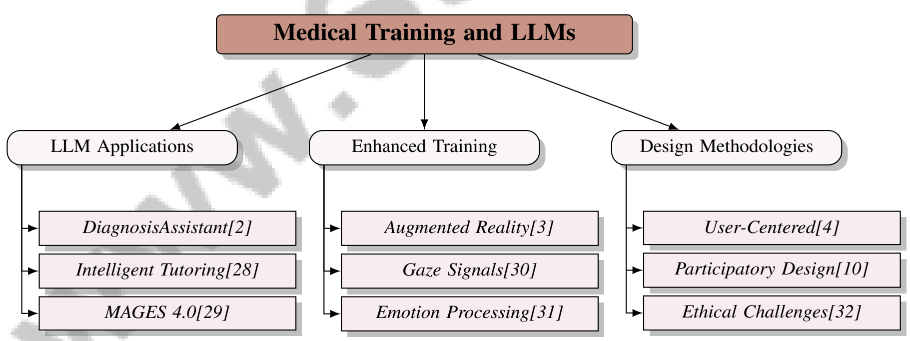

Figure 3: This figure illustrates the integration of Large Language Models (LLMs) in medical training, highlighting key applications, enhanced training methods, and design methodologies, as well as the ethical challenges involved.

3.2 Benefits of LLM Integration

Integrating LLMs into medical education offers substantial benefits, enhancing instructional methodologies and learner engagement. A key advantage is the personalization of user experiences, as LLMs tailor educational content to individual learning needs, improving efficiency in mental health triage and referral processes [2]. LLMs also foster creativity and diverse perspectives, as users may assert authorship of LLM-generated content, enriching educational experiences [18].

The combination of AR technology with LLMs enhances medical education by improving visual communication and engagement, particularly in complex procedures [3]. MedGaze's integration into training has improved diagnostic accuracy and standardized training for novice radiologists [17].

LLMs also contribute to developing user-friendly educational tools, broadening access to advanced medical training. Tools like Chat-Bot-Kit provide metrics on user interactions, enhancing research quality in computer-mediated communication (CMC) and HCI [33]. Integrating user interface design education for older adults boosts their confidence and engagement with technology, benefiting from LLM-based educational tools [10].

The benefits of LLM technology in medical education include enhanced diagnostic accuracy, personalized learning, and the promotion of creativity and innovation. These advantages underscore LLMs' transformative potential in revolutionizing healthcare education, equipping professionals to navigate contemporary medical complexities. By optimizing information retrieval, automating medical simulation scenarios, and enhancing diagnostic learning through multimodal data visualization, LLMs improve the quality and accessibility of medical training, accelerating training processes and fostering a more informed healthcare environment, ultimately leading to better patient care [34, 28, 35, 13].

3.3 Challenges in LLM Implementation

Implementing LLMs in medical training presents challenges that must be addressed to fully realize their potential. A significant challenge is evaluating generative models' complex, context-dependent nature, compounded by the lack of standardized, reproducible evaluation practices [9]. This absence complicates assessing LLMs' effectiveness in medical education, where precision and reliability are crucial.

Another hurdle is the reliance on high-quality, complete EHR data, which affects the accuracy of LLM-generated triage recommendations [2]. Robust data management is essential for reliable decision-making in medical contexts.

In AR, challenges include effectively communicating spatial orientation and tool handling, and overcoming limitations in visual information capture and display [3]. These issues are significant when integrating AR with LLMs to enhance training, as visual representation shortcomings can impede learning.

The complexity of user interactions with smart healthcare devices requires multidisciplinary collaboration to bridge the gap between technological advancements and user acceptance [19]. This complexity is compounded by the high abandonment rate of applications, such as diabetes management apps, due to poor user experience and cumbersome data input [36].

Additionally, the extensive setup required to utilize LLM features, as seen in tools like Chat-Bot-Kit, can limit accessibility for users with limited technical expertise [33]. Limited dataset sizes and diversity, as used in MedGaze, affect LLM models' generalizability, challenging their application in diverse training scenarios [17]. The increasing number of references and potential dilution of citation quality present challenges in maintaining scholarly rigor in LLM implementations [21].

Addressing the challenges of LLMs in medical training is crucial for their successful integration, ensuring these technologies enhance educational methodologies while upholding accuracy, reliability, and ethics. This involves tackling ethical dilemmas like hallucination, accountability, and bias reduction, while optimizing information retrieval in healthcare. Implementing ethical frameworks and dynamic auditing systems can promote transparency and responsible LLM use, enriching the educational experience for healthcare professionals and empowering patients with reliable health information [12, 13].

