Artificial Intelligence, Social Computing and Wearable Technologies, Vol. 113, 2023, 268–278
https://doi.org/10.54941/ahfe1004199

Design for Integrating Explainable AI for Dynamic Risk Prediction in Prehospital IT Systems

David Wallstén[1], Gregory Axton[1], Eunji Lee[2], Anna Bakidou[2,3], Bengt Arne Sjöqvist[2], and Stefan Candefjord[2]

[1]Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden
[2]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
[3]PreHospen – Centre for Prehospital Research, University of Borås, Borås, Sweden

ABSTRACT

The aging population strains healthcare systems, necessitating new technologies for patient safety. Artificial intelligence (AI) development holds promise for improving care, but the necessity of and methods for explaining AI recommendations require further research. This study proposes a design for an eXplainable AI (XAI) for prehospital IT systems as a Clinical Decision Support System (CDSS) and reports how EMS clinicians perceive it. A literature review, ethnography, and expert interviews were conducted. These methods provided a knowledge foundation on which prototype designs were based. Developed prototypes were then verified by prehospital healthcare experts. The final prototype's usability and AI-user interaction were tested and evaluated by seven EMS clinicians through think-aloud protocol and interviews. The final design was tablet-based and included an XAI interface as a detailed overlay. The overlay was accessible during system interaction and displayed risk predictions generated by an AI method for assessment of trauma patients. Indication of presence or absence of a serious condition was indicated in combination with ranked information about predictors' influence on the prediction in addition to information about any essential variables with missing data. The EMS clinicians deemed XAI necessary for trusting predictions and to enable comparisons with clinical experience and judgment. The visualized rows of predictors and information about variables with missing data served as reminders, addressing common issues in patient assessment. The divergence between AI and clinicians' assessments prompted thoughtful decision-making, likely reducing decision bias. While focused on trauma, the design can be generalized to AI models for CDSS of other patient conditions. In conclusion, the incorporation of XAI in the user interface is an important factor in increasing user trust. The user feedback regarding different design features can be used to guide future AI development in prehospital healthcare, by providing insights about potential benefits and implications.

Keywords: Artificial intelligence (AI), Explainable AI (XAI), Prehospital care, Design, Prototypes, Risk prediction, Usability

INTRODUCTION

The world population is aging (Kanasi et al., 2016), straining current healthcare systems. This creates a demand for novel solutions, where Artificial Intelligence (AI) has a potential to support and improve clinical practice (Rajpurkar et al., 2022). Because healthcare involves high-stakes decisions, the need for cognitive understanding for a human working with AI becomes pivotal (Sahoh, Choksuriwong, 2023). eXplainable AI (XAI) can bridge that gap of understanding. XAI currently lacks a proper technical definition, but key elements are interpretability, how much the model can be understood, transparency, how the model works, and explainability, providing insight into the reasons for AI predictions (Antoniadi et al., 2021). XAI is currently still in dire need of more research, in particular when it comes to usability (Antoniadi et al., 2021).

Prehospital care focuses on caring for acutely injured or ill patients by Emergency Medical Services (EMS) clinicians, before reaching a hospital, on-site or during transport. Prehospital care in particular deals with a wide variety of patient cases, where the importance of correct assessment and decision within a short time frame is often critical. Focusing on Sweden, current protocols for early assessment/triage of a patient's condition yield low triage accuracy (Magnusson, 2021). This can lead to undertriage, meaning patients at risk are assessed as less serious and thus may not receive adequate care. It can also lead to overtriage, with the patient receiving more healthcare resources than necessary. The incorporation of AI in EMS practice was recently pointed out as a promising application area (Kirubajan, 2021), and findings in empirical studies demonstrate improved EMS triaging with AI (Seki et al., 2019; Park et al., 2020; Candefjord et al., 2021). Clinical Decision Support Systems (CDSS) utilizing AI can therefore be considered as a key component in future EMS decision-making.

