Towards Human-AI Interaction in Medical Emergency Call Handling

Carsten Maletzki
German Research Center for Artificial Intelligence (DFKI)
Carsten.Maletzki@dfki.de

Christian Elsenbast
Fraunhofer-Institute for Experimental Software Engineering (IESE)
christian.elsenbast@iese.fraunhofer.de

Melanie Reuter-Oppermann
University of Twente
Technical University of Darmstadt
m.n. reuter-oppermann@utwente.nl

Abstract

Call-takers in emergency medical dispatch centers typically rely on decision-support systems that help to structure emergency call dialogues and propose appropriate responses. Current research investigates whether such systems should follow a hybrid intelligent approach, which requires their extension with interfaces and mechanisms to enable an interaction between call-takers and artificial intelligence (AI). Yet unclear is how these interfaces and mechanisms should be designed to foster call handling performances while making efficient use of call-taker's often strained mental capacities. This paper moves towards closing this gap by 1) deriving required artifacts for human-AI interaction and 2) proposing an iterative procedure for their design and evaluation. For 1), we apply the guidelines for human-AI interaction and conduct workshops with domain experts. For 2), we argue that performing a full evaluation of the artifacts is too extensive at earlier iterations of the design process, and therefore propose to enact use-case-driven lightweight evaluations instead.

Keywords: Medical Emergency Calls, Hybrid Intelligence, Mental Workload, Mental Picture

1. Introduction

Whenever citizens experience situations that require immediate medical care, skillful responses to their emergency calls are key (Zachariah and Pepe, 1995). To promote consistency in this regard, it is seen advantageous to provide emergency call-takers with decision support systems (DSS) that help to structure their conversations by displaying questions and possibly proposing appropriate emergency responses (Baumann, 2017; Mayr, 2020). Currently available systems, however, fall short in terms of adapting to the dynamic nature of emergency calls (Baumann, 2017) and therefore leave a gap that has to be bridged by the call-taker. Recent research in this area suggests that DSSs could provide a more adaptive decision support by integrating artificial intelligence (AI) (Rietzke, 2021). In this context, a close human-AI interaction appears to be a cornerstone for a successful adaptation of such an AI-driven DSS in practice (Farand et al., 1995). A conceptual framework to guide the design of this interaction can be found in hybrid intelligence – a recently emerged research focus that concentrates on combining humans and AI with the goal of fostering synergies in their respective complementary strengths (Akata et al., 2020; Dellermann et al., 2019).

A fundamental component of a hybrid intelligence system is an interface that allows for human-AI interaction. Throughout this interaction, humans and AI learn from each other to improve their performance over time (Akata et al., 2020; Dellermann et al., 2019). The design of an interface and the respective mechanisms for hybrid intelligent emergency call handling poses a widely unexplored challenge. Since emergency calls are often handled under time pressure, we assume that utilizing a call-taker's mental resources efficiently will be a crucial aspect to a successful adaptation of such an interface. This applies equally to mental overload as well as the complete lack of mental workload.

In this paper, we 1) characterize the artifacts required for human-AI interaction in a hybrid intelligent DSS for medical emergency call handling and 2) describe a procedure to design and evaluate them with a strong focus on the resulting mental workload for call-takers. For 1), we apply the *Guidelines for Human-AI Interaction* (Amershi et al., 2019) and discuss their

URI: https://hdl.handle.net/10125/106790
978-0-9981331-7-1

implications based on findings from self-conducted workshops with experts for emergency call handling from the German state of Rhineland-Palatinate. For 2), we develop a procedure to create and evaluate the required artifacts, while focusing on their impact on mental workload. In this context, we also lay out the most basic aspects of mental workload in emergency medical dispatch centers. With our work, we take steps towards hybrid intelligent emergency call handling and contribute to the current state of decision support for call-takers of medical emergency calls.

In the following, we will start by giving a general overview about DSSs in medical emergency call handling, hybrid intelligence and mental workload (Section 2). Subsequently, we describe the concept of hybrid intelligent emergency call handling (Section 3) and define the required artifacts for human-AI interaction (Section 4). After considering the aspects of mental workload in medical emergency call handling and a general description of how it can be measured (Section 5), we will introduce the procedure to develop and evaluate the required artifacts (Section 6). We will conclude by summarizing our findings and giving an outlook on future work (Section 7).

