Can Large Language Models Support Medical Facilitation Work? A Speculative Analysis

Najeeb Gambo Abdulhamid

Microsoft Africa Research Institute Kenya, nabdulhamid@microsoft.com

Millicent Ochieng

Microsoft Africa Research Institute Kenya, mochieng@microsoft.com

Kalika Bali

Microsoft Research India, kalikab@microsoft.com

Elizabeth A. Ankrah

University of California, Irvine, eankrah@uci.edu

Naveena Karusala

University of Washington, naveenak@uw.edu

Keshet Ronen

University of Washington, keshet@uw.edu

Jacki O'Neill

Microsoft Africa Research Institute Kenya, jacki.oneill@microsoft.com

Mobile messaging apps and SMS-based tools have been deployed to extend healthcare services beyond the clinic; peer support chat groups, consisting of patients and healthcare providers, can improve medication adherence. However, moderation can be burdensome for busy healthcare professionals who must respond to patients, provide accurate and timely information, and engage and build community among patients. In this paper, taking an ethnographic approach, we examine the moderation of chat groups for young people living with HIV in Kenya. We describe the roles and responsibilities of the moderator while striving to engage and build community among the participants and manage the group chat, highlighting the challenges they face. Grounded in the moderators' work, we explore how an LLM-enabled copilot could help or hinder group facilitation. In doing so, we contribute to discussions about the potential of Artificial Intelligence in supporting healthcare professionals.

CCS CONCEPTS • Human-centered computing • Human-computer interaction (HCI) • Empirical studies in HCI

Additional Keywords and Phrases: Large language models (LLMs), peer support chatgroups, ethnography, facilitators, roles and responsibility, AI copilot

ACM Reference Format:

1. INTRODUCTION

Healthcare services face global challenges, particularly in low-and-middle-income countries (LMICs), where health professional-to-population ratios are strained [22, 23, 42, 44, 70]. Mobile messaging and SMS-based services, including chat groups involving patients and medical professionals, offer promising solutions to reach and support patients [8, 9, 21, 43]. However, these groups place an additional burden on healthcare professionals. Thus, there is a need to explore the potential of artificial intelligence and natural language processing (NLP) solutions, such as ChatGPT, to streamline facilitation work [30, 31, 51].

This paper examines the interactions and activities of facilitators in a professionally supported peer chat group for young people living with HIV in Kenyan informal settlements. We employ an ethnographic approach to understand facilitators' work [6, 31, 51] and explore the potential of using large language modes (LLMs) as copilots in these groups. Additionally, we present a discussion on the benefits and risks of using LLMs as copilots for facilitators and participants [37, 38, 46]. Therefore, this paper extends prior research by providing a deeper understanding of the facilitator's role and assessing how language models might enhance their work [31, 40].

2. LITERATURE

We begin by discussing the medical group facilitation literature and then reviewing LLMs and Conversational Agents (CAs) in healthcare, focusing on professionally moderated peer chat groups.

2.1 The role of facilitators in a professionally moderated medical support group

Studies for understanding the role played by facilitators, administrators, and moderators in medical support groups have primarily focused on bulletin boards and platforms such as WebMD [28]. These studies encompass health topics, including cancer, fibromyalgia, suicide, and arthritis [28, 49, 50, 60]. They explore a range of factors contributing to facilitators' engagement, including motivations [60], benefits and strengths [3, 49], social support [50], participant experiences [13], and information exchange between participants and facilitators [28]. Studies have shed light on the intricate interplay between the benefits and challenges inherent to the role. For example, facilitator responsibilities involve recruiting participants, moderating discussions, providing clinical support, and ensuring compliance with established protocols [3, 28, 49, 50]. Facilitators face challenges, including unpredictability, limitations of written communication, balancing individual and collective needs, and interpreting comments lacking tone or non-verbal cues [3, 49, 50].

2.2 A Review of Large Language Models and Conversational Agents (CAs) in Healthcare

A substantial body of research in language technology explores applications relevant to healthcare, including text summarization, machine translation, question answering, and dialogue systems. These studies propose various models and techniques, such as Seq2Seq with copy mechanism [25], pointer-generator networks [33], transformers [65], BiDAF [29], and frameworks centered on end-to-end neural networks [11]. Additionally, research addresses ethical evaluation, mitigation, and design of Language Models (LLMs) and Conversational Agents (CAs) [16, 18, 33, 72].