4 Human-Computer Interaction and User Experience Design

4.1 User-Centered Design Principles

User-centered design (UCD) principles are crucial in developing healthcare interfaces that prioritize user needs, ensuring systems are intuitive, accessible, and effective. Continuous user involvement during the design process enhances user experiences and outcomes, particularly in healthcare settings where engagement directly impacts patient care and satisfaction [19]. UCD emphasizes inclusivity, adaptability, and empowerment, accommodating diverse user groups, including older adults and individuals with varying technological proficiency. This approach aligns with theories of joint cognitive systems, situational awareness, and intelligent agents, focusing on augmenting human capabilities rather than replacing them [7].

As illustrated in Figure 4, the hierarchical structure of user-centered design principles in healthcare interfaces highlights core principles, application areas, and challenges, along with future work directions. This visual representation underscores the necessity of integrating explainable AI systems in healthcare, promoting transparency and usability, which fosters trust among users. Leveraging multimodal systems enhances usability, flexibility, and reliability, facilitating dynamic interactions that improve task efficiency and user satisfaction. The ARISES method exemplifies UCD's effectiveness by integrating real-time data input, machine learning predictions, and user-friendly design to facilitate interaction [36].

Prioritizing user needs leads to greater acceptance and effectiveness of healthcare interfaces, ultimately enhancing patient care and satisfaction. Human-in-the-loop (HITL) systems highlight the role of UCD principles, particularly in natural language processing (NLP) and large language models (LLMs). These systems rely on well-designed interfaces to capture human feedback, essential for iterative improvements in model performance. Recent research highlights diverse methodologies integrating human feedback across various NLP tasks and explores human-centered design principles within LLMs, leading to guidelines that enhance user experience and interaction effectiveness. Studies emphasize the impact of document organization on user engagement, optimizing these elements to meet users' varied needs [37, 38]. This user-centered focus ensures that healthcare technologies are innovative, practical, and sustainable.

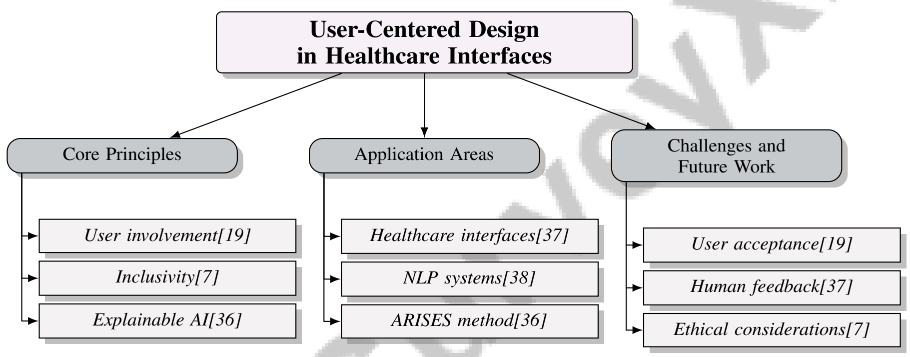

Figure 4: This figure illustrates the hierarchical structure of user-centered design principles in healthcare interfaces, highlighting core principles, application areas, and challenges with future work directions.

4.2 Multimodal Interaction Technologies

Multimodal interaction technologies enhance user experiences by integrating various input modalities, such as voice, gestures, and body language, creating natural and intuitive interfaces [23]. In healthcare, these technologies enable seamless interaction by combining multiple input recognition methods, including speech and gestures, with advanced processing techniques like signal and semantic fusion [24]. This integration is vital for developing user-friendly interfaces that support complex interactions in dynamic environments.

Incorporating physiological signals to detect cognitive states, such as stress, further amplifies the potential of these technologies in workplace settings [39]. By leveraging these signals, systems can adapt to users' emotional and cognitive needs, enhancing the overall user experience. Iterative development through virtual prototyping allows for continuous refinement, ensuring technologies meet users' evolving needs [40].

Cognitive Dimensions can be applied in both summative and formative evaluations to identify characteristics that may either impede or enhance user interactions, thereby improving the design and implementation of multimodal systems [5]. Additionally, employing Byte Pair Encoding (BPE) to analyze mouse and keyboard behavior can enhance task recognition, contributing to improved user experiences in multimodal systems [20].

Multimodal Large Language Models (LLMs) offer further opportunities to enhance user interactions by integrating auditory and visual inputs, broadening user engagement [41]. This integration is particularly beneficial in healthcare, where processing multiple modalities can significantly improve patient care and communication.