2. Foundations

Within information systems research, various concepts and classes of systems exist that aim to support users with decision-making – DSSs being one of those. DSSs are computer-based, interactive systems that align with the following aspects: 1) They assist rather than replace users in making decisions, 2) they use data and models as a basis, and 3) they focus on effectiveness rather than efficiency of the decision-making process (Eom et al., 1998). That is, DSSs aim to provide decisional advice to enable faster, better, and easier decision-making (Aronson et al., 2005). Within healthcare, DSSs are often used to support medical decision-making and suggest diagnoses (e.g. Mangiameli et al., 2004). Another important aspect of DSSs is the provision of 'decisional guidance' by explaining to the user why the system performs a certain action, suggests a specific decision, or outputs a computed result (Silver, 1991). By imitating the decision-making ability of a human expert, the class of expert systems aim to support human users during the decision-making process (Jackson, 1986). For practical use of all types of DSSs, providing explanations describing what the system knows, how it works, and why specific actions are appropriate is an important feature in order to increase the users' acceptance of the provided decisions, suggestions, or results (Gregor and Benbasat, 1999; Richardson et al., 1990; Swartout, 1987; Ye and Johnson, 1995).

In the context of information systems for emergency management, Shen et al. investigates display formats and conducted two experiments with decision makers (Shen et al., 2012). When users are supported in their decision making by provided decisional guidance, their performance, measured as decision accuracy as well as decision speed, increases (Shen et al., 2012). Applying the design science methodology, the authors Reuter-Oppermann et al., 2017 proposed a design for an assistant that can support the dispatchers in an emergency coordination center. While a system in the back contains optimization approaches to determine which ambulance to send to an incident, the assistant offers explanations for the choice of the algorithm as well as the dispatching suggestion in order to increase the dispatchers' trust and decrease their stress Reuter-Oppermann et al., 2017.

The concept of hybrid intelligence originates from a socio-technological perspective on AI-based support for humans, possibly working on cognitively demanding tasks. The main research goal of hybrid intelligence lies in augmenting human intelligence, rather than aiming at a substitution, while leveraging human strengths and compensate their weaknesses (Akata et al., 2020). This is done with the goal of achieving sophisticated goals that neither humans nor AI could reach alone (Dellermann et al., 2019). Essential to hybrid intelligence in this context is a collaboration between humans and AI that is expressed through interaction. This interaction is steered by a collaboration mechanism and forms the basis for a mutual learning process in which both learn to improve their task performance (Dellermann et al., 2019; Hemmer et al., 2021). The foundations for any collaborations in hybrid intelligence systems are interfaces through which human-AI interactions can be enacted. The design of interfaces for human-AI interaction has been addressed by many works, which has led to a widespread body of design recommendations (Amershi et al., 2019). These recommendations have been recently consolidated into the *Guidelines for Human-AI Interaction* to ease the application of the field's knowledge when designing how human-AI interactions are initiated, executed and evolved (Amershi et al., 2019).

Mental workload, a fundamental construct in human factors and cognitive psychology, plays a critical role in understanding human performance, safety, and well-being. Mental workload, in its essence, refers to the cognitive demands imposed on an individual's mental resources while performing a task. It encompasses the allocation and utilization of cognitive processes such as attention, memory, decision-making, and information processing capacity. The concept of mental workload is rooted in the understanding that human cognitive resources are finite and limited in their capacity. When individuals engage in tasks that require substantial cognitive effort, their mental workload increases as they allocate and distribute their cognitive resources to meet the demands of the task at hand. This allocation of cognitive resources is influenced by various factors, including task complexity, time pressure, task novelty, and individual differences. In complex, safety-critical environments, (e.g., in emergency call handling in dispatch centers), both underload and overload are very real concerns (Young et al., 2015). Understanding mental workload is essential for optimizing task performance, designing efficient systems, and promoting human well-being in complex and demanding environments.

3. Hybrid Intelligence in Medical Emergency Call Handling

This section outlines how we envision the paradigm of hybrid intelligence to improve current DSSs in medical emergency call handling.

According to Møller et al. (Møller et al., 2021), call-takers handle medical emergency calls by executing an iterative procedure (Figure 1) that is driven by contextual influences on themselves and on their callers. Such influences arise, for example, from a caller's ability to describe the situation to the call-taker. This determines the extent to which a call-taker is able to obtain information about the emergency. When executing the iterative procedure, call-takers interpret this information by applying their knowledge to construct a so-called mental picture of the emergency. This mental picture provides the fundament for call-takers when deciding on how to react to a given situation. Such decisions can lead to tasks that need to be managed by the call-taker. Typical examples of such tasks involve the provision of suitable emergency resources or instructions to the caller when providing first aid.