Although there have been efforts to utilize AI for improved clinical performance, research about the usability of AI applications remains scarce. This paper therefore aims to study how XAI can impact EMS clinicians' perceptions of usability and cognitive decision-making processes, by proposing and evaluating a human-machine interface CDSS design for ambulance IT systems with AI functionality. The research questions this study seeks to answer are: what design aspects are important when integrating XAI in a prehospital IT system with an integrated dynamic risk prediction, and how do EMS clinicians perceive such a system? To answer the research question, a prototype of a prehospital IT system with an integrated dynamic risk prediction supported by an XAI model was designed and tested.

To limit the scope of the study, it focuses on an On Scene Injury Severity Prediction (OSISP) model developed for all trauma events, similar to a model focused on trauma from motor vehicle crashes (Candefjord et al., 2021). The OSISP model sees no indications of significant performance benefits when comparing models with different transparency levels, e.g., logistic regression and artificial neural networks (ANN). Therefore, data that can be extracted from more transparent models, e.g., logistic regression, formed a starting point for designing the prototype.

METHODS

This study followed the double diamond model (British design council, 2019). It involves four phases: *discover*, where the problem is explored; *define*, where the aim is defined; *develop*, when potential solutions are investigated; and *deliver*, which presents a solution. Methods used in each phase are presented in Table 1.

During *discovery*, a literature review (Martin et al., 2012) was used to gain general knowledge in the fields of XAI, EMS and User Experience (UX). Articles were searched in Google Scholar, PubMed via Chalmers University of Technology library's catalogue and discovery tool EDS (EBSCO Discovery Service), with keywords including but not limited to "AI", "AI healthcare", "explainable AI", "triage models", "over-triage" and "under-triage". Ethnography (Hammersley and Atkinson, 2007) followed, involving visiting the main ambulance stations of two different Swedish regions with two different triage solutions, one digital and one based on paper, though both followed similar protocols. Time was spent alongside EMS clinicians while they were on missions, granting knowledge about their work environment, workflows, and their current challenges. Investigations were conducted into the currently most widespread EMS IT systems in Sweden, Paratus (CSAM, Lysaker, Norway) and MobiMed (Ortivus AB, Danderyd, Sweden), to study their structure, both for general inspiration and to investigate how a XAI solution could be designed to work with current systems. Interviews (Martin et al., 2012) with five domain experts were conducted to gain both general knowledge about the prehospital setting as well as specific knowledge in study-related fields. The interviewees were an EMS clinician tutor, an AI researcher, a product manager for Paratus, a process manager of implementing a new IT system in Region Västra Götaland in west of Sweden, and a prehospital researcher with industry experience, all with a focus on the Swedish prehospital setting.

During the *define* phase, hierarchical task analysis (Stanton et al., 2017) and journey maps (Kale, 2020) were conducted to understand the workflow of the users while personas (Miaskiewicz & Kozar, 2011) were used to help understand the users. Eventually, the gathered insights of possibilities and challenges in the fields of prehospital IT systems and future implementation of AI were used to define an aim for the design using MoSCoW prioritization (Must- Should- Could- and Would-have) (Ahmad et al. 2017), by setting up a list of requirements. An AI model for trauma was chosen to keep the work focused, while still having a goal to create a general interface applicable to other prehospital AI applications than trauma.

Prototyping (Houde and Hill, 1997) followed in the *develop* phase, starting with low-fidelity prototyping, which allowed gathering of quick feedback while spending fewer resources on each prototype. This was done both through sketching and using the digital design tool, Figma. Several ideas were developed, and feedback was gathered during meetings with experts in the prehospital field or from their responses to recorded videos of the prototype. When a design was regarded to meet the agreed-upon requirements, visual design was adapted based on the National Health Service design system

(2023). Final feedback was received from two UX designers, a professor in interaction design, and an industry professional before the final interactable high-fidelity prototype was developed using Figma.

