EPiC Series in Computing
Volume 102, 2024, Pages 81–89
Proceedings of Conference on Cognitive and Computational Aspects of Situation Management 2023

# Clinician-AI Collaboration for Decision Support in Telemedicine: A Randomized Controlled Trial Study

Ryan Ebnali Harari[1,2], Nima Ahmadi[3], Shiva Pourfalatoun[4], Abdullah Al-Taweel[5], Hamid Shokoohi[1, 5]
[1] Harvard Medical School, Boston, USA
[2] STRATUS Medical Simulation Center, Mass General Brigham, Boston, USA
[3] Department of System Engineering Rensselaer Polytechnic Institute, Troy, USA
[4] Department of System Engineering, Colorado State University, Fort Collins, USA
[5] Emergency Ultrasound, Mass General Brigham, Boston, USA
rhrari@bwh.harvard.edu

**Abstract**

AI systems, particularly Large Language Models (LLMs), have the potential to improve telemedicine. However, there is a need to further investigate the effectiveness of AI decision support and clinician-AI collaboration in this context. This study examines the impact of AI-only and clinician-AI support systems on trust, acceptance, usability, and cognitive load in telemedicine scenarios. In a randomized controlled study, twenty non-medical participants were randomly assigned to receive support from an AI-only or clinician-AI decision support system during cardiopulmonary resuscitation (CPR) scenarios simulated in an augmented reality (AR) headset. We used ChatGPT 3, a widely used LLM, as the AI system. Participants' responses were measured using trust, acceptance, and usability questionnaires, as well as a wearable wristband to collect physiological data. The results show that the clinician-AI scenario was perceived as more useful compared to the ChatGPT-only scenario. The collaborative approach also led to higher heart rate variability (HRV) and lower LF/HF ratio, indicating potentially lower mental effort compared to ChatGPT-only. No significant differences were found in system usability scale (SUS) and electrodermal activity (EDA) levels between the scenarios. These findings highlight the importance of involving clinicians in AI-supported telemedicine. Further research should explore real-world applications to validate the preliminary results.

K. Baclawski, M. Kozak, K. Bellman, G. D'Aniello, A. Ruvinsky and C. Da Silva Ferreira Barreto (eds.), CogSIMA 2023 (EPiC Series in Computing, vol. 102), pp. 81–89

# 1 Introduction

Telemedicine has emerged as a prominent remote healthcare delivery method with numerous benefits, including improved accessibility, convenience, and cost-effectiveness, particularly in situations where in-person consultations are challenging or not feasible (Frankenstein and Smurra 2022; Khonji 2020). As telemedicine continues to evolve, the integration of AI holds promises to enhance the telemedicine experience and address some complexities of the modern healthcare landscape (Tavares et al. 2023; Nayak et al. 2021). AI systems such as LLMs may have the capacity to revolutionize telemedicine practices by providing advanced language-related capabilities, such as question-answering, translation, and text summarization (Yeung et al. 2023). The ability of ChatGPT as a widely used LLM to generate coherent and contextually relevant responses using large textual data presents an opportunity for improving clinical support in telemedicine encounters (Rao et al. 2023; Sallam 2023).

For successful integration of AI systems in telemedicine, however, it is imperative to understand the intricacies of human trust and acceptance of AI-based clinical support technology (Hwang et al., 2019; Davis, 1989), as LLMs come with their inherent limitations such as occasional inaccuracies, nonsensical content generation, and the potential for presenting misinformation as factual information. The accuracy of LLM-generated answers is intricately linked to the quality, quantity, and nature of the data used in their training. Despite the growing prominence of AI systems in medicine, research specifically examining the acceptance and trust to LLM-powered technologies within telemedicine settings remains limited (Gagnon et al., 2012).

Moreover, most of the previous studies focus on clinical support provided by either human clinicians or AI systems alone, and there has been limited investigation into the potential of collaboration between human clinicians and AI systems (Lai, Kankanhalli, and Ong 2021; Fischer et al. 2023). As trust in AI-generated suggestions remains a challenge (Wang et al., 2020), it is crucial to examine how AI systems supervised by clinicians can affect trust and acceptance in telemedicine clinical support (Alviani et al. 2023; Orrange et al. 2020; Davenport and Kalakota, 2019).