DSSs support call-takers in executing the iterative procedure by guiding through a predefined questionnaire and possibly suggesting emergency responses. In this context, there are DSSs that substitute the need for mental pictures by forcing call-takers to follow their instructions precisely. A drawback of this approach is that it inhibits the ability of call-takers to adapt to unforeseen situations (Baumann, 2017). Less strict systems, on the other hand, allow call-takers to utilize their mental pictures for adaption but typically

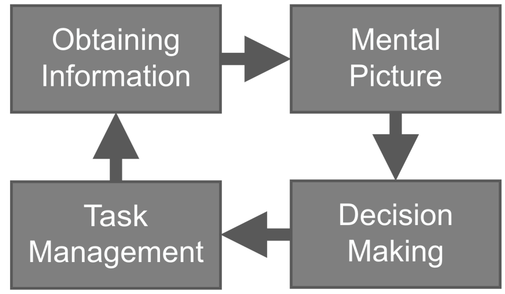

Figure 1. Iterative Procedure in Emergency Call Handling (based on Møller et al., 2021)

only provide rough suggestions. As call-takers have to analyze these suggestions manually to decide whether they are relevant to a given situation, the extensiveness of support is limited by available mental capacities. Our recent work on hybrid intelligent emergency call handling (Maletzki et al., 2023; Maletzki et al., 2022) addresses this issue by aiming at an augmentation of call-taker's mental capacities with the analytical power of AI. The aim of this approach is to achieve a wider variety of possible and greater precision of issued suggestions by utilizing AI to analyze obtained information and derive the relevance of suggestions. AI, in this context, takes over the role of a co-pilot, who judges the situation from its perspective and points in directions that could be important to consider. An AI's judgement of the situation is thereby referred to as the *artificial mental picture* of the system. It forms the basis for a human-AI interaction that aims to influence a call-taker's mental picture and, consequently, the decisions to be made. Figure 2 summarizes how the concept of hybrid intelligent emergency call handling ties in with the iterative procedure.

4. Required Artifacts for Human-AI Interaction in Emergency Call Handling

This section outlines the required artifacts for a human-AI interaction in medical emergency call handling that aims to provide call-takers with suitable suggestions about upcoming steps of the iterative procedure. We will specify the artifacts by discussing the most relevant requirements to this work defined by the *Guidelines for Human-AI Interaction* (Amershi et al., 2019). Our arguments originate from workshops we conducted with experts for emergency call handling from the German state of Rhineland-Palatinate.

Clarify the System's Functionalities in Advance: Call-takers should be able to familiarize

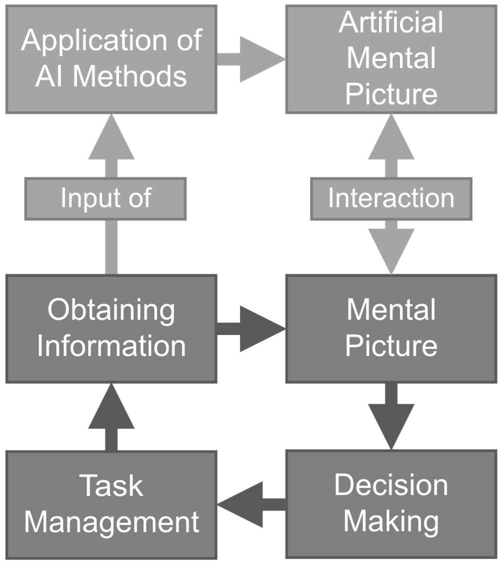

Figure 2. Hybrid Intelligent Emergency Call Handling (based on Møller et al., 2021)

with the kind of contributions of the system and their meanings before any human-AI interaction takes place. We expect that this helps them to understand the role of AI in their decision-makings and fosters efficiency in human-AI interaction. For this reason, call-takers should be provided with a concise guide that specifies the functionalities of the system. This guide should, for example, define that the system will provide suggestions about possible diagnoses and required measures. Thereby, the guide should also define how the suggestions will be displayed and how they are to be interpreted when making decisions.

Clarify the System's Performance in Advance: To build trust in the AI's influence on mental pictures of call-takers, the system should provide a dashboard to compare important key figures about call handling performances with and without AI support. Initially, this requires the definition of key figures and subsequently a visual concept for the dashboard.

Interact at the Right Time: Emergency call handling focuses on prioritizing most critical aspects for patient outcome. This can act as a guideline when identifying the right time to provide a suggestion. As a foundation to this, the system requires a mechanism to prioritize suggestions based on their criticality and a concept to visualize them accordingly.

