Towards Hybrid Intelligent Support Systems for Emergency Call Handling

Carsten Maletzki[1,*], Lisa Grumbach[1], Eric Rietzke[1,2] and Ralph Bergmann[1,3]

[1]German Research Center for Artificial Intelligence (DFKI), Branch University of Trier, Behringstraße 21, 54296 Trier, Germany
[2]LiveReader GmbH, Zur Imweiler Wies 3, 66649 Oberthal, Germany
[3]Business Information Systems II, University of Trier, Behringstraße 21, 54296 Trier, Germany

Abstract
Handling medical or firefighting-related emergency calls requires rich analytical and emotional skills to cope with callers in possibly life-threatening situations. In this context, joining human call-takers and artificial intelligence (AI) in a hybrid intelligent system promises superior call-handling performances. To our knowledge, it remains open which AI methods can be integrated into such a system to achieve a benefit and how to support call-takers when deciding how far to rely on AI. We address these gaps by analyzing emergency call handling to derive a) exemplary integrations of AI methods and b) a mechanism to calculate reliabilities of inferences based on experiences in similar situations. To this end, we build on recently introduced concepts for integrating emergency call-takers with AI based on our Ontology- and Data-Driven Expert System (ODD-ES). We identify that call-takers could benefit from applying multiple AI methods to distinct inference issues to yield comprehensive analytical support. However, applying multiple AI methods in parallel could lead to conflicting inferences. In this context, we expect that our proposed mechanism can help to resolve these conflicts as it provides comparable and quickly interpretable decision support. Future work is needed to overcome a cold-start issue of the mechanism.

Keywords
Hybrid Intelligence, Emergency Call Handling, Human-AI Collaboration

1. Introduction

Fast and precise responses to medical or firefighting-related emergency calls are crucial to overcoming worrisome and potentially life-threatening situations. However, meeting these requirements can be demanding for emergency call-takers in cases where they must apply broad analytical and emotional skills. This is, for example, when they have to assess an exceptional situation under time pressure while being confronted with a panicking caller. In such cases, we expect that joining forces of human call-takers and artificial intelligence (AI) in a hybrid

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van 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.
*Corresponding author.
Carsten.Maletzki@dfki.de (C. Maletzki); Lisa.Grumbach@dfki.de (L. Grumbach); Eric.Rietzke@dfki.de (E. Rietzke); Ralph.Bergmann@dfki.de (R. Bergmann)
0000-0003-0983-5015 (C. Maletzki); 0000-0002-2247-8270 (L. Grumbach); 0000-0001-5252-4859 (E. Rietzke); 0000-0002-5515-7158 (R. Bergmann)

intelligent system leads to increased call-handling performances. Hybrid intelligence envisions systems in which human intelligence and AI work together and learn from each other to achieve goals they could not achieve on their own [1, 2].

As a foundation for a hybrid intelligent support system for emergency call handling, we have recently introduced the Ontology- and Data-Driven Expert System (ODD-ES) [3]. As expert systems aim to mimic the thinking, skill, and intuition of experts [4], ODD-ES imitates the cognitive processes of call-takers when handling emergency calls and takes this as a basis for a process-oriented integration of human intelligence and AI. However, it remains open which AI methods to integrate into such a system and how to help call-takers decide how far to rely on AI while being under time pressure.

In this paper, we identify exemplary integrations of AI methods into ODD-ES to support emergency call-takers and sketch a mechanism to calculate reliabilities for inferences based on experiences in similar situations. We derive our solution from an analysis of emergency call handling that focuses on call-takers' skill requirements. This part of our work grounds on a qualitative study from Møller et al. [5] and self-conducted qualitative interviews with experts for emergency call handling from the German state Rhineland-Palatinate. With our work, we take further steps towards hybrid intelligent support systems for emergency call handling and address current support systems' drawbacks [6].