The exploration and implementation of multimodal interaction technologies, including virtual reality (VR) and haptic feedback devices, hold considerable potential to transform user experiences across various fields, especially in healthcare. By enabling more natural, responsive, and engaging interactions, these technologies can improve task completion rates and reduce errors, thereby enhancing user-centric training environments and cognitive engagement in medical applications. Research indicates that integrating multiple modalities not only supports effective communication but also paves the way for innovative applications that combine VR and haptic feedback, promising to revolutionize user interactions in diverse domains [24, 42, 25].

5 Systematic Review Methodology

5.1 Study Selection Criteria

This systematic review's study selection criteria aim to rigorously evaluate the integration of Large Language Models (LLMs) in medical training and human-computer interaction (HCI). It focuses on studies addressing the challenges of evaluating multimodal models, particularly their ability to learn interactions across modalities and execute complex reasoning [43]. The absence of consensus on HCI methodologies is acknowledged as a significant obstacle to creating unified usability assessment frameworks [44].

A qualitative coding scheme is employed to categorize relevant studies and establish criteria based on inter-rater reliability, emphasizing the capture of nuanced human requirements across applications. This method is supported by empirical research involving participant engagement in user studies, ensuring a thorough evaluation of user interactions with LLMs and HCI systems [45]. The structured study selection also incorporates quantitative analyses of citation trends within the HCI domain, focusing on articles from the CHI proceedings between 1981 and 2024 [21]. Additionally, studies utilizing qualitative coding schemes, such as those investigating referral usages in science fiction, are included [46].

By adhering to stringent selection criteria, this review synthesizes high-quality research findings that enhance understanding of LLMs in advancing medical training and HCI. The synthesis illuminates ethical considerations and practical applications of LLMs throughout the research pipeline, promoting a more inclusive and rigorous framework for systematic review methodologies within HCI [15, 11].

5.2 Data Extraction Process

The data extraction process in this systematic review is meticulously structured to ensure comprehensive and accurate collection and analysis of relevant information from selected studies, reflecting high standards in research synthesis within the HCI domain [15, 47]. It begins with identifying key data points aligning with research objectives, particularly regarding LLM integration in medical training and HCI.

A standardized data extraction form facilitates consistent data collection, capturing bibliographic details, study design, participant demographics, intervention specifics, outcomes, and reported challenges. Governed by a qualitative coding scheme, this approach categorizes data by thematic relevance and methodological rigor, ensuring comprehensiveness and contextual pertinence. The use of open-source tools for automatic information extraction, alongside frameworks like the Six Stages of Information Search Model, enhances query interpretation and streamlines access to critical healthcare information, supporting informed decision-making in clinical settings [15, 48, 13].

User-driven assessments, including pre- and post-prototype use questionnaires and interviews, are integral for evaluating prototype effectiveness and accuracy, offering insights into user experiences and the practical implications of LLM and HCI integration [49]. This approach enriches data extraction and ensures adequate representation of user perspectives.

Additionally, iterative cross-verification by multiple reviewers enhances reliability and minimizes bias, with discrepancies addressed in consensus meetings among domain experts. This rigorous data extraction methodology underpins the systematic review, enabling a thorough synthesis of research findings and contributing to a nuanced understanding of LLMs' role in advancing medical education and HCI [50, 51, 47, 15].

5.3 Analysis and Synthesis Methods

The analysis and synthesis methods in this systematic review are designed to rigorously evaluate the integration of Large Language Models (LLMs) in medical training and human-computer interaction (HCI). A thematic analysis approach is used to examine qualitative data collected through interactions facilitated by LLM-enabled chatbots, focusing on identifying discrepancies between human and LLM-generated outputs [1]. This method provides insights into the synergies and divergences in human-LLM interactions, enhancing educational methodologies and user experience design.

Current methods are organized into distinct stages, including an analysis of historical citation trends, the impact of editorial policy changes, and the role of collaborative practices in shaping the field [21]. This contextualization of research practices informs the deployment of LLM technology across interdisciplinary domains.

Moreover, integrating organizational accountability into AI systems is emphasized as critical for ensuring trustworthiness and reliability in healthcare applications [52]. This focus aligns AI deployment with ethical and organizational standards, fostering trust and acceptance among users.

The integration of diverse methodologies provides a robust framework for synthesizing research findings, facilitating a thorough evaluation of LLMs' impact on enhancing medical training and HCI. This includes assessing LLM-generated summaries of biomedical literature, which, while effective in summarizing individual articles, often struggle with synthesizing multiple sources accurately. Furthermore, it addresses ethical considerations surrounding LLM usage in research, highlighting the complexities and responsibilities involved in their application. Collectively, these elements contribute to a comprehensive understanding of LLMs' roles in advancing medical education and HCI methodologies [11, 15, 47]. This systematic approach ensures the review captures the complexity and multifaceted nature of LLM integration, ultimately informing decision-making and future research directions in these fields.

