The Human Side of Service Engineering, Vol. 143, 2024, 202–211
https://doi.org/10.54941/ahfe1005099

Redefining Emergency Services With Generative AI: Insights From a Preliminary Literature Review

Pavlina Kröckel[1,2]

[1]Friedrich-Alexander Universität Erlangen-Nürnberg, Nuremberg 90403, Germany
[2]Universitäres Hochschulinstitut Schaffhausen, CH-8200, Switzerland

ABSTRACT

The rapid evolution of artificial intelligence (AI) technologies, particularly generative AI (GenAI), has opened new frontiers in various sectors including emergency services. We present the results of a preliminary literature review of the current research on generative AI in emergency departments with a focus on use cases and their implications. We systematically examine peer-reviewed articles, case studies, and practical implementations to identify key trends, challenges, and opportunities in this growing field. Our findings underscore the need for ongoing research, ethical considerations, and cross-sector collaboration to fully leverage AI's capabilities to enhance the effectiveness and efficiency of emergency response.

Keywords: GenAI, ChatGPT, Triage, Emergency, SLR, LLM

INTRODUCTION

Artificial Intelligence (AI) in its various branches (e.g., robotics, machine learning) has already shown great potential in medicine and healthcare since ChatGPT was first released in November 2022. At the time of this writing, ChatGPT has already had a few upgrades and has done more to accelerate AI's growth in most industries than anything else we have seen previously. It has also pushed the regulation of AI further, especially in light of the new EU AI Act (European Commission, 2023).

Considering the rapid advancement of GenAI in the healthcare sector, our study aims to shed light on the current state of research on GenAI in emergency hospital departments (ED). ***GenAI*** is a branch of AI that can learn patterns from multimodal data (images, text, audio) and, in turn, generate new content based on these patterns (Ventura & Denton, 2023). In their review of AI in the ED, Boonstra and Laven (2022) found sufficient evidence pointing to the potential of AI tools to improve clinical decision-making in the ED. Specifically, the study found that AI support was mostly offered during the triage. Triage is the process of categorizing a condition based on the severity of the patient's symptoms, as well as the resources available for treatment (Gan, Uddin, Gan, Yew, & González, 2023). Our goal is to investigate how Generative AI, specifically, has been used in the ED. Thus, we define our **research question** as follows:

"*How have Generative AI tools been implemented in the ED so far, and what are the benefits and challenges of using such tools in this setting?*"

METHOD

We searched the following databases for relevant articles: PubMed, ScienceDirect, SpringerLink, Scopus, and Google Scholar. We defined our query as follows:

(generative AI OR GANs OR VAEs OR ChatGPT OR LLM) AND (triage OR emergency)

Initially, we restricted our search to "Title, abstract, and keywords" only, and in cases where we got no (relevant) results, we extended our search to the full-text search. An overview of the process steps is shown in Figure 1.

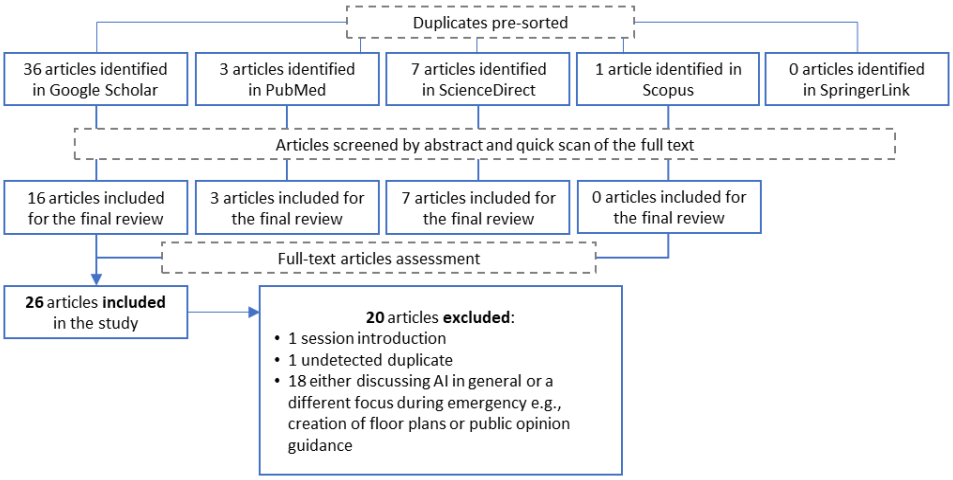

Figure 1: Literature review steps.