Within the healthcare domain, research assesses the performance of new-generation LLMs, like ChatGPT, in areas such as medical question answering, symptom checking, and health coaching [2, 4, 10, 11, 57]. While ChatGPT shows fluency, it faces limitations in domain knowledge, hallucination risk, and ethical concerns [4, 12]. Guidelines for using LLMs and CAs in healthcare are also discussed [2, 55, 72]. Furthermore, studies explore the potential and risks of LLMs in healthcare, including information provision, decision support, and ethical, legal, and environmental issues [4, 7, 29, 66, 67, 72]. Some researchers highlight concerns about LLMs' social impact and call for more diverse perspectives in their development [7, 18]. Limited research has examined the use of new-generation LLMs in group facilitation, with one study involving ChatGPT in a mental health forum revealing positive participant ratings but discomfort due to the algorithm's involvement, emphasizing authenticity challenges [63]. This highlights the need for addressing accuracy, safety, privacy, fairness, and accountability when utilizing LLMs for facilitation support [39].

2.3 Professionally moderated peer support chat groups

Our paper builds on the findings of these two studies: one at an IVF clinic in China [30] and another involving HIV support groups in Kenya [31]. The IVF clinic study revealed that nurses moderating WeChat groups faced challenges in managing patient queries and workload, leading to design recommendations. The HIV support group study emphasized the facilitator's role in providing emotional and informational support, with suggestions to ease their workload. Both studies underscore the importance of professional moderators in chat groups and offer design insights [30, 31].

3 EMPIRICAL STUDY

Nairobi has many informal settlements, such as Kibera, where over 700,000 people live, and residents face poverty and poor living conditions [1, 32]. This paper analyses the chat data from two WhatsApp groups set up to offer peer support and improve medical adherence for young people living with HIV in Nairobi's informal settlements. The groups consisted of 55 participants, assigned to one of two chat groups aged 14-18 and 18-24 years. 38/55 were active in the chat groups. The groups had more females than males, all HIV-positive and receiving treatment. Both groups were moderated by the same facilitator, who had experience in HIV testing and counselling. The study lasted for six months.

3.1 Data sources

The data sources were the chat corpus and the transcript of an interview with the facilitator. The chat corpus consisted of the original messages in Kiswahili, English, and Sheng, often codemixed; their meta-data (time and date stamp); and English translations (by two local translators). The interview had questions and answers about the facilitator's role and experience.

3.2 Method

A broad ethnographic approach was employed in this research study as a tunnel to understanding the everyday realities of our participants [24, 52], including how they accomplish their activities and interactions [36, 58]. This approach enables researchers to understand the work undertaken in and through various situated actions and interactions observed in the world [15, 52, 53]. HCI and CSCW have a long history of ethnographic research [36], going back to the early work of Suchman [58] and the Lancaster group [27]. Online traces such as chat messages are a legitimate resource for ethnographic study as they reveal the situated interaction between participants in as much as it was taking place in and through the chat groups themselves. They represent traces of naturally occurring interaction and, as such, are amenable to ethnographic analysis. We chose this approach because of its utility in informing design [14, 26, 45] and in analysing a wide range of data, including the traces left by online interactions [20, 47]. Likewise, we follow the example of Martin et al. [36], who used this approach to study the traces in Turker Nation, an online crowdsourcing forum, to provide insights into members' work and challenges.

The first author analysed all the (translated) messages from both chat groups in detail. All authors discussed emergent themes and examined different segments of chat interaction in group sessions. Additionally, the first author analysed a transcript of an interview that a research team member conducted with the facilitator after the completion of the study. The 90-minute semi-structured interview focused on understanding the facilitator's role during the study period. This role included coordination, administration, interpersonal dynamics, invisible work, challenges, and teachable moments, among other things. Our initial objective was to understand how these chat groups operated and what work was done through the interactions. Based on this understanding, we describe the work of the facilitator below. This is then used as a starting point to interrogate the idea of an LLM-enabled copilot for the facilitator.

4 BEING A FACILITATOR

We first describe the nature of facilitation, followed by the challenges of group facilitation. Facilitation work was classified into two sets of interdependent activities: 1) administration work and 2) health and well-being work.

4.1 Administration work

Administration work refers to the work of managing and organising the chat groups and includes community management, clerical and communication, and housekeeping: 1) Community management involves setting up the groups, adding members, welcoming new members, etc. 2) Clerical communication involves summarising meeting minutes and sharing resolutions, posting essential communications, and sharing resources such as links and educational materials related to HIV/AIDS. 3) Housekeeping involves moderating group discussions, managing tensions and conflicts, and ensuring compliance with established protocols, for example, reminding members to protect their phones and avoid privacy breaches.