6 Applications and Case Studies

6.1 Applications of LLMs in Medical Training

Large Language Models (LLMs) have notably advanced medical training by enhancing diagnostic accuracy and promoting effective learning. MEDCO exemplifies this by personalizing learning and improving diagnostic skills among simulated students [53]. LLMs facilitate the automation of medical simulation scenario generation, allowing educators to concentrate on content delivery [34]. The LLaVA-Med model demonstrates strong multimodal conversational skills, surpassing previous models, thus improving student engagement through realistic patient interaction simulations [54].

Serious Games for Medical Operations (SGMO) utilize LLMs to enhance training outcomes and patient experiences during perioperative periods, showcasing their role in immersive training tools [55]. Platforms like MAGES 4.0 expedite collaborative VR training simulations, enriching medical training through efficient and high-quality simulations [29]. The integration of VR with LLMs provides hands-on experience in controlled environments.

HEMM and multiconcept multivariate ELO models highlight LLMs' adaptability, leveraging various modalities to enrich medical training [43, 56]. These frameworks support adaptive learning systems tailored to diverse medical student needs. Visual query builders further enhance analytical skills and critical thinking, emphasizing LLMs' role in improving educational outcomes and diagnostic accuracy [57]. By automating scenario development, LLMs conserve resources while ensuring patient privacy and engagement, equipping healthcare professionals for modern medical practice complexities [34, 28].

6.2 Case Studies and Real-World Implementations

LLMs' transformative impact in medical training is evidenced by various case studies. LLM-powered GUI agents enhance user interactions and educational outcomes across diverse environments

[58]. In telemedicine, LLMs facilitate remote consultations by providing real-time translations and context-aware responses, improving care quality and addressing healthcare access disparities [13, 28, 59, 60, 61].

LLMs also enhance intelligent tutoring systems, improving knowledge retention and engagement through personalized feedback [62, 63, 64]. These systems cater to individual learning needs, fostering interactive educational experiences. In clinical decision support, LLMs assist in diagnosing complex cases by analyzing extensive medical literature and patient data, enhancing diagnostic accuracy and timeliness [47, 48, 35, 65].

In VR training simulations, LLMs enable dynamic scenario generation and real-time virtual patient interactions, allowing students to refine clinical skills in a risk-free environment [34, 56, 66]. These case studies highlight LLMs' potential to revolutionize medical training and practice, optimizing information search and query interpretation, and streamlining data retrieval from complex databases [34, 64]. By automating high-quality scenario generation, LLMs reduce resource demands while enhancing educational engagement. Integrating LLMs into healthcare education and communication empowers practitioners and patients, contributing to improved healthcare outcomes [13]. The exploration of LLM applications continues to reveal opportunities for advancing medical education and patient care, paving the way for future advancements in the field.

7 Challenges and Future Directions

7.1 Technological and Methodological Challenges

Integrating Large Language Models (LLMs) into medical training and human-computer interaction (HCI) presents several technological and methodological challenges. A significant technological hurdle is the extensive computational resources required for LLM deployment, which limits accessibility and scalability. The prediction of gaze patterns in medical images, distinct from natural images, complicates LLM applications in medical imaging [17]. This challenge is compounded by the necessity for accurate user input and consistent engagement, as demonstrated by the development of handheld interfaces for healthcare interactions [36].

The integration of augmented reality (AR) with LLMs introduces additional challenges, particularly in ensuring long-term skill retention among trainees and managing cognitive load during training sessions. The diverse needs of user groups and the lack of comprehensive user engagement further complicate the deployment of smart healthcare devices, necessitating intuitive and adaptable interfaces [19].

Methodologically, the lack of standardized evaluation practices for LLMs complicates their effectiveness assessment in real-world scenarios. Psychological conflicts regarding authorship and ownership of LLM-generated content raise intellectual property concerns. The interplay of cognitive dimensions in interface design further complicates usability issue identification, requiring consideration of user perspectives, cognitive processes, and the broader social context influencing technology interaction [5, 6, 67, 68].

Balancing user needs with intelligent system capabilities poses another methodological challenge, particularly regarding safety and effectiveness in user interactions. Limited technical awareness and difficulties in grasping abstract concepts among older adults hinder their involvement in technology design processes. This underscores the need for inclusive and adaptive design methodologies that cater to diverse learning styles and empower older adults in technology development. Tailored educational approaches, such as workshops on user interface design, can bridge understanding gaps, while integrating Explainable Artificial Intelligence (XAI) in e-health interfaces enhances usability for older users through intuitive visualizations and clear explanations. Culturally sensitive and common-sense-based design principles can significantly improve human-computer interactions for older adults, enhancing their engagement and well-being in a technology-driven environment [69, 10, 6, 70, 71].