We defined our inclusion and exclusion criteria based on our research goal, that is, we aim to give a thorough overview of the current state of research on using GenAI in the ED. Thus, we include any article that discusses the use of GenAI in this context regardless of the application area (e.g., patient diagnosis, treatment) and regardless of the type of GenAI that was used (e.g., ChatGPT, GANs). Although our focus was on papers implementing a GenAI model in an emergency setting, we include papers that discuss ED use of GenAI in sufficient detail, albeit not exclusively. Following this line of thought, we excluded papers that only vaguely mention that GenAI can be used for triage purposes and do not investigate this in further detail. We ended with a final selection of 26 papers. A list is available online[1].

[1] A list of all articles included in the preliminary review is available here.

DATA ANALYSIS

We collected the full-text versions of all papers to be included in our review and registered the metadata of each article on a spreadsheet by adding the following information: title, publication outlet, impact factor of the journal, type of paper, year of publication, keywords, and DOI. We reviewed each paper in detail and added additional attributes to the spreadsheet. If a new attribute was added later in the process, we returned to the previously reviewed papers to fill out the missing information (if applicable). Finally, we collected the following information for each article: aim of the research, type of study, methods used (for data analysis), dataset, participants, type of GenAI, findings, benefits of GenAI, challenges, outlook/suggestions for future research, and use cases for GenAI.

RESULTS

Meta-Data Analysis

From the 26 articles reviewed in detail, 20 were categorized as original research. Two articles were "letter to the editor" and two were categorized as a "report or opinion", meaning that it was an expert's opinion on GenAI submitted to the editor of the respective journal. We chose to include these articles in our preliminary analysis, as they were written by medical practitioners, and it is important to convey the message by the primary users of GenAI in clinical settings. Finally, one article was a technical report as per the journal's own categorization scheme.

Most articles (19) were published in 2023 and 2024 (6). One article was published in 2022. This demonstrates the speed at which researchers have investigated the effects of GenAI in the ED following the first public release of the ChatGPT model in November 2022.

Regarding the quality of the published papers, six are from journals with an impact factor higher than 5, while for four studies we could not extract an impact factor (two are from the non-peer-reviewed site medRxiv, and two studies are from journals with no declared impact factor - the Canadian Journal of Health Technologies and the Journal of Ambient Intelligence and Humanized Computing). The median impact factor of all papers in this review is 3.6.

Goals of Existing Research

The research objectives of studies on GenAI in the ED reveal a relatively broad range of research focuses. These can be grouped into the following categories.

Triage Optimization and Assistance

A primary focus of many studies is assessing the accuracy of GenAI tools, such as ChatGPT and GPT variants, in diagnosing health conditions, triaging patients, and comparing their performance to that of healthcare professionals and other AI tools. This theme includes articles that consider the following aspects:

- The assessment of ChatGPT's performance in mass casualty incident triage and emergency department triage using various scales such as the Simple Triage And Rapid Treatment (START) triage, Korean Triage and Acuity Scale (KTAS), Canadian Triage and Acuity Scale (CTAS), and in diagnosing specific conditions such as stroke in young females and ophthalmological emergencies (Franc, Cheng, Hart, Hata, & Hertelendy, 2024; Gan et al., 2023; Gan, Ogbodo, Wee, Gan, & González, 2024; Gebrael et al., 2023; Ito et al., 2023; Kim, Kim, Choi, & Lee, 2024; Knebel et al., 2023; Rosen & Saban, 2023).
- The diagnostic and triage capabilities of GPT-3 and ChatGPT with attending physicians, lay adults, Google Bard, medical students, and existing medical AI tools, such as Ada Health and WebMD Symptom Checkers (Fraser et al., 2023; Gan et al., 2024; Levine et al., 2023).
- Evaluation of the ability of ChatGPT to aid in diagnostics and provide accurate triage in emergency scenarios, including prehospital basic life support and pediatric advanced life support scenarios (Bushuven et al., 2023; Dahdah et al., 2023; Jacob, 2023).