4.2 Health and well-being work

Karusala et al. [31] described how groups provided members with emotional support and practical information. In this work, we found that facilitators provided similar support to members.

Emotional support in the chat group involves empathetic responses, offering choices, and using comforting and encouraging language, sometimes even incorporating spiritual elements. When a member, P25 (14, M, G1), expressed a challenge with medication, the facilitator responded empathetically and referred P25 to the clinic while encouraging others to share similar experiences for collective solutions. The facilitator's primary focus is supporting members in need, either through group chat or private 1-on-1 chats. For instance, when P28 (17, M, G1) mentioned difficult situations, the facilitator offered a chance to discuss it privately to maintain sensitivity and ensure support for all members experiencing challenging life events. This approach balances creating an active and supportive group environment while respecting individual privacy and needs.

Information and education encompass the facilitator's essential role, involving the provision of accurate health advice, member query resolution, and encouragement of active participation. Key mechanisms include initiating discussions through weekly messages, revisiting unresolved questions to provide context, addressing member concerns, and soliciting community feedback before offering responses. Despite occasional delays due to chat flow, these efforts aim to inform, engage, and promote interaction among members.

4.3 The work to make chat groups work.

In the following sections, we describe how the work above is collaboratively achieved in and through the chat turns between the facilitator and the participants.

4.3.1 Timing and message design for engagement.

A crucial aspect of the facilitator's role is sending weekly messages, but participants may not always be online when they are sent. To encourage discussion, the facilitator adapts by following up when activity is observed. For instance, if the initial message sent at 8 a.m. receives no replies, the facilitator revisits the topic later, breaking it down into specific questions, often leading to engaging discussions. This approach involves rephrasing, timing follow-ups, and responding to participants' replies to make the message effective. Beyond weekly messages, the facilitator invests effort in keeping the groups active, occasionally using humor to prompt participation. For example, jokes like "I don't know if I should throw stones at you so that when you see my texts you reply" were employed. These tactics, including encouraging silent participants to join the discussion, contribute to group vitality.

4.3.2 Setting the tone.

The facilitator's key role in the chat groups is to create a friendly and supportive atmosphere by positively interacting with participants, motivating, appreciating, acknowledging, and encouraging group discussions. Participants who share updates or provide peer support receive recognition and gratitude from the facilitator. Group members actively contribute to discussions and often pose open-ended questions to promote conversation and debate. Occasionally, the conversation extends beyond HIV-related topics, with the facilitator offering practical advice, even on matters outside his usual purview, such as addressing unemployment among group members.

4.4 Challenges of Group Facilitation

The facilitator encountered a range of challenges while managing a medical chat group, including sustaining dynamic discussions during late-night message surges, encouraging active participation, coping with a high volume of message that led to response delays, addressing technical questions that required external consultations causing substantial hold-ups, and maintaining a balanced response tone to suit the diverse group demographics and experiences. These difficulties were evident in the struggle to keep discussions engaging despite crafting appealing messages. For example, while an initial member comment sparked lively interaction, subsequent questions from the facilitator on the same topic often received minimal responses. Furthermore, the facilitator's response times were affected by the sheer volume of messages, sometimes resulting in delays of up to 24 hours or more, prompting participants to seek urgent assistance and inquire about the facilitator's whereabouts.

5 POTENTIALS AND PERILS OF AN LLM-ENABLED COPILOT TO SUPPORT THE FACILITATOR

In this section, we present a discussion exploring the potential of Large Language Models (LLMs) to support facilitator’s work. Prior to the emergence of ChatGPT, NLP's use in medical interactions had garnered interest, including chatbots and aiding human patient support. Recent LLM iterations have excelled in mimicking human-like text and interaction, marking a promising development. Our paper, however, narrows its focus, employing empirical study to examine the specific benefits and challenges of integrating LLMs into facilitators' roles. These challenges include sustaining engagement, managing chat volume, and effectively communicating sensitive health information. Additionally, we acknowledge the language complexities within our chat corpus, encompassing code-mixed English, Swahili, and Sheng, replete with emojis, abbreviations, and misspellings. While preliminary qualitative tests indicate improved language understanding by LLMs, experimental validation is pending [40]. Nevertheless, we remain hopeful about LLMs' ability to handle such data, recognizing the need to identify and manage potential errors in future work.