The objective of this study is to evaluate the usability and physiological responses of participants to an AI-clinician support system during telemedicine emergency scenarios. The significance of this research lies in its contribution to the existing body of knowledge on AI adoption in healthcare and its potential to inform the development of guidelines and best practices for telemedicine implementation.

# 2 Methods

## 2.1 Participants

A total of 20 undergraduate and graduate students, with no medical background or prior CPR experience, were recruited. Participants were randomly assigned to (a) ChatGPT-only or (b) clinician-ChatGPT collaboration groups in a balanced randomized setup (Figure 1). Before signing the informed consent form, all participants received study information and underwent a screening process to ensure they met the eligibility criteria.

## 2.2 Telemedicine Scenarios

We simulated a CPR scenario involving a virtual patient experiencing cardiac arrest. This scenario was presented to participants in a virtual environment using a Microsoft HoloLense 2 augmented reality (AR) headset, providing a 3D and immersive experience. The CPR scenario was rendered at a smooth frame rate of 30 frames per second (30 fps). The experimental room was well-illuminated, for optimal visibility for participants (Figure 2).

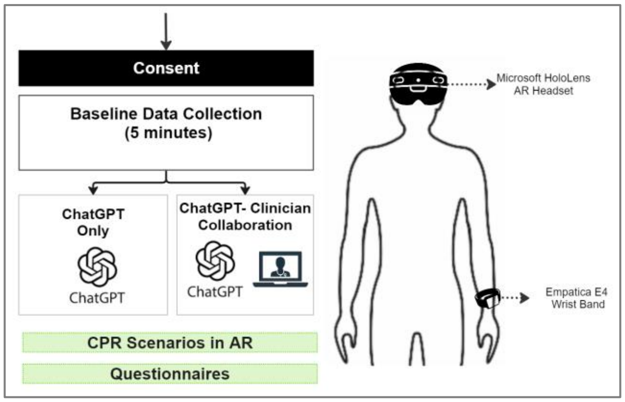


**Figure 1:** Study design and procedure used for the experiment including consent, baseline data collection for physiological data, and the scenario.

Participants were assigned to either the ChatGPT-only group or the clinician-ChatGPT collaboration group in a balanced manner (Figure 1). The ChatGPT-only group was provided with a Dell 15-inch laptop placed on a nearby table, allowing them to access the ChatGPT website and ask questions. The clinician-ChatGPT group had an additional component to their experience; they participated in a one-minute audio call with a remote clinician who had access to the conversation between the participant and ChatGPT. This setup was designed to simulate situations where direct access to healthcare professionals is limited, enabling the participants to receive guidance and support from a remote clinician through collaboration with ChatGPT. Each participant was given ten minutes (based on our pilot testing) to complete the CPR scenario, which allowed sufficient time to engage with the tasks, follow CPR protocols, and make decisions based on the options that were provided in the AR display. The focus of our study was on evaluating participants' trust and acceptance of the technology, and clinical outcomes were not measured as part of our research objectives because of difficulty of performing CPR on the AR environment.

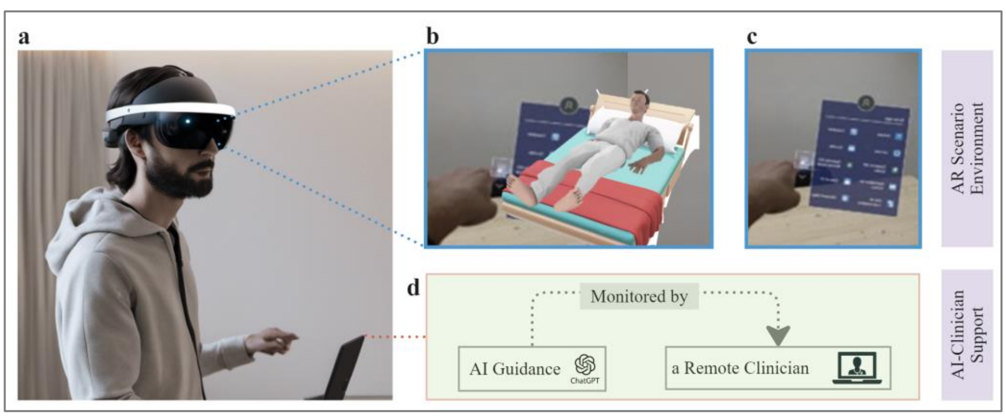


**Figure 2:** a) A participant during a CPR scenario wearing an AR headset, b) the CPR scenario projected in the AR headset, c) the participant interacting with the AR UIs to choose actions to be performed during the CPR scenario, d) type of clinical support provided to the participant during the scenario.