Provide Contextually Relevant Information: In addition to suggestions about suspected diagnoses and required measures, the system should be able to provide contextually relevant information that may affect decision-making. Such contextual information could be, for example, a weather warning at the emergency site that may influence the feasibility of a suggested measure. This results in the need for artifacts to identify and visualize contextually relevant information that possibly influence decision-making based on provided suggestions.

Enable Efficient Dismissal and Correction: In its role as a co-pilot, AI will regularly refer to the call-taker with suggestions to accept, reject or correct. To allow for efficiency during this interaction, an appropriate visual concept is required. Further artifacts to enable efficient corrections can be identified when considering that emergency-relevant information can change during the call. As a result, it could be possible that a decision was made based on out-of-date information. In such cases, an efficient correction requires a visual concept to provide the call-taker with a procedure that needs to be executed in reaction to these changes. Further, also a mechanism to identify this procedure is required.

Provide Alternatives when in Doubt: If it is not clear, for example, which is the ideal rescue resource or the most suitable suspected diagnosis for a call, the system should offer alternatives. At the same time, the system should also provide contextually relevant information that facilitates the selection of a suggestion. Thus, there is a need for a mechanism to recognize alternatives and a visualization concept that allows these alternatives to be weighed up efficiently by regarding contextually relevant information.

Provide Explanations of AI: In case call-takers require a thorough understanding of why AI behaved in a certain way, they should be provided with suitable explanations. Due to the time-pressure of medical emergency calls, these explanations should initially focus on the necessities and allow for extension. Therefore, a visualization concept for explanations is needed that allows for extension according to the needs of a call-taker.

Learn from User-Behavior and Provide Feedback Mechanisms: In hybrid intelligent emergency call handling, learning could be applied to optimize how suggestions are handled during human-AI interaction. This could lead to an adaption of the priority of a suggestion, or even to an automation of a reaction to a suggestion. As a fundament to this, the system requires a concept to collect appropriate data and a mechanism for its evaluation. To collect data, the system could, for example, ask call-takers for their explicit feedback or log how interactions are being executed. The system further also requires a mechanism to ensure that automations and alterations of priorities are only performed if they are safe for a patient. Another safeguard that should be in place is a concept for visualizing automations that allows for efficient correction by the call-taker.

Clarify the Impact of User Actions on Future Interactions: The described concept to collect data for optimizing upcoming human-AI interactions could, for example, ask call-takers for their explicit feedback or log their interactions with the system. Regardless of how this feedback is acquired, the system should make it transparent how feedback will affect future interactions. Therefore, the system should provide conspicuously placed descriptions about how feedback influences the behavior of the system. If necessary, this explanation should also provide a general description of the mechanism to evaluate the feedback.

Keep Adaptions Manageable to the User: Hybrid intelligent emergency call handling, as introduced in Section 3, aims at adaptive suggestions on how to enact the iterative procedure of medical emergency call handling based on available information. Whenever information gets available that leads to an adaption of the system's suggestions, it must be ensured that call-takers can maintain an overview. This results in the need for a visualization concept that allows call-takers to keep up with changes in the system's suggestions.

By applying the *Guidelines for Human-AI Interaction*, we have identified various artifacts to initialize, execute, and evolve human-AI interactions in hybrid intelligent medical emergency call handling. Among these artifacts are mechanisms and visual concepts to facilitate familiarization with the system, provide contextual information and explanations, and handle reactions. To evaluate these artifacts, the mental workload they produce for call-takers is especially important for artifacts related to the execution and evolution of human-AI interactions. As a foundation for defining a procedure to perform an evaluation of these artifacts, we will subsequently describe the characteristics of mental workload of call-takers and lay out how it can be measured during medical emergency call handling.

5. Measuring Mental Workload in Emergency Call Handling

The work of dispatchers in emergency medical dispatch centers is characterized by complexity. This is primarily due to the large number of variables that influence the correct tactical decision. However, some of the variables needed to make a safe decision, such as the number of casualties and injury patterns, are largely unknown at the time of the call. Other components of the work environment arise from technology, such as the interaction of data transmission via telephony or telemetry with the DSS and its subsystems. The operation of the DSS must take place parallel to the emergency call in order to be able to trigger an alert of the required emergency resource (e.g., ambulance) already during the emergency call dialog. If this interface is considered in a human-centered way, it can be assumed (Venkatesh and Davis, 2000), that a multitude of different variables influence the adaptation and readiness to use the technology, which must also be transferred to the DSS and its subsystems (Elsenbast and Hagemann, 2023). Furthermore, this human-computer interaction cannot be achieved without the use of mental resources for the operation of the DSS. Therefore, according to cognitive load theory (Paas and Sweller, 2010), there is a mental workload due to cognitively managing the mission and due to operating the DSS. Consequently, it is also necessary to analyze mental underload situations (Hancock, 1989), if applicable.