Applications in Radiology

Several studies focus on the application of GenAI in interpreting radiology reports and images, aiming to enhance the emergency department's radiology workflow and accuracy in anomaly detection. These include:

- Extracting emergency data from radiology reports and interpreting chest radiographs (Huang et al., 2023; Infante et al., 2024).
- Developing and validating algorithms for anomaly detection in brain CT images and stable chest radiographs during longitudinal follow-up (Seungjun Lee et al., 2022; Yun et al., 2023).

Clinical Workflow, Documentation, and Overall Decision Support

Articles in this category aimed to evaluate the real-world clinical impact of GenAI tools in the ED, including assessing the performance of GPT-3.5 in identifying higher acuity patients and the overall potential of GenAI to improve healthcare delivery in Canadian hospitals (Clark & Severn, 2023; Williams et al., 2023). The rest of the studies focused on the development of deep learning models for automating electronic health records by processing clinician-patient conversations and predicting patient length of stay (LOS) in the ED; the assessment of GenAI's potential in healthcare improvement through medical triage optimization and support in emergency care settings (Bhattaram, Shinde, & Khumujam, 2023; Kadri, Dairi, Harrou, & Sun, 2022; Siryeol Lee et al., 2023).

Advanced Support in Specific Emergencies

We found two studies evaluating GenAI's capability in specific emergency scenarios, such as managing metastatic prostate cancer patients in the ED and triaging polytrauma patients, indicating a push towards specialized applications of AI in emergency medicine (Gebrael et al., 2023; Jacob, 2023).

GenAI Implementations

In our study selection, ChatGPT was the most prominent form of GenAI. An overview is presented in Figure 2.

Six articles tested customized models such as a deep learning algorithm using thoracic cage registration and subtraction (Yun et al., 2023), a deep learning-driven GAN model (Kadri et al., 2022), a transformer-based encoder-decoder model (Huang et al., 2023), a deep generative model called the closest normal-style-based generative adversarial network (Seungjun Lee et al., 2022), pre-trained transformer-based LLMs specifically designed for Korean language (Siryeol Lee et al., 2023), and Perplexity and Bard (Infante et al., 2024).

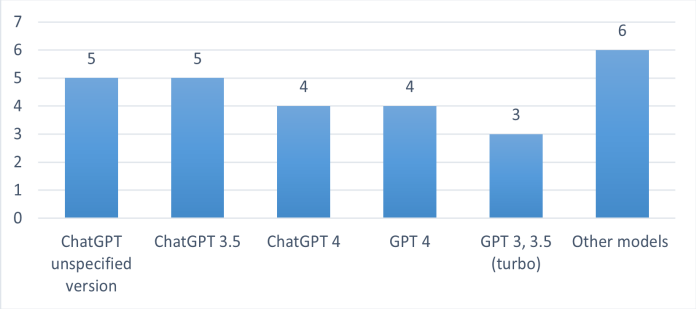

Figure 2: Implementations of GenAI used in existing research (the numbers represent the frequency of studies in which the respective type of GenAI was used).

Datasets

Various datasets were used in our selection of articles to examine the potential of GenAI in emergency scenarios. The results are summarized in the table below.

Table 1. Overview of datasets used in existing research.