5.1 Supporting existing work.

One way an LLM-enabled copilot could support the facilitator is to co-produce compelling content for his weekly messages. Co-production would involve an iterative model, where the facilitator uses the copilot to help produce engaging content likely to elicit responses from group members. One challenge now, however, is that LLMs cannot learn from the chat history, unless it is inserted into every prompt, which would be costly. What is needed is a copilot who can learn and remember what tactics are most likely to elicit responses, adapt over time, and even suggest topics the facilitator might focus on.

A second functionality that a copilot could bring is summarisation. LLMs are highly effective at summarising real-world text [59, 71]. The facilitator and the participants could benefit from summarisation, which could help the tension between the desirability and challenges of active groups. The facilitator sometimes finds the message volume overwhelming, especially when the group has been highly active during the night or when he is at work. A quick at-a-glance summary of the main topics and those involved could help him get an overview of what he has missed and potentially reduce the burden of work. Participants who spend time offline because of school, lack of data or sharing phones often struggle to catch up on what they have missed – trawling through the entire chat history is burdensome – to have a summary of the core topics discussed could help them catch up and enable them to dive deeper into topics of interest to them. Such summaries might be shared by push notification (e.g., by SMS) as a prompt for them to check out exciting discussions.

Additionally, the copilot might assist the facilitator in accelerating his research by recommending solutions to questions put forth by the participants. However, caution must be taken here to verify the suggested solutions, given the known tendency for LLMs to produce false responses (hallucinations). Nonetheless, with appropriate checks and balances in place, the copilot could help the facilitator co-produce responses and offer recommendations on the appropriate tone and style for conveying messages to participants, aiding facilitators in crafting effective communication. One caution, however, is that LLMs are trained on data vastly from the Global North – an issue referred to as the data divide [64]- so their ability to produce a human-like text which reflects the lived experience of the participants, even if it is in the language of the participants (Swahili/Sheng/English) is likely to be limited and would remain a core point of further investigation.

5.1.1 Additional services and functionality.

Copilot might also enable additional services of functionality to be provided at low overhead to the facilitator but benefit the group. The copilot could support the facilitator in assigning karma-like badges or scores to helpful or empathetic participants. Whilst gamification has often been suggested to support engagement, some suggest it acts as a motivator [17, 34, 56]; it must be carefully evaluated for appropriateness and impact in such a sensitive setting.

Another additional service is sentiment analysis. Many previous studies have suggested that this could be a desirable aid to facilitators in such groups [30, 31, 40], and these new LLMs take us a leap closer to achieving this to an elevated level of accuracy [61]. For example, a copilot could give the facilitator an overview of the day's sentiment, flag messages or discussions with negative sentiment for attention or even detect long-term declines in the mental well-being of certain participants by employing sentiment analysis of their posts [62]. This could allow the facilitator to promptly intervene and provide referrals to mental health counsellors, as necessary. By monitoring the emotional tone expressed in participants' contributions, the copilot could serve as a vigilant ally, supporting the holistic care of individuals within the group. However, for this functionality to be most effective, some historical "memory" would be invaluable; otherwise, as mentioned above, the entire chat history would need to be run as a prompt each time, which would be costly and inefficient.

5.2 Human-like vs. human language

In the previous section, we demonstrated the potential of employing a Large Language Model (LLM)-enabled copilot to support facilitation. We now turn to the potential limitations. Even though previous studies advocated designing out facilitators by automating their work, we think that is missing the point of what facilitators do. To this end, we described what and how facilitators work. One thing that stood out was how they provide emotional support. As we explained in Section 4, emotional support is not just about sending nice messages; it is also about following up, offering spiritual guidance, giving one-to-one support, and building personal relationships. It is dangerous to suggest that the human-like language produced by LLMs can replace this. Human-like is different from human, and human language is a means of expressing ourselves – from friendship to anger, from respect to care. In comparison, LLMs do not have anything to express. Interactions, even if they are task-based – much more so in a sensitive group like this - along with shared experiences, are what build relationships, and relationships are the glue that holds society together. If we remove those human interactions, what happens to relationships? LLMs can have no understanding of the context, the culture, or the feelings of people living with HIV/AIDS. A recent study investigated the effects of employing GPT-3 to co-create empathic messages for individuals seeking mental health support on a peer-to-peer platform [63]. This study found that messages composed of AI and supervised by humans received significantly higher ratings than those written solely by humans ($p < .001$), and response times were reduced by 50%. However, when individuals discovered that AI partially generated the messages, they experienced a sense of unease and emptiness. They did not feel genuinely understood or cared for. While AI can write helpful and supportive messages, people want real and genuine human interaction. They also want to know those to whom they are talking. This shows that using AI for emotional support needs useful design and honesty, and it cannot replace human empathy and connection.