Regarding mental workload, there are many different explanatory models, but they all have in common the finite nature of human cognition (Kramer and Spinks, 1991). It is worth mentioning that it is obvious that both mental overload (hyperstress) and mental underload (hypostress) lead to performance losses, as Hankock and Warm state in the Dynamic Adaptive Theory (Hancock, 1989). Figure 3 depicts how mental activations imposed by a task at hand relate to mental workload and influence task performance. Notice that the areas of mental underload and overload are marked red. As a consequence, both hyperstress and hypostress should be considered with respect to emergency medical dispatch centers and high responsibility. The latter could occur, for example, when call volume is low or when simple,

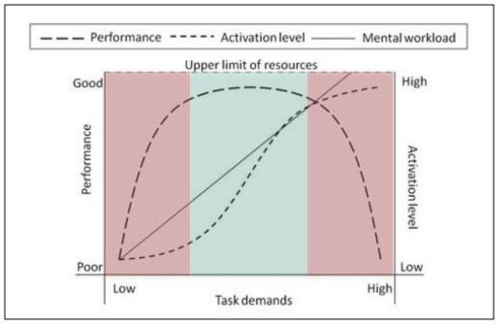

Figure 3. Relationship between Performance, Activation Level and Mental Workload (based on Young et al., 2015)

Figure 4. EDA Measurement with Movisens EDAMove 4

repetitive activities accumulate.

Consequently, the question arises as to what level of mental workload dispatchers should have in order to make safe and correct decisions on the one hand and to achieve health resilience on the other.

Unlike, for example, arterial blood pressure, which can be measured directly with an invasive probe in an artery, mental workload as a latent construct cannot be measured directly. Therefore, reflective parameters, such as physiological correlates, are used for measurement. These include heart rate variability (HRV; especially Root Mean Square of Successive Differences, RMSSD, and Standard Deviation of the NN Intervall, SDNN), electrodermal activity (EDA), and, to a limited extent, pupillometry (pupil diameter), which can be subject to validity loss due to external influences (e.g., changing brightness of the environment and screen). In particular, the first two methods are well established for assessing stress in an occupational context. EDA is measured by surface electrodes and sensor [Movisens, EDAMove 4, figure 4] on the right foot, HRV by chest strap with sensor [Movisens, ECGMove 4] and pupillometry by eye tracking [Pupil Labs Core, figure 5].

Pupillometry is only an ancillary finding of the eye movement analysis and is therefore evaluated additionally. At the end of each test, participants are asked to complete a questionnaire that contains the Raw-TLX, a validated psychometric instrument for estimating workload, including mental workload. Note that TLX stands for Task Load Index. The Raw-TLX is the one-step abbreviated form of the two-step NASA-TLX. While the NASA-TLX addresses the different facets of workload in the first step and applies a weighting in the second step (Hart, 2006), the Raw-TLX omits the weighting and thus simplifies the survey. The Raw TLX is highly correlated with the NASA TLX and can be considered equivalent (Byers et al., 1989).

In order to generate a comparison between the status quo in the dispatch center, tests were first conducted with the conventional DSS. In a second phase, the conventional DSS will be compared with the innovative DSS in a within-subjects design. A within-subjects design has a higher sensitivity because there is no confounding due to individual differences. Since the same participants are studied in all conditions,

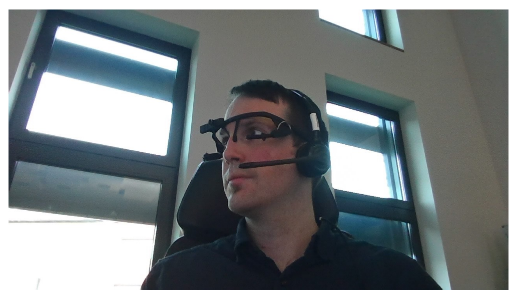

Figure 5. Pupillometry Measurement with Pupil Labs Core

all person-related confounding variables are perfectly paralleled. This allows each participant to determine how the independent variable affects his or her behavior. Demand characteristics and position effects must be taken into account. Demand characteristics are the subjectively perceived demand contents of the respective experimental situation and are connected with the hypothesis of what the subject thinks he has to do. On the one hand, it should be pointed out that demand characteristics, as a complex interaction, require an independent and profound consideration. On the other hand, the physiological parameters used here are not expected to be biased to any relevant degree. Position effects are countered by alternating the position of the independent variables in the within-subjects design (1st: A-B, 2nd B-A, 2nd A-B, etc.). An overview of the methods used is provided in Table 1.