In the final *deliver* phase, usability testing (Interaction design foundation, 2002) was conducted. During the testing, the think-aloud protocol (Hanington & Martin, 2012) was utilized, followed by semi-structured interviews to gather information on the participants' thoughts on the effectiveness of the prototype (Dumas and Redish, 1999). The usability testing involved seven EMS clinicians who were aged between 25–48 (M = 39.3, SD = 7.6). The participants were given a short video introduction about a scenario where they were for the first time to use an AI application that had been tested and certified for clinical effectiveness. They were then given two patient cases defined by the authors, one that represented a common case of overtriage for injury sustained in motor vehicle crash, and another that represented a case of undertriage where an 80-year-old woman fell on a hard floor. The participants were instructed to fill in information about the patient according to current EMS protocol and use the AI to help them triage the patient.

Table 1. Methods used during the different phases of the double diamond model.

Design phases	Methods
Discover	Literature Review, Ethnography, Expert Interview
Define	Hierarchical task analysis, Journey maps, Personas, MoSCoW (Must- Should- Could- and Would-have)
Develop	Low-fidelity prototyping (Paper, Figma), High-fidelity prototyping (Figma), Expert review
Deliver	Usability testing: Think-aloud protocol & User interview

RESULTS

Key findings from the literature review included that current IT systems are rarely used by EMS clinicians to support decisions on patient handling and treatment, but usually only for journaling (Porter et al., 2020). Furthermore, results from the review confirmed that XAI user research in healthcare is currently scarce (Antoniadi, 2021).

Ethnography resulted in information about the EMS workflow. EMS clinicians first gain limited information from a call from the Public Safety Answering Point (PSAP), given a priority based on the acuteness. On scene, they first conduct an early assessment of the most important factors, such as if the patient is breathing, before doing a more detailed analysis of several parameters, such as measuring blood pressure. The patient is then triaged and can either stay at home, be referred to primary care, be transported to a nearby hospital or be sent directly for emergency care. Additionally, ethnography results correlated with the findings from our literature review regarding EMS clinicians not using the EMS IT systems to support diagnosis, instead they are registering most Electronic Patient Records (EPR) data when the patient has already been triaged.

From the interviews, key findings included that there are current issues with data transfer and interoperability between different EPR systems, but that new technological solutions and revised regulations may allow for increased real-time data transfers and access to more contextual EPR data. Another finding was that AI applications should be able to always work in the background, giving live updates to the EMS clinicians as risk predictions are altered based on newly entered information. From the AI application used as a reference in this study, there were no indications of significant performance differences when comparing ANN to simpler models, e.g., logistic regression. Thus, simpler models may be preferred because of increased transparency, according to interviews.

Based on findings from the *discover* phase, a list of requirements was constructed. This was used to define the aim of our study, which became to develop and evaluate a prototype for XAI with a focus on trauma, but that should be adaptable enough to be used with models for other conditions. The prototype should be based on current EMS IT systems to simplify future integration, with local adaptions to Region Västra Götaland in Sweden to enable testing. It should also always be accessible by the user and give notifications to the user. During low-fidelity prototyping, careful consideration was taken to design the system interaction with as few necessary clicks as possible to complete an action, and to use icons, colors, and contrast to help readability and guide users' attention.

The final prototype is shown in Figure 1, and its workflow is shown in Figure 2. To the left of the Figure 1 is a navigation bar, allowing access to the different pages of the system, and the top bar shows which clinician is logged in, patient info, and current risk prediction, followed by additional icons, such as guidelines, battery, and connectivity. These are inspired by current IT systems, such as MobiMed or Paratus.

The prototype uses an overlay page to display the AI risk prediction, meaning that an interactive page appears on top of the base page when the user selects protocol parts (patient info, triage protocol, Vital parameters, Scenario, Care and Medications, Non-medical journal, or Summary/send-off) or application domain (chest pain, stroke, trauma). The interactive overlay allows the user to interact with the XAI for the selected purpose while keeping the base page untouched. It can be accessed at all times, either by swiping from the right, clicking a notification or the button in the top bar, next to the patient info. Two additional input pages were also included, not included in Figure 1, to enable the overlay to be tested during a normal EMS clinician workflow. At any point during the workflow, when given sufficient input to calculate a risk prediction, a notification is displayed, giving information about a change of prediction or whether the system urgently needs more data.