## 2.3 Data Collection

Our study used multiple instruments and physiological measurements to gather data on participants' trust, and technology acceptance as primary outcomes and system usability and cognitive load levels as secondary outcomes. We employed the following measures:

Participants' trust in the AI system was assessed using a trust scale based on a Likert scale of 0-5 (0 representing no trust and 5 representing full trust). Technology Acceptance was evaluated using the Technology Acceptance Model (TAM), which included items related to perceived usefulness, perceived ease of use, attitude towards use, and behavioral intention to use the AI system. System usability was examined using the System Usability Scale (SUS), a reliable and widely used questionnaire for measuring usability in various domains, including technology and software (Brooke, 1996). To assess cognitive load levels, we collected physiological measurements, including HRV, low-frequency to high-frequency ratio (LF/HF), and Electrodermal Activity (EDA) based on data extracted using an Empatica device (Figure 1).

## 2.4 Data Analysis:

The normality of the data was assessed using the Shapiro-Wilk test. As the technology acceptance variables were not normally distributed, we employed the Mann-Whitney U test to compare groups. For the remaining variables, we used the t-test to compare differences between the ChatGPT and clinician-ChatGPT collaboration scenarios. A significance level of $p < .05$ was applied to all analyses. Python 3.3 and relevant statistical packages were utilized in Jupyter Notebook for pre-processing and data analysis.

# 3 Results:

## 3.1 Demographic data:

Among the participants, 70% were undergraduates, with an age range of 18 to 30 years (mean = 21.5, SD = 2.5), and 60% were female. Approximately 10% of the participants had prior experience with AR hardware or software. None of the participants reported experiences with CPR nor had prior medical training and experiences.

## 3.2 Trust and Technology Acceptance:

*Trust*: There was a significant difference in the trust scores between the clinician-ChatGPT collaboration scenario (mean = 4.3, SD=0.46) and the ChatGPT-only scenario (median = 3.8, SD=0.74) ($p = .027$). *Perceived Usefulness*: The median score for perceived usefulness was significantly higher in the clinician-ChatGPT collaboration scenario (5.6, Q1 = 5.0, Q3 = 6.2) compared to the ChatGPT-only scenario (4.3, Q1 = 3.8, Q3 = 4.8) ($p < .001$). *Perceived Ease of Use:* There was no significant difference in the median scores for perceived ease of use between the clinician-ChatGPT collaboration scenario (4.8, Q1 = 4.3, Q3 = 5.2) and the ChatGPT-only scenario (4.6, Q1 = 4.1, Q3 = 5.0) ($p > .05$). *Attitude Towards Use:* The median score for attitude towards use was significantly more positive in the clinician-ChatGPT collaboration scenario (6.2, Q1 = 5.7, Q3 = 6.6) compared to the ChatGPT-only scenario (5.1, Q1 = 4.7, Q3 = 5.4) ($p < .001$). *Behavioral Intention to Use:* The median score for behavioral intention to use was significantly higher in the clinician-ChatGPT collaboration scenario (5.7, Q1 = 5.3, Q3 = 6.2) compared to the ChatGPT-only scenario (4.5, Q1 = 4.1, Q3 = 4.8) ($p < .01$) (Figure 3).