Measurement	Parameter
Electrodermal Activity	Skin Conductance Level
Heartrate Variability	RMSSD, SDNN
Eyetracking	Pupil Diameter
Questionnaire	Raw-TLX

Table 1. Measurements and Parameter of Mental Workload Assesment

The time required is 60–65 minutes per person, starting with an informational interview where the preliminary information is expanded to include, for example, privacy aspects. After verbal and written consent, the sensors are applied and calibrated. The actual measurement process begins with a relaxation sequence to establish a baseline. During this phase, a relaxation video and relaxing music are played. Different videos and music can be selected by the participants to suit their individual preferences. This is followed by the measurement phase at the control center workstation. The study ends with the questionnaire described above and a debriefing interview. The different phases are shown in Figure 6. This measurement protocol is based on the presented aspects of the dispatch center work, but also on the psychophysiological principles already mentioned. In addition, this excerpt represents a best practice which, at the time of its submission, is significantly based on several pre-tests in three different dispatch centers as well as two series of measurements carried out in two dispatch centers. Additional measurement series in other dispatch centers will follow. The raw data and derived findings will be published after analysis.

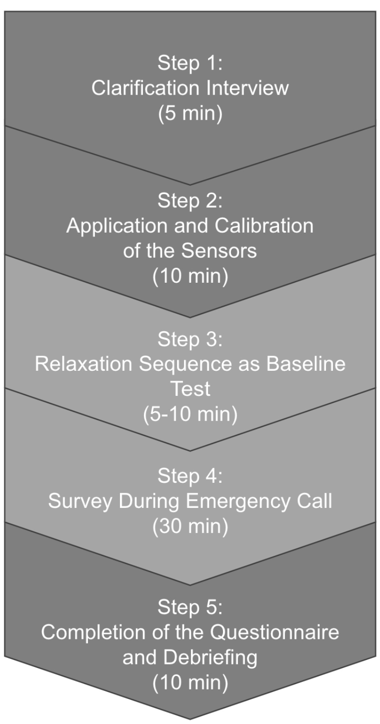

Figure 6. Measurement Phases

6. Procedure to Develop Required Artifacts

A quantification of mental workload is an important key figure to evaluate many of the outlined artifacts for human-AI interaction in medical emergency call handling. However, a thorough evaluation in the context of the overall approach of hybrid intelligent medical emergency call handling requires further key figures, like a quantification of the call handling performance. Examples for such key figures are call-duration, the number of dispatched emergency resources and the appropriate reliance (Schemmer et al., 2022) of call-takers on AI. An evaluation that quantifies all such aspects is an elaborate procedure that requires extensive resources during its execution. Since we expect the design of identified artifacts to require several attempts to yield the best possible results, evaluating each design thoroughly each time would produce immense evaluation efforts. Therefore, this section proposes to include a lightweight preliminary evaluation into the development procedure that can be executed with manageable effort. Figure 7 depicts the proposed procedure to develop the identified artifacts. The proposed procedure consists of two