As timeliness was a challenge for the facilitators and members, this might seem like a place where LLMs could help. Being online with the participants encourages active interaction, and one advantage of technology, like an LLM-based assistant, is that it can be (almost) always online. However, the purpose of these groups was not simply about providing information but about creating engagement and learning opportunities through the relationships between the group members and the members and the facilitators. Automated answers, unless they serve a clear purpose and are clearly marked as such, cannot and should not be considered to replace human interaction, caring and empathy.

It is facilitators' empathy, humour and care, as well as their knowledge, which helps these groups be successful. Indeed, the facilitator cannot be replaced by automatic AI services, but even how such services support him needs to be carefully considered.

5.3 Data Divide, Ethics, and Privacy

Many challenges still exist in supporting Large Language Models (LLMs) pertaining to the data divide, ethics, privacy, and security. First, challenges exist around the data on which the models are trained. The models are often trained on data from the Global North and English-speaking content found on the internet, which results in biased outputs [19, 48, 54], enlarging the data divide. Whilst GPT4 is becoming an impressive translator and linguist, it is still not inclusive of low-resource languages and other Indigenous languages such as Swahili or Sheng.[1]. Thus, the current attempt to use these models in multilingual chat rooms will result in erroneous outputs. For example, incorrect natural language generations, such as translations or summarising forum content, may lead to inaccurate responses by the facilitators or confusion among group members. By training LLMs in low-resource languages, these models can improve their performance and support multilingual contexts, thus bridging the data divide.

Additionally, challenges may exist concerning the ethics of using LLMs for professionally supported peer chat groups. For example, when left unchecked, LLMs can hallucinate, making up persuasive yet incorrect or inaccurate information, which can propagate misinformation and disinformation. Therefore, facilitators must maintain close supervision of their copilots and make final decisions and approvals before disseminating AI-generated text. While copiloting is a critical task we argue for, it is paramount that we minimise the invisible work facilitators must engage in. By training LLMs on internal, deidentified data that complies with health and privacy regulations and has been curated by health professionals reduces the ethical concerns associated with the use of LLMs in professionally supported peer chat groups. As the generated text, it is based on organizationally vetted information.

6 LIMITATION AND FUTURE WORK

This study has limitations to consider. It is based on a single case study involving two WhatsApp groups in Kenya, which may not be representative of other contexts, cultures, or populations. Participant characteristics, facilitator roles, and health domain dynamics can vary widely in different settings. Therefore, our analysis and design suggestions may not apply universally to other peer support chat groups or medical facilitation contexts. Future studies should explore how LLM-enabled chat groups can benefit patients across diverse health domains and cultural settings. Our paper serves as a starting point for further research and development rather than a definitive proposal.

7 CONCLUSIONS

Through trace ethnography, we analysed the multifaceted role of peer support chat group facilitators, shedding light on their administrative, emotional, and informational responsibilities. We also identified the challenges they encounter in maintaining these groups. As a contribution to the literature on mobile messaging and SMS-based medical support, we advocate for the adoption of LLM-Enabled copilots in peer-supported chat groups, outlining potential benefits and challenges. While advances in LLMs enhance their capacity to support group work, particularly in copiloting chat groups, addressing the complexities of sensitive medical and emotional contexts is paramount. Design principles, transparency, and genuine human empathy should guide AI-based solutions for facilitation support. Our paper serves as a thought-provoking exploration of how such services can be designed effectively.

---

[1] Sheng is the local language spoken in Kenya. It is a Swahili-English code-mixed language spoken in urban areas in Kenya, most common and popular with the youths, such as those who participated in this study.

REFERENCES

[1] Amuyunzu-Nyamongo, M. et al. 2007. Putting on a brave face: The experiences of women living with HIV and AIDS in informal settlements of Nairobi, Kenya. AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV. 19, SUPPL.1 (2007), 25–34. DOI:https://doi.org/10.1080/09540120601114618.

[2] Arora, A. and Arora, A. 2023. The promise of large language models in health care. The Lancet. 401, 10377 (Feb. 2023), 641. DOI:https://doi.org/10.1016/S0140-6736(23)00216-7.