Type of data	Used in
Clinical vignettes	• Ito et al., 2023 • Levine et al., 2023 • Bhattaram et al., 2023 • Franc et al., 2024 • Knebel et al., 2023 • Bushuven et al., 2023
Electronic Medical Records (EMR) and clinical data, ED visits	• Rosen & Saban, 2023 • Gebrael et al., 2023 • Fraser et al., 2023 • Kadri et al., 2022 • Williams et al., 2023

(Continued)

Table 1. Continued

Type of data	Used in
Radiology reports and imaging data	• Infante et al., 2024 • Huang et al., 2023 • Yun et al., 2023 • Seungjun Lee et al., 2022
Clinical conversations, consult queries and questionnaires	• Siryeol Lee et al., 2023 • Le & Amendola, 2023 • Gan et al., 2023 • Gan et al., 2024
Other (scenarios, virtual patient cases)	• Kim et al., 2024 • Jacob, 2023

Benefits

The reviewed studies highlighted several benefits of GenAI in the ED. Most of these revolve around improving patient care and optimizing the clinical workflow as a support for the healthcare workforce. The benefits are summarized as follows:

- **Rapid and accurate triage** – especially in situations with mass casualties and high patient loads, by quickly identifying critical conditions and optimizing patient flow (Gan et al., 2023; Gebrael et al., 2023; Jacob, 2023; Paslı et al., 2024).
- **Diagnostic precision** - AI models, such as GPT-4, demonstrate diagnostic accuracy comparable to that of physicians, enhancing diagnosis without introducing significant biases, and supporting rapid deployment of new AI applications due to their broad training (Ito et al., 2023; Levine et al., 2023).
- **Operational efficiency** - GenAI can be beneficial in ensuring healthcare service efficiency, supporting clinicians in diagnosis and triage decisions, enhancing patient experience, and reducing costs. It also offers potential for administrative tasks, such as electronic health record (EHR) management (Dahdah et al., 2023; Ito et al., 2023; Seungjun Lee et al., 2022).
- **Clinical decision support** – GenAI provides immediate medical guidance, supports early intervention, and assists in clinical reasoning, aiding healthcare professionals in decision-making and potentially leading to improved patient outcomes (Bhattaram et al., 2023; Huang et al., 2023; Infante et al., 2024; Seungjun Lee et al., 2022; Rosen & Saban, 2023; Williams et al., 2023).
- **Resource optimization and management** – GenAI enhances emergency department workflows, reduces radiologists' workload, and aids in managing overcrowding and resource allocation through predictive accuracy and real-time notifications (Kadri et al., 2022; Yun et al., 2023).
- **Better patient care** – by automating clinical and administrative processes, GenAI reduces healthcare providers' burden, allowing for a focus on direct

patient care and streamlining health services (Clark & Severn, 2023; Franc et al., 2024; Fraser et al., 2023; Gan et al., 2024; Kim et al., 2024; Knebel et al., 2023; Nashwan & Abujaber, 2023).

Challenges

Despite the many benefits of GenAI in the ER, there are significant challenges that need to be overcome for this type of AI to make a difference in clinical settings. The following is a summary of the challenges we extracted:

- **Technical and data limitations** – concerns over data dependency, model specificity, and adequacy of training data lead to issues such as bias and misclassification (Bhattaram et al., 2023; Franc et al., 2024; Ito et al., 2023; Kadri et al., 2022; Levine et al., 2023; Paslı et al., 2024; Ventura & Denton, 2023; Williams et al., 2023; Yun et al., 2023).
- **Operational and integration challenges** – difficulty in integrating AI into existing healthcare systems, ensuring that AI's recommendations are up-to-date and ethically sound, and managing the complexity of emergency settings (Bhattaram et al., 2023; Dahdah et al., 2023; Infante et al., 2024; Jacob, 2023; Kim et al., 2024; Seungjun Lee et al., 2022; Nashwan & Abujaber, 2023; Rosen & Saban, 2023; Yun et al., 2023).
- **Reliability and safety concerns** – risks associated with system hacking, misdiagnosis, and liability, along with the need for rigorous validation to confirm AI's utility and safety of AI in real-life scenarios (Bushuven et al., 2023; Clark & Severn, 2023; Fraser et al., 2023; Gan et al., 2023; Gan et al., 2024; Gebrael et al., 2023; Huang et al., 2023; Knebel et al., 2023; Seungjun Lee et al., 2022; Siryeol Lee et al., 2023; Rosen & Saban, 2023).
- **Ethical and legal considerations** – navigating the overall ethical requirements for developing responsible AI tools, addressing privacy concerns, and the potential for legal challenges due to negative health outcomes (Dahdah et al., 2023; Gan et al., 2023; Gan et al., 2024; Gebrael et al., 2023; Jacob, 2023; Nashwan & Abujaber, 2023).