The overlay is divided into three main sections: risk prediction and guidelines, predictors for and against a serious condition, and most important missing variables. Risk prediction displays the current triage color based on the commonly used EMS South African Triage System (SATS) (EMSSA, 2017), the confidence the model has in this prediction on a 5-grade scale, and a box for adaptable treatment guidelines depending on the prediction. This is followed by the predictors, where two rows display either predictors for

a serious condition, which could be high age, or predictors against a serious condition, which could be normal brain function. The degree to how much they contribute to either direction is ordered from left to right. These are both scrollable from the side to be able to view more predictors. They can also be changed into a bar chart to get a more detailed overview. The intention of displaying the predictors in this way is to give the EMS clinicians an explanation for what predictors the AI considers to be most important for its prediction, and to allow for the personnel to make their own assessment based on that data. Finally, the most important missing variables allow the user to add data the model considers important to increase prediction confidence. It thus allows for easy addition of variables, without forcing the user to find specific pages for different parameters.

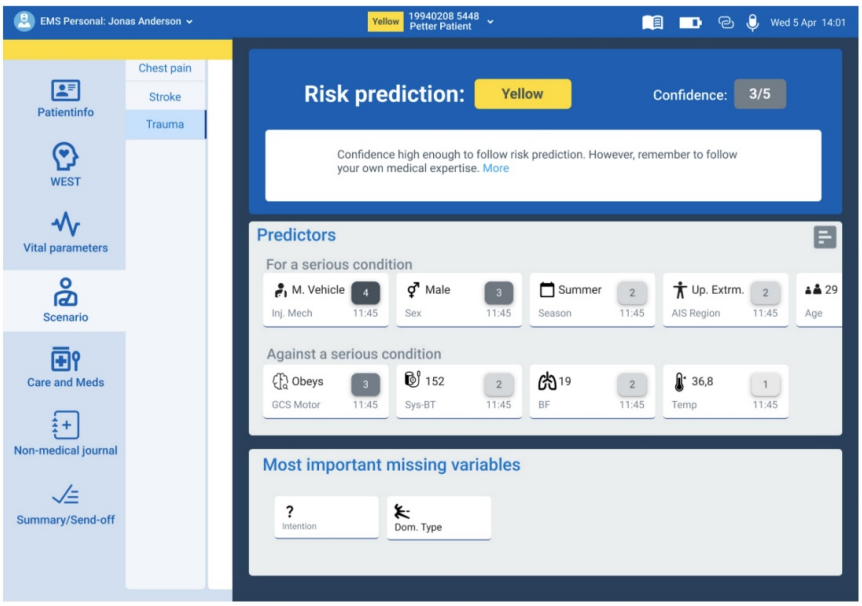

Figure 1: XAI overlay. Left: navigation bar of the main system. Top outside the overlay: EMS clinician info, status bar and patient info. Top of the overlay: risk prediction, confidence level and recommendation. Middle of the overlay: predictors for and against a serious condition. Bottom of the overlay: interactive list of missing variables.

During the usability testing, most of the participants did not interact directly with the XAI output when given the risk prediction notification, instead continued to input all the information that was on their current page, before interacting with the overlay. When using the overlay, the prediction appeared clear. If they agreed with the prediction, it helped them feel more confident in their own assessment based on their expertise. When it was different, it made the participants think an extra time if there might have been something they had missed, and some commented that this might help reduce bias. If deciding between two levels of severity, such as orange or yellow, the risk prediction result might nudge them in either direction, similar to when they confer with their colleague as to what they should do.