The initial section of this work gives an overview of relevant foundations and related work. As a basis for our work, we will then analyze emergency call handling with a focus on the skill requirements of call-takers. Afterward, we will introduce ODD-ES and sketch exemplary integrations of AI methods that we expect to benefit call-takers in terms of their skill requirements. Further, we will sketch the mechanism for calculating experience-based reliability values for given inferences. We conclude this paper by summarizing our findings and providing an outlook on future work.

2. Foundations and Related Work

As for now, emergency call centers rarely use intelligent technologies to support emergency call-takers. Instead, they often rely on support systems that utilize handcrafted questionnaires designed by experts and of varying strictness to help call-takers structure the emergency call dialogue. These systems are generally deemed helpful but criticized for being too strict or leading to forgotten questions when following a paradigm with lower strictness [6]. Recent research addressed this issue by utilizing rule-based expert knowledge to deduce adaptive questionnaires, which produced promising results [7, 8]. To further support the skill requirements of call-takers with intelligent systems, few solutions exist. Recent research in this direction has led to a system that utilizes artificial neural networks to identify cardiac arrests based on textual representations of emergency calls [9]. While this has been shown to reduce recognition times for cardiac arrests, in practice, it results in a significant number of false alerts [10, 11]. To our knowledge, no work has yet been done on integrating this or other approaches with different AI methods in a hybrid intelligent system.

The idea of hybrid intelligence postulates that the proficient combination of human intelligence and AI can yield superior task performance due to synergies in complementary skills

[1]. Hybrid intelligent systems aim to foster these synergies in a close human-AI collaboration from which all actors constantly improve by mutual learning [1, 2]. In this context, humans are associated with skills like flexibility, creativity, empathy, and common sense, while AI adds fast, scalable, and consistent analytical abilities [12, 13]. How the resulting set of skills can be exploited best is addressed by various open research questions of hybrid intelligence [1]. Since humans in this context are seen as verifiers of AI-based inferences, their role is particularly decisive toward successful applications of hybrid intelligent systems [13, 14]. To fulfill their role, humans must know to what extent to trust AI-based inferences and when to overrule them [14]. To this end, the concept of appropriate reliance provides a means to measure the degree to which a human actor correctly resists an AI's opinion [15].

3. Analysis of Emergency Call Handling

In case of exceptional medical or firefighting-related situations, emergency numbers like 911 or 112 promise immediate access to qualified help from professionals. Calls to these numbers are answered by specially trained emergency call-takers that assess the caller's situation and decide on a suitable response. In high-urgency cases, this leads to dispatching the closest emergency resources like ambulances and emergency doctors. In Germany, emergency calls relevant to police duties are usually handled by call-takers in separate call centers with different emergency numbers. Although we do not regard this area of emergency call handling, this does not preclude the possibility that some aspects of this work could be transferred.

Møller et al. recently summarized their findings about handling medical emergency calls in a conceptual model, highlighting that contextual influences on callers and call-takers drive their course of execution [5]. Such influences arise, for example, from the ability of callers to observe and describe the situations to call-takers. On the other hand, call-takers are influenced by the degree of expertise they need to assess the callers' descriptions. This is reflected in the core of the conceptual model covering an iterative procedure executed by call-takers when handling an emergency call. Figure 1 displays this iterative procedure. Thereby, call-takers iteratively ask questions to obtain information from callers about the emergency. Afterward, they assess

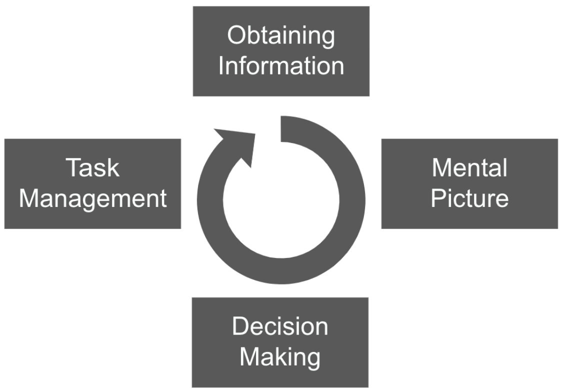

Figure 1: Iterative Procedure in Medical Emergency Call Handling based on Møller et al. [5]

the received information with their expertise to form a mental picture of the situation. This mental picture is the basis for decisions that may lead to tasks that call-takers must manage. An example of such a task would be dispatching the closest available emergency resource. If the emergency call center has dedicated dispatchers to perform this task, the call taker must manage the case handover instead.