CONCLUSION

Existing research on the application of GenAI in emergency medicine demonstrates a growing interest in exploring the capabilities of technologies, such as ChatGPT, in improving patient triage, diagnosis accuracy, and overall emergency care efficiency. Studies have assessed various aspects of GenAI's performance, including its ability to accurately triage patients in mass casualty incidents, diagnose health conditions considering patient race and ethnicity, and compare the diagnostic and triage effectiveness of attending physicians. The findings generally indicate that GenAI, particularly newer versions such as GPT-4, can perform comparably to healthcare professionals in diagnosing and triaging patients, showing high accuracy rates in identifying emergency conditions, and suggesting the potential to significantly support emergency medical decision-making. Additionally, the research explores GenAI's utility in specific medical scenarios, such as stroke diagnosis in young females and

chest radiograph interpretation, highlighting its promising role in enhancing clinical decision support systems and potentially improving medical education and emergency response efficiency.

However, along with the acknowledged potential of GenAI in transforming emergency healthcare practices, existing research also outlines significant challenges and areas for future investigation. Concerns, such as the risk of amplifying existing biases, the necessity for rigorous validation to ensure safety and efficacy, and the limitations posed by GenAI's current training data and adaptability, are frequently mentioned. The studies call for further research to address these challenges, emphasizing the importance of continuous monitoring, evaluation, and enhancement of GenAI models to ensure their responsible and ethical use in clinical settings. Considering the technical, ethical, and clinical implications, there is a consensus on the critical need for multidisciplinary efforts to effectively integrate GenAI into healthcare systems. Future research directions include improving GenAI's diagnostic and triage accuracy, exploring its application across a broader range of medical emergencies, and assessing its real-world impact on patient outcomes, healthcare efficiency, and the workload of emergency department staff.

REFERENCES

Bhattaram, S., Shinde, V. S., & Khumujam, P. P. (2023). Chatgpt: The next-gen tool for triaging? *The American Journal of Emergency Medicine*, *69*, 215–217. https://doi.org/10.1016/j.ajem.2023.03.027

Boonstra, A., & Laven, M. (2022). Influence of artificial intelligence on the work design of emergency department clinicians a systematic literature review. *BMC Health Services Research*, *22*(1), 669. https://doi.org/10.1186/s12913-022-08070-7

Bushuven, S., Bentele, M., Bentele, S., Gerber, B., Bansbach, J., Ganter, J., . . . Ranisch, R. (2023). "ChatGPT, Can You Help Me Save My Child's Life?" - Diagnostic Accuracy and Supportive Capabilities to Lay Rescuers by ChatGPT in Prehospital Basic Life Support and Paediatric Advanced Life Support Cases - An In-silico Analysis. *Journal of Medical Systems*, *47*(1), 123. https://doi.org/10.1007/s10916-023-02019-x

Clark, M., & Severn, M. (2023). Artificial Intelligence in Prehospital Emergency Health Care: CADTH Horizon Scan. *Canadian Journal of Health Technologies*, *3*(8).