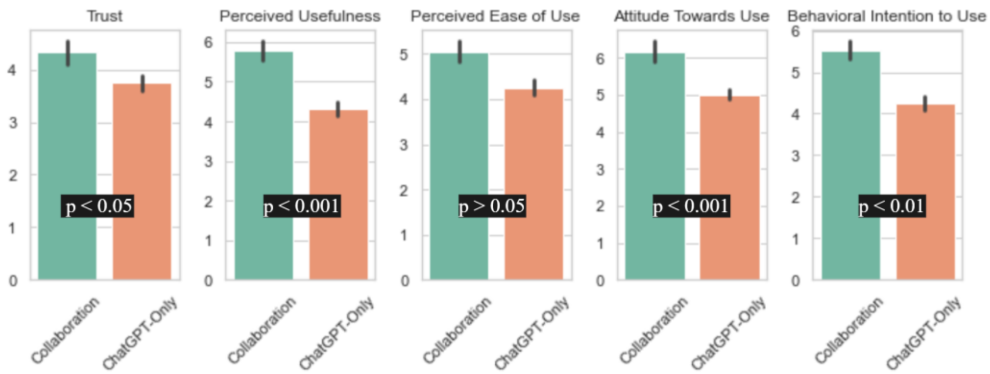


**Figure 3:** the result of comparing trust, perceived usefulness, perceived ease of use, attitude towards use, and behavioral intention to use between ChatGPT-only and Collaboration (clinician-ChatGPT) support during the CPR scenario.

## 3.3 System Usability Scale:

There was no significant difference in the SUS scores between the clinician-ChatGPT collaboration scenario (mean = 75.2, SD=10.41) and the ChatGPT-only scenario (median = 74.8, SD=8.94)($p = .827$).

## 3.4 Cognitive Load:

*HRV*: Participants exhibited significantly higher HRV in the clinician-ChatGPT collaboration scenario (mean = 76.2ms, SD=11.52ms) compared to the ChatGPT-only scenario (median = 68.7ms, SD=8.21ms) ($p < .05$). *LF/HF ratio*: The LF/HF ratio was significantly lower in the clinician-ChatGPT collaboration scenario (mean = 1.25, SD=0.54) compared to the ChatGPT-only scenario (median = 1.87, SD=0.93) ($p < .01$). *EDA*: There was no significant difference in EDA levels between the clinician-ChatGPT collaboration scenario (mean = 4.3 µS, SD=1.09 µS) and the ChatGPT-only scenario (mean = 4.4 µS, SD=1.35 µS) ($p= 1.49$).

# 4 Discussion:

This study examined trust, acceptance, usability, and physiological metrics associated with two clinical guidance interventions (ChatGPT-only or clinician-ChatGPT) in a telemedicine practice performing CPR scenarios using a randomized controlled design. The study highlighted higher trust and acceptance in using ChatGPT as instruction coupled with access to a clinician. This finding is aligned with prior studies emphasizing the importance of human expertise in trust and acceptance of AI integration (Choung, David, and Ross 2022; Ebnali et al. 2019; Ebnali et al. 2020; Tocchetti and Brambilla 2022; Knop et al. 2022).

The results of the TAM showed that the clinician-ChatGPT scenario was associated with higher scores for perceived usefulness when compared to the ChatGPT-only. This suggests that limited access (1 minute) to a clinician's guidance enhanced participants' perception of the AI system's usefulness in providing CPR with ChatGPT instructions. One explanation for this result could be that the inclusion of two verification methods, including asking experts as one of the validation methods, contributes to the validity of the data and its usefulness. This finding is supported by other studies indicating that crowdsourced data contributed by experts have higher quality than data contributed by non-experts, suggesting that involving experts in the validation process can improve the quality and usefulness of the data (See et al. 2013; Rovinelli and Hambleton 1976).

No significant difference was observed in the scores for perceived ease of use, indicating that both scenarios were similarly perceived in terms of their ease of use. Since approximately 10% of participants had prior experience with augmented reality (AR), and the limited amount of time each participant had to complete the task, it is possible that participants were more influenced by the novelty of the scenario and concept rather than by the addition of a clinician's opinion to their experiment when determining their perceived ease of use. This could lead to both groups experiencing these scenarios equally challenging. This finding is supported by Palma et al. suggesting that information overload is a major contributor to users' performance when completing a task (Palma, Seeger, and Heinzl 2019).