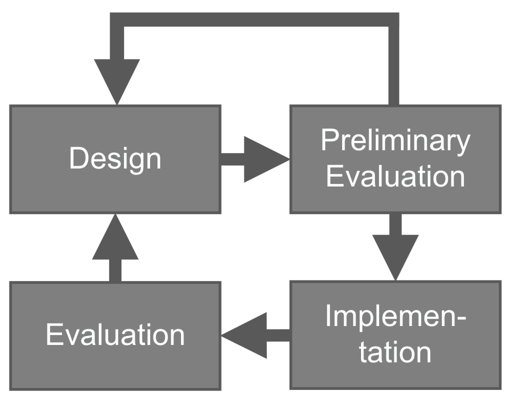

Figure 7. Iterative Development Procedure

design-evaluation cycles – the preliminary cycle and the main cycle. Compared to the main cycle, the preliminary cycle is characterized by the omission of an implementation activity and the inclusion of a lean preliminary evaluation. This preliminary cycle aims to iteratively assess and refine the output of the design step, which is expected to produce interface mockups and pseudocode. As a means to identify required refinements, expert workshops are conducted during the preliminary evaluation. In these workshops, domain experts are requested to roughly estimate how designed artifacts might influence the key figures for evaluation compared to the support of current DSSs. To make it as easy as possible for experts to provide assessments of these influences, the workshops will build on use cases designed to simulate the behavior of designed artifacts. Designs that experts expect to be beneficial towards their key figures will be considered in the main design-evaluation cycle, where they will be implemented to extend a DSS for medical emergency calls. In the following main evaluation, the resulting prototypical system will be applied to simulated emergency calls while relevant key figures will be collected. Following the evaluation of collected key figures, the whole procedure can be executed again in case critical weaknesses have been identified. Since we expect that multiple main design-evaluation cycles will be required and measuring mental workload precisely implies enormous efforts, earlier iterations will focus solely on providing participants with a questionnaire. Later iterations of the design-evaluation cycle, however, will include the full described measuring procedure. Thereby, a between-subject approach is used, in which each call-taker works on each case only once – either with or without AI support.

A major advantage of the proposed development procedure is that it enables lightweight design-evaluation cycles through the preliminary evaluation. However, we expect that gains in terms of quickness come at the cost of precision compared to a full evaluation in the main cycle. Due to an incorrect assessment during a preliminary evaluation workshop, it could happen that a design will be discarded that would have led to solid results in the main evaluation. Therefore, when applying the procedure, vigilance should be exercised towards the risk of discarding a design mistakenly and should only be done with a reasonable explanation.

7. Conclusion and Outlook

In this paper, we outlined the required artifacts for human-AI interaction in hybrid intelligent emergency call handling and defined a procedure for their design and evaluation. By applying the *Guidelines for Human-AI Interaction* (Amershi et al., 2019), we were able to identify artifacts to initiate, execute, and evolve human-AI interactions. To design and evaluate identified artifacts, we proposed an iterative development procedure that integrates a lightweight preliminary evaluation step. This preliminary evaluation aims to gather expert assessments about the influence of designed artifacts on relevant key figures. In comparison to the full evaluation that is executed subsequently for promising artifacts and aims at measuring these influences, we expect the preliminary evaluation to enable quicker design-evaluation cycles in earlier iterations of the procedure. Despite the benefits of this approach, it comes with the risk of wrong expert assessments that can lead to the dismissal of designs that would achieve good results in the full evaluation. This risk must be kept in mind when executing the procedure.

Future work will be concerned about designing the identified artifacts while potentially applying the design science research (DSR) methodology. Then, we will first derive a set of design requirements, followed by design principles and leading to design features and a matching prototype as a conceptual model.

This work has a strong focus on medical emergency calls – future work will also consider the transferability to other related domains, such as firefighting-related emergency calls.

Furthermore, in relation to the aim of DSR to generate new design knowledge, we aim to further extend our research to the more general issue of decision support and human-computer interaction, inside and outside of emergency coordination centers.

8. Acknowledgments

This work is funded by the Federal Ministry for Economic Affairs and Climate Action under grant No. 22973 SPELL. https://spell-plattform.de/en

References

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., ... Welling, M. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer*, *53*(8), 18–28. https://doi.org/10.1109/MC.2020.2996587

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

Aronson, J. E., Liang, T.-P., & MacCarthy, R. V. (2005). *Decision support systems and intelligent systems* (Vol. 4). Pearson Prentice-Hall Upper Saddle River, NJ, USA:

Baumann, A. (2017). Leitlinien oder Protokolle zur Notrufabfrage. In A. Hackstein & H. Sudowe (Eds.), *Handbuch Leitstelle* (pp. 189–193). Stumpf + Kossendey.

Byers, J. C., Bittner, A., & Hill, S. G. (1989). Traditional and raw task load index (tlx) correlations: Are paired comparisons necessary. *Advances in industrial ergonomics and safety*, *1*, 481–485.

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In T. Bui (Ed.), *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 1–10). ScholarSpace.

Elsenbast, C., & Hagemann, V. (2023). Assessment of emergency medical service providers' and dispatchers' technology commitment to the use of intelligent mission support systems. *International Paramedic Pratice*, *13(3)*, 59–67. https://doi.org/10.12968/ippr.2023.13.3.59

Eom, S., Lee, S., Kim, E., & Somarajan, C. (1998). A survey of decision support system applications (1988-1994). *Journal of the Operational Research Society*, *49*, 109–120. https://doi.org/10.1057/palgrave.jors.2600507

Farand, L., Leprohon, J., Kalina, M., Champagne, F., Contandriopoulos, A. P., & Preker, A. S. (1995). The role of protocols and professional judgement in emergency medical dispatching. *European Journal of Emergency Medicine*, *2*(3), 136–148.