Dahdah, J. E., Kassab, J., Helou, M. C. E., Gaballa, A., Sayles, S., & Phelan, M. P. (2023). Chatgpt: A Valuable Tool for Emergency Medical Assistance. *Annals of Emergency Medicine*, *82*(3), 411–413. https://doi.org/10.1016/j.annemergmed.2023.04.027

European Commission (2023). AI Act. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Franc, J. M., Cheng, L., Hart, A., Hata, R., & Hertelendy, A. (2024). Repeatability, reproducibility, and diagnostic accuracy of a commercial large language model (ChatGPT) to perform emergency department triage using the Canadian triage and acuity scale. *CJEM*, *26*(1), 40–46. https://doi.org/10.1007/s43678-023-00616-w

Fraser, H., Crossland, D., Bacher, I., Ranney, M., Madsen, T., & Hilliard, R. (2023). Comparison of Diagnostic and Triage Accuracy of Ada Health and WebMD Symptom Checkers, ChatGPT, and Physicians for Patients in an Emergency Department: Clinical Data Analysis Study. *JMIR MHealth and UHealth*, *11*, e49995. https://doi.org/10.2196/49995

Gan, R. K., Ogbodo, J. C., Wee, Y. Z., Gan, A. Z., & González, P. A. (2024). Performance of Google bard and ChatGPT in mass casualty incidents triage. *The American Journal of Emergency Medicine*, *75*, 72–78. https://doi.org/10.1016/j.ajem.2023.10.034

Gan, R. K., Uddin, H., Gan, A. Z., Yew, Y. Y., & González, P. A. (2023). Chatgpt's performance before and after teaching in mass casualty incident triage. *Scientific Reports*, *13*(1), 20350. https://doi.org/10.1038/s41598-023-46986-0

Gebrael, G., Sahu, K. K., Chigarira, B., Tripathi, N., Mathew Thomas, V., Sayegh, N., . . . Li, H. (2023). Enhancing Triage Efficiency and Accuracy in Emergency Rooms for Patients with Metastatic Prostate Cancer: A Retrospective Analysis of Artificial Intelligence-Assisted Triage Using ChatGPT 4.0. *Cancers*, *15*(14). https://doi.org/10.3390/cancers15143717

Huang, J., Neill, L., Wittbrodt, M., Melnick, D., Klug, M., Thompson, M., . . . Etemadi, M. (2023). Generative Artificial Intelligence for Chest Radiograph Interpretation in the Emergency Department. *JAMA Network Open*, *6*(10), e2336100. https://doi.org/10.1001/jamanetworkopen.2023.36100

Infante, A., Gaudino, S., Orsini, F., Del Ciello, A., Gullì, C., Merlino, B., . . . Sala, E. (2024). Large language models (LLMs) in the evaluation of emergency radiology reports: Performance of ChatGPT-4, Perplexity, and Bard. *Clinical Radiology*, *79*(2), 102–106. https://doi.org/10.1016/j.crad.2023.11.011

Ito, N., Kadomatsu, S., Fujisawa, M., Fukaguchi, K., Ishizawa, R., Kanda, N., . . . Tsugawa, Y. (2023). The Accuracy and Potential Racial and Ethnic Biases of GPT-4 in the Diagnosis and Triage of Health Conditions: Evaluation Study. *JMIR Medical Education*, *9*, e47532. https://doi.org/10.2196/47532

Jacob, J. (2023). Chatgpt: Friend or Foe?-Utility in Trauma Triage. *Indian Journal of Critical Care Medicine : Peer-Reviewed, Official Publication of Indian Society of Critical Care Medicine*, *27*(8), 563–566. https://doi.org/10.5005/jp-journals-10071-24498

Kadri, F., Dairi, A., Harrou, F., & Sun, Y. (2022). Towards accurate prediction of patient length of stay at emergency department: A GAN-driven deep learning framework. *Journal of Ambient Intelligence and Humanized Computing*, 1–15. https://doi.org/10.1007/s12652-022-03717-z