[3] Atanasova, S. et al. 2017. Exploring the benefits and challenges of health professionals' participation in online health communities: Emergence of (dis)empowerment processes and outcomes. International Journal of Medical Informatics. 98, (Feb. 2017), 13–21. DOI:https://doi.org/10.1016/j.ijmedinf.2016.11.005.

[4] Au Yeung, J. et al. 2023. AI chatbots not yet ready for clinical use. Frontiers in Digital Health. 5, (Apr. 2023), 1161098. DOI:https://doi.org/10.3389/FDGTH.2023.1161098.

[5] Ayers, J.W. et al. 2023. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Internal Medicine. 183, 6 (Jun. 2023), 589–596. DOI:https://doi.org/10.1001/JAMAINTERNMED.2023.1838.

[6] Beaglehole, R. et al. 2008. Improving the prevention and management of chronic disease in low-income and middle-income countries: a priority for primary health care. Elsevier. 372, 9642 (2008), 940–949.

[7] Bender, E.M. et al. 2021. On the dangers of stochastic parrots: Can language models be too big? FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. (Mar. 2021), 610–623. DOI:https://doi.org/10.1145/3442188.3445922.

[8] Boulos, M.K. et al. 2016. Instagram and WhatsApp in health and healthcare: An overview. Future Internet. 8, 3 (2016), 37. DOI:https://doi.org/10.3390/fi8030037.

[9] Calleja-Castillo, J.M. and Gonzalez-Calderon, G. 2018. WhatsApp in stroke systems: Current use and regulatory concerns. Frontiers in Neurology. 9, MAY (May 2018). DOI:https://doi.org/10.3389/FNEUR.2018.00388/FULL.

[10] Callejas, Z. and Griol, D. 2021. Conversational Agents for Mental Health and Wellbeing. Logic, Argumentation and Reasoning. 22, (2021), 219–244. DOI:https://doi.org/10.1007/978-3-030-61438-6_11/COVER.

[11] Cascella, M. et al. 2023. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. Journal of Medical Systems. 47, 1 (Dec. 2023), 1–5. DOI:https://doi.org/10.1007/S10916-023-01925-4/TABLES/2.

[12] Chan, A. 2022. GPT-3 and InstructGPT: technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry. AI and Ethics 2022 3:1. 3, 1 (Apr. 2022), 53–64. DOI:https://doi.org/10.1007/S43681-022-00148-6.

[13] Coulson, N.S. and Shaw, R.L. 2013. Nurturing health-related online support groups: Exploring the experiences of patient moderators. Computers in Human Behaviour. (2013), 1695–1701.

[14] Crabtree, A. et al. 2012. Doing design ethnography. Springer Science & Business Media.

[15] Crabtree, A. et al. 2000. Ethnomethodologically Informed Ethnography and Information System Design. Journal of the American Society for Information Science. 51, 7 (2000), 666–682.

[16] Curto, G. et al. 2022. Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. AI & SOCIETY. 1, (2022), 3. DOI:https://doi.org/10.1007/s00146-022-01494-z.

[17] Facey-Shaw, L. et al. 2020. Do Badges Affect Intrinsic Motivation in Introductory Programming Students? Simulation and Gaming. 51, 1 (Feb. 2020), 33–54. DOI:https://doi.org/10.1177/1046878119884996/ASSET/IMAGES/LARGE/10.1177_1046878119884996-FIG3.JPEG.

[18] Fiske, A. et al. 2019. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. J Med Internet Res 2019;21(5):e13216 https://www.jmir.org/2019/5/e13216. 21, 5 (May 2019), e13216. DOI:https://doi.org/10.2196/13216.

[19] Gadiraju, V. et al. 2023. "I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. 2023 ACM Conference on Fairness, Accountability, and Transparency. (Jun. 2023), 205–216. DOI:https://doi.org/10.1145/3593013.3593989.

[20] Gail, J. 1992. Harvey Sacks Lectures on Conversation (2 vols) . Oxford: Basil Blackwell.

[21] Ganasegeran, K. et al. 2017. The m-Health revolution: Exploring perceived benefits of WhatsApp use in clinical practice. Elsevier. 97, (2017), 145–151.

[22] H Alderwick 2022. Is the NHS overwhelmed? bmj. (2022), 376.

[23] H Salisbury 2020. Helen Salisbury: GPs are stretched too thin. bmj. (2020), 368.

[24] Harper, R.H.R. 2000. The organisation in ethnography–a discussion of ethnographic fieldwork programs in CSCW. Computer Supported Cooperative Work (CSCW). 9, (2000), 239–264.