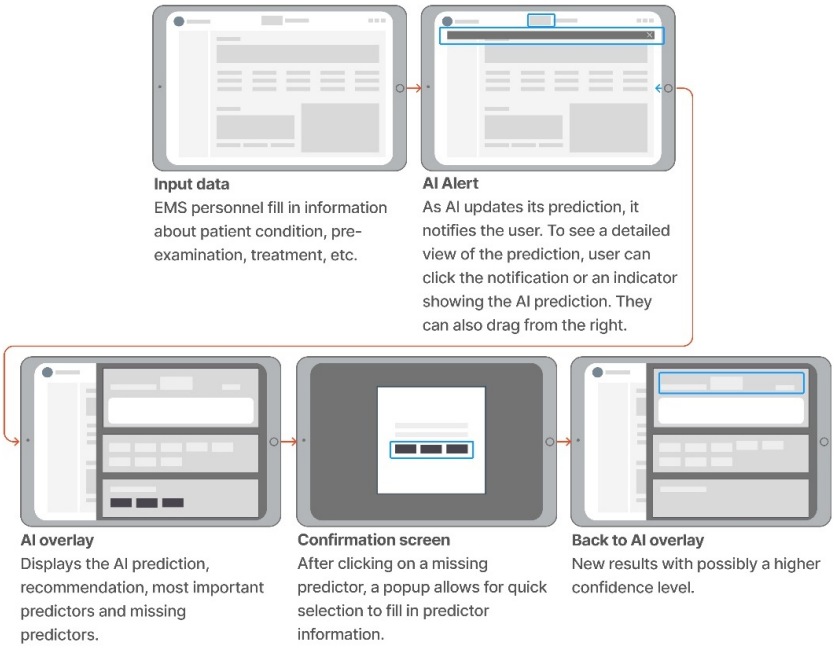

Figure 2: XAI prototype interactive workflow.

The participants understood the predictors' field to be what the risk model deemed most important, though some initially had issues discerning the difference between the two rows before being given hints. All participants appeared to appreciate the display of current predictors and missing predictors, since it may act as a reminder to control if they have missed measuring vitals or to reconsider the importance of a factor compared to the initial assessment. However, in a few observations, the model was not perceived to take enough predictors into account, which reduced the trust of the AI.

Overall, the design was perceived to have a high degree of usability. Some mentioned that the benefits of the design would likely increase reliance compared to current systems, increasing patient safety. Cognitive benefits, especially for helping memory, were noted by having access to the most important predictors and what parts of the assessment they might have missed. All EMS clinicians agreed that the design features used to describe why a prediction was made were essential to build trust for the system. They were also positive about AI in general and thought that it will be part of future care.

DISCUSSION

The prehospital field is complex. Technical solutions need to be adapted for a wide variety of conditions, be quick due to the time-critical nature and be easy to interpret correctly. CDSS that are not AI-based have the potential to increase reliance on guidelines but can often increase time spent on the scene (Andersson et al., 2019). Furthermore, too cumbersome systems often result in not being used at all (Porter et al., 2020). This study proposes a general

design for XAI integration in current ambulance IT systems. It is designed to be quick to use, and the added benefit of AI could increase the accuracy of predicting patient risk and deciding an appropriate transport destination, contributing to increased patient safety. Out of the current commercial IT systems explored in this study, the design should be able to be implemented with only minor design adaptations.

Our study indicates that XAI is essential to building trust in AI for EMS clinicians. They are making the final decision, and they are responsible. The EMS clinicians thus have a need to be able to justify that decision and know what it is based on. A considerable benefit of the XAI design suggested here was how the clinicians perceived it to help them remember what to focus on. When making quick decisions in high-stakes and stressful situations, cognitive biases can be potentiated (Yu, 2016). Having a reminder as to what predictors are most important to focus on should help in diverting from an intuitive focus on predictors that might be misleading.

The prediction itself might also help cognitive decision-making. When the AI suggests an opposing view, it can make the clinician think more analytically about what decision to make. In a situation when the clinicians can't decide between two different assessments, the AI might help them decide in either direction. Since AI has the potential to improve accuracy compared to clinical state-of-the-practice (Candefjord et al., 2021), these effects should mostly be positive. However, there is also a risk of the clinicians becoming over-reliant on the AI, while clinicians are still responsible for the decisions.