The participants' attitudes to using the AI system were more positive in the clinician-ChatGPT scenario when compared to the ChatGPT-only. This may in part be due to the fact that the presence of a clinician in the collaboration reflects a sense of reliability and real-time affirmation to use the ChatGPT for CPR instruction. Considering potential drawbacks associated with the use of AI systems, including legal and ethical concerns (Bozorgi 2023; Kooli and Muftah 2022), it is possible that the involvement of a clinician may have mitigated such issues resulting in a more positive intention for those participants in the clinician-ChatGPT scenario. This finding is consistent with other recent case studies, which emphasized the necessity of involving clinicians in AI development and implementation for healthcare to ensure systems are effective, trustworthy, safe, and compatible with ethical principles (Brown et al. 2023; Strohm et al. 2020). It is important to note that the SUS scores did not significantly differ between the two scenarios, suggesting that both the clinician-ChatGPT and ChatGPT-only scenarios were perceived similarly regarding the system usability. This indicates that participants found both scenarios equally user-friendly and efficient.

The participants exhibited higher HRV in the clinician-ChatGPT scenario compared to the ChatGPT-only. This suggests that the collaboration scenario may have resulted in a lower cognitive load among participants. The lower LF/HF ratio in the clinician-ChatGPT scenario indicates a shift towards a more parasympathetic state, which is associated with relaxation and reduced mental effort. In line with previous findings, experts' opinions incorporated into AI systems may facilitate the decision-making process (Zhang, Liao, and Bellamy 2020), leading to a lower LF/HF ratio. In addition, experts' decisions are less influenced by time-related stress, suggesting that expertise can compensate for cognitive demands imposed by stress (Morrow et al. 2004). However, no significant difference was found in EDA levels between the two scenarios, indicating similar levels of physiological excitement.

The findings suggest that the presence of a clinician in the ChatGPT-based telemedicine scenarios positively influenced participants' perception of the system's usefulness, attitude towards use, behavioral intentions, and cognitive load. These results highlight the potential of clinician-AI collaboration in enhancing the acceptance and effectiveness of AI systems in telemedicine support.

This study has some limitations to be considered. The sample consisted of undergraduate and graduate students, with a small number of participants which may limit the generalizability of the findings to a broader population. However, this study aims to provide preliminary insights into the trust, acceptance, and usability of AI-enhanced telemedicine support. Additionally, the study focused on simulated scenarios, and clinical outcomes were not evaluated. Therefore, further research on evaluating the performance of the ChatGPT system and clinician-ChatGPT collaboration in real-world clinical settings is warranted. Another limitation of the study is the use of AR to simulate telemedicine scenarios. While AR can provide a realistic visual representation of the interaction in human-human and human-system communication (Burian et al. 2023; Ebnali et al. 2023; Beadle et al. 2020; Ebnali et al, 2022), it may not fully replicate the actual experience of telemedicine.

# 5 Conclusion:

This study demonstrated the potential benefits of integrating AI technology, such as ChatGPT, with clinician collaboration in telemedicine scenarios. It appears that receiving clinical support from a clinician-AI system can improve trust and acceptance and may reduce cognitive load compared to receiving telemedicine support from an AI (ChatGPT)-only system. Future research on clinical scenarios and other potential factors that affect the adoption and implementation of AI-based decision support systems is needed.

# 6 References

Alviani, Rima, Betty Purwandari, Imairi Eitiveni, and Mardiana Purwaningsih. 2023. “Factors Affecting Adoption of Telemedicine for Virtual Healthcare Services in Indonesia.” Journal of Information Systems Engineering and Business Intelligence 9 (1): 47–69.

Beadle, S., Spain, R., Goldberg, B., Ebnali, P. M., Bailey, S. K. T., Ciccone, B., Wilson, R., Rubensky, S., Carroll, M., Bennett, W., Hu, X., Reiner, A. J., Vasquez, H., Wooldridge, A. R., Lee, L., Verma, P., Jeong, H., Gisick, L. M., French, J., & Keebler, J. R. (2020). Virtual reality, augmented reality, and virtual environments: demonstrations of current technologies and future directions. Proceedings of the Human Factors and Ergonomics Society . . . Annual Meeting, 64(1), 2119–2123. https://doi.org/10.1177/1071181320641514

Bozorgi, Majedeh. 2023. “The Impact of Procedural Law on Artificial Intelligence by Improving the Healthcare Systems.” Advances in Healthcare Information Systems and Administration, 172–92.