Kim, J. H., Kim, S. K., Choi, J., & Lee, Y. (2024). Reliability of ChatGPT for performing triage task in the emergency department using the Korean Triage and Acuity Scale. *Digital Health*, *10*, 20552076241227132. https://doi.org/10.1177/20552076241227132

Knebel, D., Priglinger, S., Scherer, N., Klaas, J., Siedlecki, J., & Schworm, B. (2023). Chatgpt in der präklinischen Versorgung augenärztlicher Notfälle – eine Untersuchung von 10 fiktiven Fallvignetten [Assessment of ChatGPT in the Prehospital Management of Ophthalmological Emergencies - An Analysis of 10 Fictional Case Vignettes]. *Klinische Monatsblatter Fur Augenheilkunde*. Advance online publication. https://doi.org/10.1055/a-2149-0447

Le, Q., & Amendola, M. (2023). Performance of ChatGPT on Fictional Vascular Surgery Consult Queries: A Potential Artificial Intelligence-Assisted Tool. *Journal of Vascular Surgery*, *78*(4), e110. https://doi.org/10.1016/j.jvs.2023.08.042

Lee, S. [Seungjun], Jeong, B., Kim, M., Jang, R., Paik, W., Kang, J., . . . Kim, N. (2022). Emergency triage of brain computed tomography via anomaly detection with a deep generative model. *Nature Communications*, *13*(1), 4251. https://doi.org/10.1038/s41467-022-31808-0

Lee, S. [Siryeol], Lee, J. [Juncheol], Park, J. [Juntae], Park, J. [Jiwoo], Kim, D., Lee, J. [Joohyun], & Oh, J. (2023). Deep learning-based natural language processing for

detecting medical symptoms and histories in emergency patient triage. *The American Journal of Emergency Medicine*, *77*, 29–38. https://doi.org/10.1016/j.ajem.2023.11.063

Levine, D. M., Tuwani, R., Kompa, B., Varma, A., Finlayson, S. G., Mehrotra, A., & Beam, A. (2023). The Diagnostic and Triage Accuracy of the GPT-3 Artificial Intelligence Model. *MedRxiv : The Preprint Server for Health Sciences*. Advance online publication. https://doi.org/10.1101/2023.01.30.23285067

Nashwan, A. J., & Abujaber, A. (2023). Leveraging Large Language Models to Improve Triage Accuracy in Emergency Departments. *Journal of Emergency Nursing*, *49*(5), 651–653. https://doi.org/10.1016/j.jen.2023.05.002

Paslı, S., Şahin, A. S., Beşer, M. F., Topçuoğlu, H., Yadigaroğlu, M., & İmamoğlu, M. (2024). Assessing the precision of artificial intelligence in emergency department triage decisions: Insights from a study with ChatGPT. *The American Journal of Emergency Medicine*. Advance online publication. https://doi.org/10.1016/j.ajem.2024.01.037

Rosen, S., & Saban, M. (2023). Can ChatGPT assist with the initial triage? A case study of stroke in young females. *International Emergency Nursing*, *70*, 101340. https://doi.org/10.1016/j.ienj.2023.101340

Ventura, C. A. I., & Denton, E. E. (2023). Artificial Intelligence Chatbots and Emergency Medical Services: Perspectives on the Implications of Generative AI in Prehospital Care. *Open Access Emergency Medicine: OAEM*, *15*, 289–292. https://doi.org/10.2147/OAEM. S420764

Williams, C. Y., Zack, T., Miao, B. Y., Sushil, M., Wang, M., & Butte, A. J. (2023). *Assessing clinical acuity in the Emergency Department using the GPT-3.5 Artificial Intelligence Model*. https://doi.org/10.1101/2023.08.09.23293795

Yun, J., Ahn, Y., Cho, K., Oh, S. Y., Lee, S. M., Kim, N., & Seo, J. B. (2023). Deep Learning for Automated Triaging of Stable Chest Radiographs in a Follow-up Setting. *Radiology*, *309*(1), e230606. https://doi.org/10.1148/radiol.230606