Even though Møller et al. focused on medical emergency calls in their study, domain experts from the German state of Rhineland Palatinate confirmed in our qualitative interviews that the iterative procedure also reflects firefighting-related emergency call handling. In the context of our interviews, we picked up on the findings of Møller et al. [5] and focused on expanding their insights about the mental picture of call takers. In summary, call-takers address the following topics within their mental pictures:

Suspected Event: Call-takers require a suspicion of the happenings at the emergency site. In medical cases, this includes mapping reported symptoms to diagnoses. In the context of firefighting-related cases, observations are mapped to threats. This includes, e.g., mapping reported unusual smells to possible leakages of hazardous substances.

Risk Assessment: When having an impression about the reported event, call-takers have to assess the risks for patients and the population that arises from the situation. In the medical area, this leads to a triage of patients based on their symptoms and suspected diagnoses. When handling firefighting-related cases, risk assessment is often related to investigating the context of an event. In the case of a fire, this means identifying further combustible material in the immediate vicinity which threatens to catch fire.

Required Measures, Material and Resources: After understanding the current risks, call-takers must decide what needs to be done and by whom to solve the issue. This includes deciding about the required material for the effectuation of the measure. For example, in the medical area, a call-taker could define that a medication is required that can only be provided and administered by an emergency doctor and an emergency medical vehicle.

Ideal Available Resource: When having a general intention about what kind of resource should be appropriate to address the current situation, it is to be identified how and to what extent these requirements can be met in the current operational situation. This means selecting appropriate available resources while being aware of tactical considerations regarding the availability of emergency resources for upcoming incidents.

Call-takers need comprehensive skills in various areas to work on the listed topics during the iterative procedure. Fundamental to this, call-takers need extensive expertise in medical or firefighting contexts, e.g., to make decisions on the suspected event, risk assessment, and required measures. From the study of Møller et al. can further be derived that call-takers also draw their decisions from experiences in previous cases [5]. This helps them to assess a situation more accurately and to fine-tune their reactions. In the context of possibly frightened or uncooperative callers, call-takers also need empathy to recognize emotional states to guide them expediently. Due to wide-ranging influences on the exact characteristics of an emergency call, call-takers are regularly confronted with unique situations whose solution requires creativity.

This is, e.g., needed when a caller needs to be involved to quickly solve a problem like putting out a small fire. The following example illustrates that the application of skills is thereby closely interlinked: Given that a call-taker detects that a caller is calm and rational (empathy) in the face of a small fire that appears not to be related to hazardous substances (expertise), the caller could be asked to look out for a suitable source of water to put out the fire (creativity).

4. ODD-ES – Ontology- and Data-Driven Expert System

This section introduces ODD-ES – a process-oriented Ontology- and Data-Driven Expert System that aims at a foundation for a hybrid intelligent support system for emergency call handling. In the following, we will first introduce the basic concepts of ODD-ES and then elaborate on its metamodel, which poses a first step towards realizing the envisioned integration of human call-takers and AI.

4.1. Hybrid Intelligent Emergency Call Handling with ODD-ES

As a foundation for hybrid intelligent emergency call handling, ODD-ES provides concepts to combine human call-takers and AI in a close collaboration continuously improved by mutual learning. Figure 2 depicts the basic concepts of ODD-ES and shows how its application extends the iterative procedure of Møller et al. [5]. ODD-ES applies AI methods to assess emergency-