An important consideration for AI development is the number of predictors. Candefjord et al., (2021) showed that using merely 5–7 predictors can be sufficient to provide high prediction accuracy. While fewer predictors make the input interaction quicker, which is paramount, it might also be perceived to overlook conditions the clinicians consider important, reducing their trust. Although adding further parameters might help, the added information might not contribute much to prediction accuracy. A potentially good alternative is to educate the nurses as to why not all predictors are needed for the algorithm to make a robust prediction. This study was limited in that it tested the design in a situation where clinicians only were given a bare-bones understanding of the AI. While this tests the limits of the designs' explainability, it does not investigate the influence of educating users on AI theory, which would be a requirement for it to be used and understood (Kim, 2020). Studies of long-term usage and scenarios where education on the system has been given are required. In the long term, clinicians are likely to adapt differently to the system depending on AI performance. This could both enhance or reduce trust. Further, the study was done in a lab setting, and a more real-life scenario might have given different findings.

While the current design shows which predictors are most important, it does not show why. Most of this is up to the clinical expertise of the nurses to interpret. Future development should look into how this information could be portrayed. Using the current design, clicking on a predictor could give this information. At the same time, it is important not to overcomplicate the design since there is a risk of increased time demand.

The XAI design presented here would currently not be possible to implement using ANN due to the lack of transparency. Future and current models might benefit from using ANN despite their black-box nature. To maintain trust from EMS clinicians, tools to make these models should be more transparent or extended education in how they work could be a necessity.

Future implications of the design remain unexplored. How EMS clinicians would trust the system and rely on it over time likely changes their interaction with the system and their decisions. If the AI is perceived as being consistently correct, EMS clinicians might become so reliant on it that it is essentially the AI making medical decisions. Generally, people appear to prefer when a human makes the decision (Leyer & Schneider, 2019). An argument can be made that many medical decisions are already made using simplified triage models (i.e., current state-of-the-practice) that do not interpret complex patient data like AI risk prediction, which could yield higher triage accuracy.

CONCLUSION

This study presents a design for integrating XAI into current ambulance IT systems, its design process, and its test results by EMS clinicians. Results indicate a high perception of usability of the design. It allows clinicians to receive continuous feedback, likely enhancing their ability to make correct clinical decisions in critical situations. The study reinforces that XAI is much needed in the medical field.

Future studies should quantitatively look at what is the final triage decision compared to AI with a large sample. Future design considerations could allow the user to drill down to get more information about why the AI considers certain predictors to be important, as well as which combinations of variables provides most accurate prediction with high confidence. When and how to notify the user and its impact on workflow needs further investigation.

ACKNOWLEDGMENT

The authors would like to acknowledge all EMS clinicians that contributed to this study. Without them this work would not have been possible. The same applies to: Elin Maxstad and Andreas Dehre from PICTA, Stefan Jönsson, Region Västra Götaland, and Jonas Borgström, Paratus. This study was financially supported by Vinnova IoT Sweden through the project *ASAP PoC Improved Civil and Military prehospital Point-of-Care decisions through Data Fusion and AI*, Ref. No. 2022-03748.

REFERENCES

Ahmad, K. S., Ahmad, N., Tahir, H., & Khan, S. (2017, July). Fuzzy_MoSCoW: A fuzzy based MoSCoW method for the prioritization of software requirements. In 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (pp. 433-437). IEEE.

Andersson Hagiwara, M., Lundberg, L., Sjöqvist, B. A., & Maurin Söderholm, H. (2019). The effects of integrated IT support on the prehospital stroke process: Results from a realistic experiment. Journal of Healthcare Informatics Research, 3, 300–328.

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Applied Sciences, 11(11), 5088.