Brooke, John. 1996. “SUS: A ‘Quick and Dirty’ Usability Scale.” In Usability Evaluation In Industry, 1st Edition, 207–12. CRC Press.

Brown, Chris, Rayiz Nazeer, Austin Gibbs, Pierre Le Page, and Andrew R. J. Mitchell. 2023. “Breaking Bias: The Role of Artificial Intelligence in Improving Clinical Decision-Making.” Cureus. https://doi.org/10.7759/cureus.36415.

Burian, B. K., Ebnali, M., Robertson, J., Musson, D., Pozner, C. N., Doyle, T. E., Smink, D. S., Miccile, C., Paladugu, P., Atamna, B., Lipsitz, S., Yule, S., & Dias, R. D. (2023). Using extended reality (XR) for medical training and real-time clinical support during deep space missions. Applied Ergonomics, 106, 103902. https://doi.org/10.1016/j.apergo.2022.103902

Choung, Hyesun, Prabu David, and Arun Ross. 2022. “Trust in AI and Its Role in the Acceptance of AI Technologies.” International Journal of Human–Computer Interaction, April. https://doi.org/10.1080/10447318.2022.2050543.

Ebnali, M., Hulme, K. F., Ebnali-Heidari, A., & Mazloumi, A. (2019). How does training effect users’ attitudes and skills needed for highly automated driving? Transportation Research Part F-traffic Psychology and Behaviour, 66, 184–195. https://doi.org/10.1016/j.trf.2019.09.001

Ebnali, M., Fathi, R., Lamb, R., Pourfalatoun, S., & Motamedi, S. (2020, April). Using Augmented Holographic UIs to Communicate Automation Reliability in Partially Automated Driving. In AutomationXP@ CHI.

Ebnali, M., Paladugu, P., Miccile, C., Park, S. H., Burian, B. K., Yule, S., & Dias, R. D. (2023). Extended reality applications for space health. Aerospace Medicine and Human Performance, 94(3), 122–130. https://doi.org/10.3357/amhp.6131.2023

Ebnali, M., Goldsmith, A. J., Burian, B., Atamna, B., Duggan, N. M., Fischetti, C., Dias, R. (2022). AR-coach: using augmented reality (AR) for real-time clinical guidance during medical emergencies on deep space exploration missions. Healthcare and Medical Devices, 51, 67. 10.54941/ahfe1002100

Fischer, A. M., Anna Rietveld, P. W. Teunissen, M. Hoogendoorn, and Petra Bakker. 2023. "What Is the Future of Artificial Intelligence in Obstetrics? A Qualitative Study among Healthcare Professionals." https://doi.org/10.22541/au.168441546.62793973/v1.

Franceschini, Carlos Maria, and Marcela Viviana Smurra. 2022. "Telemedicine in Sleep-Related Breathing Disorders and Treatment with Positive Airway Pressure Devices. Learnings from SARS-CoV-2 Pandemic Times." Sleep Science (Sao Paulo, Brazil) 15 (1): 118–27.

Hallowell, Nina, Shirlene Badger, Aurelia Sauerbrei, Christoffer Nellåker, and Angeliki Kerasidou. 2022. "'I Don't Think People Are Ready to Trust These Algorithms at Face Value': Trust and the Use of Machine Learning Algorithms in the Diagnosis of Rare Disease." BMC Medical Ethics 23 (1). https://doi.org/10.1186/s12910-022-00842-4.

Khanijo, S. 2020. "The Use of Telemedicine in Pulmonary Conditions." Journal of Quality in Health Care &Amp; Economics 3 (4). https://doi.org/10.23880/jqhe-16000173.

Knop, Michael, Sebastian Weber, Marius Mueller, and Bjoern Niehaves. 2022. "Human Factors and Technological Characteristics Influencing the Interaction of Medical Professionals With Artificial Intelligence–Enabled Clinical Decision Support Systems: Literature Review." JMIR Human Factors 9 (1): e28639.