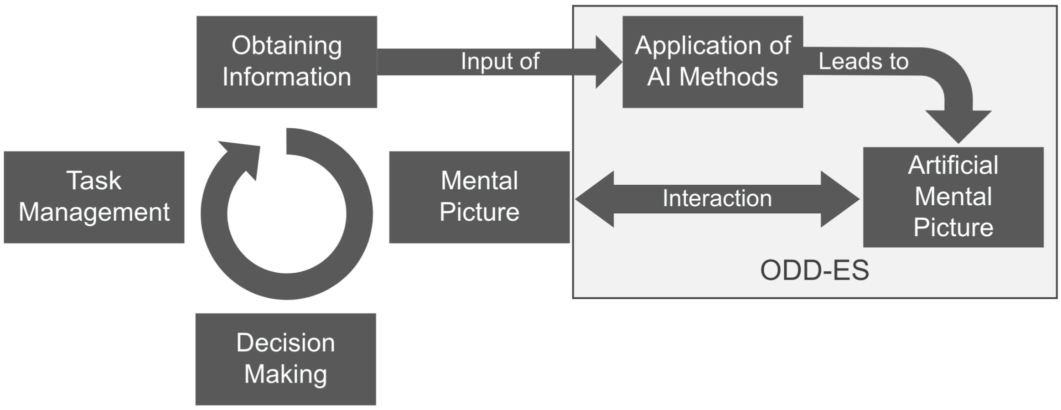

Figure 2: Basic Concepts of ODD-ES in Context of the Iterative Procedure from Møller et al. [5]

relevant information to infer an artificial mental picture. This artificial mental picture represents the view of integrated AI methods on the reported emergency and lays the foundation for a human-AI interaction driven by mental pictures. As mental pictures of call-takers and AI originate from diverging sets of skills, they are likely to portray different perspectives on the same case. The human-AI interaction is, therefore, primarily concerned with negotiating a shared view to find reasonable decisions. When making these decisions, however, it is down to the call-taker in the role of an AI verifier to either approve the view of AI or overrule it. Therefore, decisions are only made if the mental picture of the call taker has been altered in a way that allows for a specific decision. As AI-based inferences may have been generated from analytical skills that exceed the ones of call-takers, the need for an overview arises on whether AI-based results can be trusted. The mechanism sketched in this paper takes the first step to provide such insights.

Since emergency call handling spans a time-critical framework, ODD-ES also addresses optimizing the time required for human-AI interaction. Mutual learning of the two actors is used to achieve this. As an example, suppose that specific alterations in the mental picture of the call-taker occur regularly. In that case, call-takers will learn from the analytical skills of AI to some degree and eventually improve their analytical skills. On the other hand, if the call-taker gives the AI the feedback that it performs well, it could learn to make autonomous decisions in this area but within an ethically acceptable framework. Increased skills of call-takers could thereby help to identify outliers for an intervention. However, if the AI performs poorly in certain areas, it could restrain itself from communicating its opinion. Thus, the hybrid intelligent system relies more on the call-taker when handling emergency calls until its AI methods are updated.

4.2. Metamodel of ODD-ES

As a first step towards the envisioned integration of human call-takers and AI in ODD-ES, we have recently introduced a metamodel [3] that builds on multiple approaches of our Ontology- and Data-Driven Principle (ODD-Principle)[1]. One approach of the ODD-Principle that is essential to ODD-ES is the Ontology- and Data-Driven Business Process Model (ODD-BP) [7]. As ODD-BP is crucial for a thorough understanding of ODD-ES, we will first elaborate on the relevant aspects of ODD-BP and only afterward introduce the metamodel of ODD-ES.

ODD-BP provides a metamodel that forms the foundation for a data-driven process system based on semantic technologies. A significant difference between ODD-BP and typical process execution systems is that it aims at a semantic integration of all process-relevant knowledge into a unified knowledge base that is accessible for ontology-based reasoning. Following the idea of semantic process modeling [16], ODD-BP describes data flow relationships of tasks by linking them to data elements they consume or produce. This puts process-relevant data elements at the core of the resulting process models and enables them to drive the process execution. In this context, ODD-BP also includes concepts that enable ontology-based reasoners to infer the influence of data on the process execution and provide corresponding recommendations. In emergency call handling, we have already shown how ODD-BP can represent emergency relevant information and questions that may need to be asked to obtain them [3]. However, to identify the influences of process-relevant data on emergency call handling, domain knowledge is required that exceeds the expressiveness of semantic technologies. For this reason, ODD-ES was designed as an extension of ODD-BP.