[25] Harrer, S. 2023. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine Personal View. www.thelancet.com. 90, (2023).

[26] Hughes, J. et al. 1994. Moving out from the control room: Ethnography in system design. Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW 1994. (Oct. 1994), 429–439. DOI:https://doi.org/10.1145/192844.193065.

[27] Hughes, J.A. et al. 1992. Faltering from ethnography to design. Proceedings of the Conference on Computer-Supported Cooperative Work. (1992), 115–122. DOI:https://doi.org/10.1145/143457.143469.

[28] Huh, J. et al. 2013. Patient Moderator Interaction in Online Health Communities. AMIA Annual Symposium Proceedings (Vol. 2013) (2013), 627.

[29] Jovanovic, M. et al. 2020. Chatbots as Conversational Healthcare Services. IEEE Internet Computing. 25, 3 (May 2020), 44–51. DOI:https://doi.org/10.1109/MIC.2020.3037151.

[30] Karusala, N. et al. 2020. Making Chat at Home in the Hospital: Exploring Chat Use by Nurses. Conference on Human Factors in Computing Systems - Proceedings (Apr. 2020).

[31] Karusala, N. et al. 2021. "That courage to encourage": Participation and Aspirations in Chat-based Peer Support for Youth Living with HIV. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (2021), 1–17.

[32] Kenya: Nairobi Urban Profile | UN-Habitat: 2005. https://unhabitat.org/kenya-nairobi-urban-profile. Accessed: 2022-05-30.

[33] Kidwai, B. and Rk, N. 2020. Design and Development of Diagnostic Chabot for supporting Primary Health Care Systems. Procedia Computer Science.

167, (Jan. 2020), 75–84. DOI:https://doi.org/10.1016/J.PROCS.2020.03.184.

[34] Kyewski, E. and Krämer, N.C. 2018. To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. Computers & Education. 118, (Mar. 2018), 25–37. DOI:https://doi.org/10.1016/J.COMPEDU.2017.11.006.

[35] Lee, P. et al. 2023. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. The New England journal of medicine. 388, 13 (Mar. 2023), 1233–1239. DOI:https://doi.org/10.1056/NEJMSR2214184/SUPPL_FILE/NEJMSR2214184_DISCLOSURES.PDF.

[36] Martin, D. et al. 2014. Being a turker. Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW (2014), 224–235.

[37] Mehra, A. et al. 2017. Prayana: A Journey Towards Financial Inclusion.

[38] Mehra, A. et al. 2018. Prayana: Intermediated Financial Management in Resource-Constrained Settings. Conference on Human Factors in Computing Systems - Proceedings. 2018-April, (Apr. 2018). DOI:https://doi.org/10.1145/3173574.3173963.

[39] Mittelstadt, B. 2019. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence. 1, 11 (Nov. 2019), 501–507. DOI:https://doi.org/10.1038/S42256-019-0114-4.

[40] Mondal, I. et al. 2021. A Linguistic Annotation Framework to Study Interactions in Multilingual Healthcare Conversational Forums. In Proceedings of The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop (2021), 66–77.

[41] Mwongela, S. et al. 2023. Data-Efficient Learning for Healthcare Queries In Low-Resource And Code Mixed Language Settings. International Conference on Learning Representations. (2023).

[42] N Triggle 2015. Staff shortage puts neonatal services under strain. Nursing Children and Young People (2014+). 27, 9 (2015), 7.

[43] Nardo, B. et al. 2016. Optimizing Patient Surgical Management Using WhatsApp Application in the Italian Healthcare System. Telemedicine and e-Health. 22, 9 (Sep. 2016), 718–725. DOI:https://doi.org/10.1089/tmj.2015.0219.

[44] O'connell, T. et al. 2008. Health services under siege: the case for clinical process redesign. Medical Journal of Australia . 188, 6 (2008), 9–13.

[45] O'neill, J. et al. 2011. From ethnographic study to mixed reality: a remote collaborative troubleshooting system. Proceedings of the ACM 2011 conference on Computer supported cooperative work. (2011), 225–234.

[46] O'Neill, J. et al. 2017. Working Digital Money into a Cash Economy: The Collaborative Work of Loan Payment. Journal of Computer Supported Cooperative Work (CSCW). 26, 4 (Jul. 2017), 733–768. DOI:https://doi.org/10.1007/S10606-017-9289-6.