Kooli, Chokri, and Hend Abdalrahman Al Muftah. 2022. "Artificial Intelligence in Healthcare: A Comprehensive Review of Its Ethical Concerns." Technological Sustainability 1 (2): 121–31.

Lai, Yi, Atreyi Kankanhalli, and Desmond C. Ong. 2021. "Human-AI Collaboration in Healthcare: A Review and Research Agenda." Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2021.046.

Morrow, Daniel G., Lisa M. Soederberg Miller, Heather E. Ridolfo, Nina K. Kokayeff, Dervon Chang, Ute Fischer, and Elizabeth A. L. Stine-Morrow. 2004. "Expertise and Aging in a Pilot Decision-Making Task." Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting 48 (2): 228–32.

Nayak, Priyakanta, Anil Purohit, Ayushi Shrivastava, Mathew Samuel, Jagdish Harsh, and Rahul Singhi. 2021. "A Webinar Training on Digitally Transforming the Future of Global Public Health." Indian Journal of Public Health Research &Amp; Development. https://doi.org/10.37506/ijphrd.v12i4.16560.

Orrange, Sharon, Wendy J. Mack, Julia Ann Cassetta, and Arpna S. Patel. 2020. "Patient Satisfaction and Trust in Telemedicine during the COVID-19 Pandemic (Preprint)." https://doi.org/10.2196/preprints.25509.

Palma, Maria del Carmen Ocón, Anna-Maria Seeger, and Armin Heinzl. 2019. "Mitigating Information Overload in E-Commerce Interactions with Conversational Agents." Information Systems and Neuroscience, 221–28.

Raeve, Paul De, Patricia M. Davidson, Franklin A. Shaffer, Eric Pol, Amit Kumar Pandey, and Elizabeth Adams. 2021. "Leveraging the Trust of Nurses to Advance a Digital Agenda in Europe: A Critical Review of Health Policy Literature." Open Research Europe 1: 26.

Rao, Arya S., John Kim, Meghana Kamineni, Michael Pang, Winston Lie, and Marc Succi. 2023. "Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making." https://doi.org/10.1101/2023.02.02.23285399.

Rovinelli, Richard J., and Ronald K. Hambleton. 1976. "On the Use of Content Specialists in the Assessment of Criterion-Referenced Test Item Validity," April. http://files.eric.ed.gov/fulltext/ED121845.pdf.

Sallam, Malik. 2023. "The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations." https://doi.org/10.1101/2023.02.19.23286155.

See, Linda, Alexis Comber, Carl F. Salk, Marijn van der Velde, Christoph Perger, Christian Schill, Ian McCallum, Florian Kraxner, and Michael Obersteiner. 2013. "Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts." PloS One 8 (7): e69958.

Strohm, Lea, Charisma Hehakaya, Erik R. Ranschaert, Wouter P. C. Boon, and Ellen H. M. Moors. 2020. "Implementation of Artificial Intelligence (AI) Applications in Radiology: Hindering and Facilitating Factors." European Radiology 30 (10): 5525–32.

Tavares, Diana, Ana Isabel Lopes, Catarina Castro, Gisela Maia, Liliana Leite, and Mónica Quintas. 2023. "The Intersection of Artificial Intelligence, Telemedicine, and Neurophysiology." Handbook of Research on Instructional Technologies in Health Education and Allied Disciplines, 130–52.

Tiwari, Rudra. 2023. "Explainable AI (XAI) and Its Applications in Building Trust and Understanding in AI Decision Making." Interantional Journal of Scientific Research in Engineering and Management 07 (01). https://doi.org/10.55041/ijsrem17592.

Tocchetti, Andrea, and Marco Brambilla. 2022. "The Role of Human Knowledge in Explainable AI." Brown University Digest of Addiction Theory and Application: DATA 7 (7): 93.

Yeung, Joshua Au, Zeljko Kraljevic, Akish Luintel, Alfred Balston, Esther Idowu, Richard J. Dobson, and James T. Teo. 2023. "AI Chatbots Not yet Ready for Clinical Use." https://doi.org/10.1101/2023.03.02.23286705.

Zhang, Yunfeng, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. "Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3351095.3372852.