Addressing these challenges requires developing standardized evaluation methods and frameworks that consider the unique complexities of LLMs. Promoting interdisciplinary collaboration is essential for translating LLM advancements into practical solutions, addressing ethical challenges such as accountability, bias reduction, and transparency, while also creating tailored frameworks and dynamic auditing systems. By fostering partnerships across various fields, we can responsibly integrate LLM innovations into society, enhancing information dissemination and human-computer interaction [72, 12, 73, 11, 64]. Continued exploration of these challenges will be pivotal for advancing LLM integration in medical training and HCI, ultimately improving the effectiveness and reliability of these technologies.

7.2 Future Directions and Innovations

The future of Large Language Models (LLMs) in medical training and human-computer interaction (HCI) is set for significant advancements through various promising innovations. One key direction involves incorporating advanced technologies, such as avatar representations and synthesized gestures, within augmented reality (AR) environments to enrich the learning experience in medical training [3]. This integration promises more immersive educational experiences, facilitating the effective simulation of medical procedures.

Exploring uncontrolled settings for data collection and investigating diverse natural language processing (NLP) methods can enhance understanding of the relationship between different input modalities [20]. Such exploration may yield robust models capable of managing complex interactions in real-world scenarios, thus broadening the applicability of LLMs across healthcare contexts.

Establishing robust ethical guidelines and enhancing training for researchers on LLM ethics are critical future directions. Interdisciplinary collaboration is vital for addressing ethical challenges associated with LLM deployment, ensuring responsible and equitable technology use [11]. Furthermore, refining user intention recognition techniques and improving collaborative interfaces are essential for advancing intelligent human-computer interaction (iHCI), with a focus on the ethical implications of these technologies [7].

Expanding dataset diversity and applying methodologies like MedGaze to various imaging modalities can enhance clinical applications, providing comprehensive insights into medical imaging processes [17]. This expansion will enable LLMs to support a broader range of diagnostic and educational tasks, ultimately improving patient care and medical training.

Future research should also emphasize developing comprehensive methodologies for teaching design to older adults, incorporating role models and real-life examples of successful senior designers [10]. This approach will ensure technological advancements are inclusive and accessible to all user groups, promoting greater engagement and participation.

Lastly, developing more intuitive interfaces and enhancing user engagement in the design process will be crucial for leveraging emerging technologies in healthcare [19]. By prioritizing user-centered design principles, future innovations in LLM technology can lead to more effective and user-friendly healthcare solutions, ultimately transforming medical education and practice.

8 Conclusion

This survey illustrates the profound impact of Large Language Models (LLMs) on medical training and human-computer interaction (HCI), highlighting their potential to transform educational frameworks and enhance user engagement. LLMs play a pivotal role in elevating diagnostic precision and fostering adaptive learning environments, thus addressing a wide spectrum of educational demands. Their integration with technologies like augmented reality (AR) and virtual reality (VR) enriches the immersive experience of medical training, equipping healthcare professionals to adeptly manage intricate clinical situations.

The survey underscores the necessity of aligning LLM innovations with evolving human values and ethical standards, advocating for a dynamic balance between technological advancement and the preservation of core human principles. This balance is crucial for cultivating meaningful interactions and enhancing user well-being. Understanding the significance of meaning within HCI systems empowers users and prompts a reassessment of institutional objectives that may limit user autonomy.

Moreover, the survey emphasizes the importance of continued research and collaboration to address the challenges of LLM deployment, particularly concerning technological resource demands and the development of standardized evaluation frameworks. A comprehensive understanding of interactive machine learning (iML) systems necessitates integrating both human-centered and algorithm-centered evaluations, ensuring a holistic perspective that incorporates human and algorithmic insights.

The findings also highlight the substantial impact of user expectations on HCI, where subjective performance assessments can be influenced by perceived effects, signaling the need for further investigation into this area. Additionally, the survey stresses the importance of inclusive design in creating digital tools that cater to diverse user needs, ensuring accessibility and usability across various demographics.

References

[1] Han Meng, Yitian Yang, Yunan Li, Jungup Lee, and Yi-Chieh Lee. Exploring the potential of human-llm synergy in advancing qualitative analysis: A case study on mental-illness stigma, 2024.

[2] Niall Taylor, Andrey Kormilitzin, Isabelle Lorge, Alejo Nevado-Holgado, and Dan W Joyce. Bespoke large language models for digital triage assistance in mental health care, 2024.