British Design Council. (2019) https://www.designcouncil.org.uk/our-resources/framework-for-innovation/

Dumas, J. S., & Redish, J. (1999). A practical guide to usability testing. Intellect

E. Park, K. Lee, T. Han, and H. S. Nam, "Automatic Grading of Stroke Symptoms for Rapid Assessment Using Optimized Machine Learning and 4-Limb Kinematics: Clinical Validation Study," J Med Internet Res, vol. 22, no. 9, p. e20641, Sep. 2020, doi: 10.2196/20641

Emergency Medicine Society of South Africa, EMSSA (2017), The South African Triage Scale (SATS) [Internet]. Available from: https://emssa.org.za/special-interest-groups/the-south-african-triage-scale-sats/#objectiveshttps://www.designcouncil.org.uk/our-work/skills-learning/tools

Hammersley, M., & Atkinson, P. (2007). Ethnography: Principles in practice (3rd ed.). Routledge.

Houde, S., & Hill, C. (1997). What do prototypes prototype?. In Handbook of human-computer interaction (pp. 367–381). North-Holland.

Kale, P. (2020). User journey maps: What they are and how to create one. InVision.

Kanasi, E., Ayilavarapu, S., & Jones, J. (2016). The aging population: demographics and the biology of aging. Periodontology 2000, 72(1), 13–18.

Kim, H. S. (2020). Decision-making in artificial intelligence: is it always correct?. Journal of Korean Medical Science, 35(1).

Kirubarajan, A., Taher, A., Khan, S., & Masood, S. (2020). Artificial intelligence in emergency medicine: a scoping review. Journal of the American College of Emergency Physicians Open, 1(6), 1691–1702.

Magnusson, C. (2021) Patient Assessment and Triage in Emergency Medical Services.

Martin, B., Hanington, B., & Hanington, B. M. (2012). Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Pub.

Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes?. Design studies, 32(5), 417–430.

National Health Service (NHS). (2023). Design and build digital services for the NHS. https://service-manual.nhs.uk/design-system/design-principles

Porter A, Badshah A, Black S, Fitzpatrick D, Harris-Mayes R, Islam S, Jones M, Kingston M, LaFlamme-Williams Y, Mason S, McNee K, Morgan H, Morrison Z, Mountain P, Potts H, Rees N, Shaw D, Siriwardena N, Snooks H, Spaight R, Williams V. Electronic health records in ambulances: the ERA multiple-methods study. Southampton (UK): NIHR Journals Library; 2020 Feb. PMID: 32119231.

Rajpurkar, P., Chen, E., Banerjee, O. et al. AI in health and medicine. Nat Med28, 31–38 (2022). https://doi.org/10.1038/s41591-021-01614-0

Retrieved from https://www.invisionapp.com/inside-design/user-journey-maps/

S. Candefjord, A. Sheikh Muhammad, P. Bangalore, and R. Buendia, "On Scene Injury Severity Prediction (OSISP) machine learning algorithms for motor vehicle crash occupants in US," Journal of Transport & Health, vol. 22, p. 101124, Sep. 2021, doi: 10.1016/j.jth.2021.101124

Sahoh, B., Choksuriwong, A. The role of explainable Artificial Intelligence in high-stakes decision-making systems: a systematic review. J Ambient Intell Human Comput 14, 7827–7843 (2023). https://doi.org/10.1007/s12652-023-04594-w

Seki, T., Tamura, T., Suzuki, M. Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique. Resuscitation 141, pp. 128–35 (2019). Doi: https://doi.org/10.1016/j.resuscitation.2019.06.006

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. pp. 45–76. (2017). Human factors methods: a practical guide for engineering and design. CRC Press.

The Swedish EMS nurse in a new role. [Doctoral dissertation, University of Gothenburg. Sahlgrenska Academy]. URL: https://gupea.ub.gu.se/handle/2077/67134

Yu, R. (2016). Stress potentiates decision biases: A stress induced deliberation-to-intuition (SIDI) model. Neurobiology of stress, 3, 83–95.