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly*, 497–530.

Hancock, P. A. (1989). A dynamic model of stress and sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *31*(5), 519–537. https://doi.org/10.1177/001872088903100503

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. In D. Vogel, K. N. Shen, P. S. Ling, M. N. Ravishankar, & J. X. Zhang (Eds.), *25th Pacific Asia Conference on Information Systems, PACIS 2021, Virtual Event / Dubai, UAE, July 12-14, 2021*.

Jackson, P. (1986). Introduction to expert systems.

Kramer, A., & Spinks, J. (1991). Capacity views of human information processing. *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches*, 179–249.

Maletzki, C., Grumbach, L., Rietzke, E., & Bergmann, R. (2023). Towards Hybrid Intelligent Support Systems for Emergency Call Handling. In A. Martin, H.-G. Fill, A. Gerber, K. Hinkelmann, D. Lenat, R. Stolle, & F. von Harmelen (Eds.), *Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023), Hyatt Regency, San Francisco Airport, California, USA, March 27-29, 2023*.

Maletzki, C., Rietzke, E., & Bergmann, R. (2022). Utilizing Expert Knowledge to Support Medical Emergency Call Handling. In C. Beierle, M. Ragni, K. Sauerwald, F. Stolzenburg, & M. Thimm (Eds.), *Proceedings of the 8th Workshop on Formal and Cognitive Reasoning co-located with the 45th German Conference on Artificial Intelligence (KI 2022), Virtual Event, Trier, Germany, September 19, 2022* (pp. 79–89).

Mangiameli, P., West, D., & Rampal, R. (2004). Model selection for medical diagnosis decision support systems. *Decision Support Systems*, *36*(3), 247–259.

Mayr, B. (2020). Strukturierte bzw. standardisierte Notrufabfrage: Leisten die Systeme tatsächlich, was sie vorgeben zu leisten? *Notfall und Rettungsmedizin*, *23*, 505–512. https://doi.org/10.1007/s10049-020-00733-4

Møller, T. P., Jensen, H. G., Viereck, S., Lippert, F., & Østergaaard, D. (2021). Medical dispatchers' perception of the interaction with the caller during emergency calls - a qualitative study. *Scandinavian Journal of Trauma,* *Resuscitation and Emergency Medicine*, *29*. https://doi.org/10.1186/s13049-021-00860-y

Paas, T., Fred; van Gog, & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educ Psychol Rev*, (21(11-12)), 1461–1485.

Reuter-Oppermann, M., Morana, S., & Hottum, P. (2017). Towards designing an assistant for semi-automatic ems dispatching. *Proceedings of the HICSS conference 2017*.

Richardson, G. L., Jackson, B. M., & Dickson, G. W. (1990). A principles-based enterprise architecture: Lessons from texaco and star enterprise. *MIS quarterly*, 385–403.

Rietzke, E. (2021). *Modellierung, Visualisierung und Ausführung wissensintensiver Prozesse unter Verwendung Semantischer Technologien* (Doctoral thesis). Universität Trier.

Schemmer, M., Hemmer, P., Kühl, N., Benz, C., & Satzger, G. (2022). Should I Follow AI-based Advice? Measuring Appropriate Reliance in Human-AI Decision-Making. *arXiv preprint*. https://doi.org/10.48550/arXiv.2204.06916

Shen, M., Carswell, M., Santhanam, R., & Bailey, K. (2012). Emergency management information systems: Could decision makers be supported in choosing display formats? *Decision support systems*, *52*(2), 318–330.

Silver, M. S. (1991). Decisional guidance for computer-based decision support. *MIS quarterly*, 105–122.

Swartout, W. (1987). Explanation. In *Encyclopedia of artificial intelligence*.

Venkatesh, V., & Davis, F. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*, 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Ye, L. R., & Johnson, P. E. (1995). The impact of explanation facilities on user acceptance of expert systems advice. *Mis Quarterly*, 157–172.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics [PMID: 25442818]. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151

Zachariah, B. S., & Pepe, P. E. (1995). The development of emergency medical dispatch in the USA: a historical perspective. *European Journal of Emergency Medicine*, *2*, 109–112.