[3] Manuel Rebol, Krzysztof Pietroszek, Neal Sikka, Claudia Ranniger, Colton Hood, Adam Rutenberg, Puja Sasankan, and Christian Gütl. Evaluating augmented reality communication: How can we teach procedural skill in ar?, 2023.

[4] Frédéric Vella, Flavien Clastres-Babou, Nadine Vigouroux, Philippe Truillet, Charline Calmels, Caroline Mercadier, Karine Gigaud, Margot Issanchou, Kristina Gourinovitch, and Anne Garaix. User centred method to design a platform to design augmentative and alternative communication assistive technologies, 2022.

[5] Gene Golovchinsky. Cognitive dimensions analysis of interfaces for information seeking, 2009.

[6] Sara Kingsley. A cognitive science perspective for learning how to design meaningful user experiences and human-centered technology, 2021.

[7] Wei Xu. User centered design (vi): Human factors approaches for intelligent human-computer interaction, 2022.

[8] Richard V. Diamond. User experience, software interfaces, and the unconscious, 2009.

[9] Q. Vera Liao and Ziang Xiao. Rethinking model evaluation as narrowing the socio-technical gap, 2025.

[10] Dorota Orzeszek, Wieslaw Kopec, Marcin Wichrowski, Radoslaw Nielek, Bartlomiej Balcerzak, Grzegorz Kowalik, and Malwina Puchalska-Kaminska. Beyond participatory design: Towards a model for teaching seniors application design, 2017.

[11] Shivani Kapania, Ruiyi Wang, Toby Jia-Jun Li, Tianshi Li, and Hong Shen. "i'm categorizing llm as a productivity tool": Examining ethics of llm use in hci research practices, 2024.

[12] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. Navigating llm ethics: Advancements, challenges, and future directions. *arXiv preprint arXiv:2406.18841*, 2024.

[13] Forhan Bin Emdad and Mohammad Ishtiaque Rahman. Potential renovation of information search process with the power of large language model for healthcare, 2024.

[14] Ikpe Justice Akpan, Yawo M. Kobara, Josiah Owolabi, Asuama Akpam, and Onyebuchi Felix Offodile. An investigation into the scientific landscape of the conversational and generative artificial intelligence, and human-chatbot interaction in education and research, 2024.

[15] Katja Rogers and Katie Seaborn. The systematic review-lution: A manifesto to promote rigour and inclusivity in research synthesis, 2023.

[16] Katie Shilton et al. Values and ethics in human-computer interaction. *Foundations and Trends® in Human–Computer Interaction*, 12(2):107–171, 2018.

[17] Akash Awasthi, Ngan Le, Zhigang Deng, Rishi Agrawal, Carol C. Wu, and Hien Van Nguyen. Multimodal learning and cognitive processes in radiology: Medgaze for chest x-ray scanpath prediction, 2024.

[18] Azmine Toushik Wasi, Mst Rafia Islam, and Raima Islam. Llms as writing assistants: Exploring perspectives on sense of ownership and reasoning, 2024.

[19] Pu Liu, Sidney Fels, Nicholas West, and Matthias Görges. Human computer interaction design for mobile devices based on a smart healthcare architecture, 2019.

[20] Guanhua Zhang, Matteo Bortoletto, Zhiming Hu, Lei Shi, Mihai Bâce, and Andreas Bulling. Exploring natural language processing methods for interactive behaviour modelling, 2023.

[21] Jonas Oppenlaender. Past, present, and future of citation practices in hci, 2025.

[22] Vita Hinze-Hoare. Further evaluationh of vles using hci and educational metrics, 2006.

[23] Hong Lin, Zirun Gan, Wensheng Gan, Zhenlian Qi, Yuehua Wang, and Philip S. Yu. Interaction in metaverse: A survey, 2023.

[24] Muhammad Zeeshan Baig and Manolya Kavakli. Multimodal systems: Taxonomy, methods, and challenges, 2020.

[25] Alireza Sadeghi Milani, Aaron Cecil-Xavier, Avinash Gupta, Joe Cecil, and Shelia Kennison. A systematic review of human–computer interaction (hci) research in medical and other engineering fields. *International Journal of Human–Computer Interaction*, 40(3):515–536, 2024.

[26] Sean Wu, Michael Koo, Lesley Blum, Andy Black, Liyo Kao, Fabien Scalzo, and Ira Kurtz. A comparative study of open-source large language models, gpt-4 and claude 2: Multiple-choice test taking in nephrology, 2023.