[47] O'Neill, J. and David, M. 2003. Text chat in action. Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work (2003), 40–49.

[48] Ovalle, A. et al. 2023. "I'm fully who I am": Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation. 2023 ACM Conference on Fairness, Accountability, and Transparency. 23 (May 2023), 1246–1266. DOI:https://doi.org/10.1145/3593013.3594078.

[49] Owen, J.E. et al. 2009. Benefits and challenges experienced by professional facilitators of online support groups for cancer survivors. Psycho-Oncology. 18, 2 (2009), 144–155. DOI:https://doi.org/10.1002/pon.1374.

[50] Perry, A. et al. 2022. "I Want to Be Stepping in More" - Professional Online Forum Moderators' Experiences of Supporting Individuals in a Suicide Crisis. Frontiers in Psychiatry. 13, (Jun. 2022). DOI:https://doi.org/10.3389/fpsyt.2022.863509.

[51] Pruitt, S. and JAE Epping-Jordan 2005. Preparing the 21st century global healthcare workforce. Bmj. 330, 7492 (2005), 637–639.

[52] Randall, D. et al. 2021. Ethnography, CSCW and Ethnomethodology. Computer Supported Cooperative Work (CSCW) . 30, (2021), 189–214. DOI:https://doi.org/10.1007/s10606-020-09388-8.

[53] Rouncefield, M. 2011. Fieldwork, Ethnography and Ethnomethodology. LSCITS Socio-Technical Systems Engineering Handbook. University of St Andrews. 44–48.

[54] Salewski, L. et al. 2023. In-Context Impersonation Reveals Large Language Models' Strengths and Biases. (May 2023).

[55] Sallam, M. 2023. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare. 11, 6 (2023), 887.

[56] Shin, M. and Bolkan, S. 2020. Intellectually stimulating students' intrinsic motivation: the mediating influence of student engagement, self-efficacy, and student academic support. https://doi.org/10.1080/03634523.2020.1828959. 70, 2 (2020), 146–164. DOI:https://doi.org/10.1080/03634523.2020.1828959.

[57] Singhal, K. et al. 2022. Large Language Models Encode Clinical Knowledge. (2022).

[58] Suchman, L. 1987. Plans and situated actions: The problem of human-machine communication. Cambridge: Cambridge University Press.

[59] Tam, D. et al. 2022. Evaluating the Factual Consistency of Large Language Models Through Summarization. (Nov. 2022).

[60] van Uden-Kraan, C.F. et al. 2010. Patient-initiated online support groups: Motives for initiation, extent of success and success factors. Journal of Telemedicine and Telecare. 16, 1 (Jan. 2010), 30–34. DOI:https://doi.org/10.1258/jtt.2009.001009.

[61] Wang, Z. et al. 2023. Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study. (2023).

[62] Wang, Z. et al. 2023. Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study. (Apr. 2023).

[63] We provided mental health support to about 4,000 people — using GPT-3. Here's what happened？ 2023. https://twitter.com/RobertRMorris/status/1611450197707464706. Accessed: 2023-01-06.

[64] Were, A. 2023. Ethical AI Needed Now More Than Ever: A view on the intersection between AI and economic development.

[65] WHO calls for safe and ethical AI for health: 2023. https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health. Accessed: 2023-05-16.

[66] Wilson, L. and Marasoiu, M. 2022. The Development and Use of Chatbots in Public Health: Scoping Review. JMIR human factors. 9, 4 (Oct. 2022). DOI:https://doi.org/10.2196/35882.

[67] Xu, L. et al. 2021. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. JMIR cancer. 7, 4 (Oct. 2021). DOI:https://doi.org/10.2196/27850.

[68] Xu, W. 2023. AI in HCI Design and User Experience. (Jan. 2023).

[69] Yap, C.E.L. et al. 2021. How HCI Interprets Service Design: A Systematic Literature Review. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 12933 LNCS, (2021), 259–280. DOI:https://doi.org/10.1007/978-3-030-85616-8_16/TABLES/3.

[70] Zarocostas, J. 2011. Libyan health system is "absolutely stretched," says UN. BMJ: British Medical Journal (Online). 343, (2011).

DOI:https://doi.org/10.1136/BMJ.D4326.

[71] Zhang, T. et al. 2023. Benchmarking Large Language Models for News Summarization. (Jan. 2023).

[72] Zheng, Q. et al. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. Conference on Human Factors in Computing Systems - Proceedings. (Apr. 2022). DOI:https://doi.org/10.1145/3491102.3501855.