[27] Qiuyu Lu, Jiawei Fang, Zhihao Yao, Yue Yang, Shiqing Lyu, Haipeng Mi, and Lining Yao. Enabling generative design tools with llm agents for mechanical computation devices: A case study, 2024.

[28] Jinqiang Wang, Huansheng Ning, Yi Peng, Qikai Wei, Daniel Tesfai, Wenwei Mao, Tao Zhu, and Runhe Huang. A survey on large language models from general purpose to medical applications: Datasets, methodologies, and evaluations, 2024.

[29] Paul Zikas, Antonis Protopsaltis, Nick Lydatakis, Mike Kentros, Stratos Geronikolakis, Steve Kateros, Manos Kamarianakis, Giannis Evangelou, Achilleas Filippidis, Eleni Grigoriou, Dimitris Angelis, Michail Tamiolakis, Michael Dodis, George Kokiadis, John Petropoulos, Maria Pateraki, and George Papagiannakis. Mages 4.0: Accelerating the world's transition to vr training and democratizing the authoring of the medical metaverse, 2023.

[30] Tejaswini Ganapathi, David Vining, Roland Bassett, Naveen Garg, and Mia Markey. A human computer interaction solution for radiology reporting: Evaluation of the factors of variation, 2016.

[31] Yuyan Chen and Yanghua Xiao. Recent advancement of emotion cognition in large language models, 2024.

[32] Wei Xu and Marvin Dainoff. Enabling human-centered ai: A new junction and shared journey between ai and hci communities, 2023.

[33] Kyoko Sugisaki. Chat-bot-kit: A web-based tool to simulate text-based interactions between humans and with computers, 2019.

[34] Scott Sumpter. Automated generation of high-quality medical simulation scenarios through integration of semi-structured data and large language models, 2024.

[35] Yang Ouyang, Yuchen Wu, He Wang, Chenyang Zhang, Furui Cheng, Chang Jiang, Lixia Jin, Yuanwu Cao, and Quan Li. Leveraging historical medical records as a proxy via multimodal modeling and visualization to enrich medical diagnostic learning, 2023.

[36] Robert Spence, Chukwuma Uduku, Kezhi Li, Nick Oliver, and Pantelis Georgiou. A novel hand-held interface supporting the self-management of type 1 diabetes, 2020.

[37] Zijie J. Wang, Dongjin Choi, Shenyu Xu, and Diyi Yang. Putting humans in the natural language processing loop: A survey, 2021.

[38] Yuchen Wang, Yin-Shan Lin, Ruixin Huang, Jinyin Wang, and Sensen Liu. Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse software learning needs with a single document knowledge base, 2024.

AI-generated, for reference only.

[39] Marco Filetti and Jari Torniainen. A short review and primer on multimodal psychophysiological applications in work-related human computer interaction, 2016.

[40] Florian Fischer, Arthur Fleig, Markus Klar, Viktorija Paneva, and Jörg Müller. Towards a deep(er) understanding of interaction through modeling, simulation, and optimization, 2023.

[41] Efe Bozkir, Süleyman Özdel, Ka Hei Carrie Lau, Mengdi Wang, Hong Gao, and Enkelejda Kasneci. Embedding large language models into extended reality: Opportunities and challenges for inclusion, engagement, and privacy, 2024.

[42] Jose Daniel Azofeifa, Julieta Noguez, Sergio Ruiz, José Martín Molina-Espinosa, Alejandra J Magana, and Bedrich Benes. Systematic review of multimodal human–computer interaction. In *Informatics*, volume 9, page 13. MDPI, 2022.

[43] Paul Pu Liang, Akshay Goindani, Talha Chafekar, Leena Mathur, Haofei Yu, Ruslan Salakhutdinov, and Louis-Philippe Morency. Hemm: Holistic evaluation of multimodal foundation models, 2024.

[44] V. Hinze-Hoare. The review and analysis of human computer interaction (hci) principles, 2007.

[45] I. Scott MacKenzie, Janet C. Read, and Matthew Horton. Empirical research methods for human-computer interaction, 2024.

[46] Philipp Jordan, Omar Mubin, Mohammad Obaid, and Paula Alexandra Silva. Exploring the referral and usage of science fiction in hci literature, 2018.

[47] Chantal Shaib, Millicent L. Li, Sebastian Joseph, Iain J. Marshall, Junyi Jessy Li, and Byron C. Wallace. Summarizing, simplifying, and synthesizing medical evidence using gpt-3 (with varying success), 2023.

[48] Daniel Sonntag and Hans-Jürgen Profitlich. An architecture of open-source tools to combine textual information extraction, faceted search and information visualisation, 2018.

[49] Anastasia Zhukova, Lukas von Sperl, Christian E. Matt, and Bela Gipp. Generative user-experience research for developing domain-specific natural language processing applications, 2024.

[50] Naama Rozen, Liat Bezalel, Gal Elidan, Amir Globerson, and Ella Daniel. Do llms have consistent values?, 2024.

[51] Zijie J. Wang, Alex Kale, Harsha Nori, Peter Stella, Mark E. Nunnally, Duen Horng Chau, Mihaela Vorvoreanu, Jennifer Wortman Vaughan, and Rich Caruana. Interpretability, then what? editing machine learning models to reflect human knowledge and values, 2022.

[52] Rob Procter, Peter Tolmie, and Mark Rouncefield. Holding ai to account: Challenges for the delivery of trustworthy ai in healthcare, 2022.

[53] Hao Wei, Jianing Qiu, Haibao Yu, and Wu Yuan. Medco: Medical education copilots based on a multi-agent framework, 2024.

[54] Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems*, 36:28541–28564, 2023.

[55] Huansheng Ning, Zhijie Guo, Raouf Hamzaoui, Rongyang Li, Fadi Farha, and Lingfeng Mao. Review of serious games for medical operation, 2022.

[56] Erva Nihan Kandemir, Jill-Jenn Vie, Adam Sanchez-Ayte, Olivier Palombi, and Franck Ramus. Adaptation of the multi-concept multivariate elo rating system to medical students training data, 2024.

[57] Tanja Svarre and Tony Russell-Rose. Think outside the search box: A comparative study of visual and form-based query builders, 2022.

[58] Chaoyun Zhang, Shilin He, Jiaxu Qian, Bowen Li, Liqun Li, Si Qin, Yu Kang, Minghua Ma, Guyue Liu, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, and Qi Zhang. Large language model-brained gui agents: A survey, 2025.

[59] Yuanhe Tian, Ruyi Gan, Yan Song, Jiaxing Zhang, and Yongdong Zhang. Chimed-gpt: A chinese medical large language model with full training regime and better alignment to human preferences. *arXiv preprint arXiv:2311.06025*, 2023.

[60] Jiajun Xu, Qun Wang, Yuhang Cao, Baitao Zeng, and Sicheng Liu. A general-purpose device for interaction with llms, 2024.

[61] Avishek Choudhury and Zaria Chaudhry. Large language models and user trust: Consequence of self-referential learning loop and the deskilling of healthcare professionals, 2024.

[62] Sanjay Oruganti, Sergei Nirenburg, Jesse English, and Marjorie McShane. Automating knowledge acquisition for content-centric cognitive agents using llms, 2023.

[63] John Stamper, Ruiwei Xiao, and Xinying Hou. Enhancing llm-based feedback: Insights from intelligent tutoring systems and the learning sciences. In *International Conference on Artificial Intelligence in Education*, pages 32–43. Springer, 2024.

[64] Cheonsu Jeong. A study on the implementation of generative ai services using an enterprise data-based llm application architecture. *arXiv preprint arXiv:2309.01105*, 2023.

[65] Francois Barnard, Marlize Van Sittert, and Sirisha Rambhatla. Self-diagnosis and large language models: A new front for medical misinformation, 2023.

[66] Manuel Rebol, Krzysztof Pietroszek, Claudia Ranniger, Colton Hood, Adam Rutenberg, Neal Sikka, and Christian Guetl. Collaborative system design of mixed reality communication for medical training, 2023.

[67] Malin Eiband, Daniel Buschek, and Heinrich Hussmann. How to support users in understanding intelligent systems? structuring the discussion, 2021.

[68] Asad Sayeed. Usability in the larger reality: A contrarian argument for the importance of social and political considerations, 2019.

[69] Claudia Bianchini, Fabrizio Borgia, and Maria de Marsico. A concrete example of inclusive design: deaf-oriented accessibility, 2019.

[70] Xueting Huang, Zhibo Zhang, Fusen Guo, Xianghao Wang, Kun Chi, and Kexin Wu. Research on older adults' interaction with e-health interface based on explainable artificial intelligence, 2024.

[71] Junia Coutinho Anacleto and Aparecido Fabiano Pinatti de Carvalho. Improving human-computer interaction by developing culture-sensitive applications based on common sense knowledge, 2010.

[72] Yingqing He, Zhaoyang Liu, Jingye Chen, Zeyue Tian, Hongyu Liu, Xiaowei Chi, Runtao Liu, Ruibin Yuan, Yazhou Xing, Wenhai Wang, Jifeng Dai, Yong Zhang, Wei Xue, Qifeng Liu, Yike Guo, and Qifeng Chen. Llms meet multimodal generation and editing: A survey, 2024.

[73] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.