ODD-ES achieves the required expressiveness for emergency call handling by extending ODD-BP with semantically modeled functions that a broad spectrum of AI methods can implement. A semantically modeled function interprets its implementation as a ‘black box’ about which primarily only its input and output parameters are known. These input and output parameters are a means for the function to tie in with the semantically modeled data elements of ODD-BP. Functions take data elements from ODD-BP as input or return them as their inferencing

[1]https://odd-principle.org

result. Returned inferencing results are then accessible to ODD-BP to handle their influence on the further process execution. Since ODD-ES aims to provide the basis for a mental picture-driven human-AI interaction, outputs are not always returned immediately as they may have to be verified by the call-taker. A detail to add in this context is that ODD-BP describes the data elements that are processed by functions in ODD-ES as attributes of dataobjects. While dataobjects represent relevant entities within a process (e.g., a patient), attributes are represented by individuals describing their characteristics (e.g., a reported symptom).

Figure 3 depicts how the described characteristics are picked up by the metamodel of ODD-ES and portrays how semantically modeled functions tie in with ODD-BP. Functions in

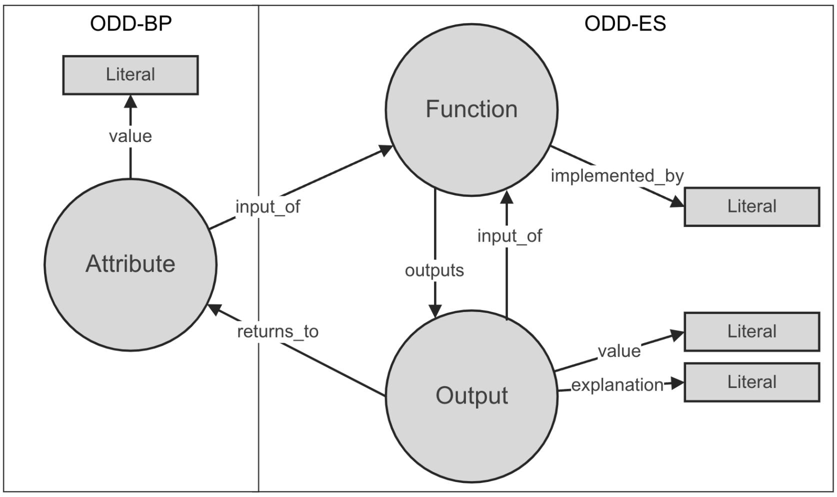

Figure 3: Metamodel of ODD-ES

ODD-ES are represented by individuals that point to external implementations of AI methods ($Function \xrightarrow{implemented\_by} Literal$) and are linked to attributes in ODD-BP to input their respective values ($Attribute \xrightarrow{value} Literal$, $Attribute \xrightarrow{input\_of} Function$). The results of a function's inferencing are returned into the knowledge base by adding them as values to linked dedicated output individuals ($Function \xrightarrow{outputs} Output$, $Output \xrightarrow{value} Literal$). These dedicated output individuals materialize the artificial mental picture in the knowledge base. Thus, the artificial mental picture is made up of the sum of all output individuals and their corresponding values and metadata like explanations ($Output \xrightarrow{explanation} Literal$). Output individuals are further linked to attributes in ODD-BP to which the resulting inferences are returned when verified by a call-taker ($Output \xrightarrow{returns\_to} Attribute$). Typing output individuals with classes from a domain ontology further defines the return values' general semantics. This allows, for example, to express that a function can raise a suspected diagnosis. Another feature of ODD-ES is that functions can be linked to each other to subsequently extend their outputs ($Output \xrightarrow{input\_of} Function$). Therefore, ODD-ES aims not only to integrate humans and AI but also to integrate various AI-Methods among each other, as this allows for drawing complex inferences from the combination of different AI methods.

5. Applying ODD-ES to Emergency Call Handling

In this section, we will identify AI methods that have the potential to support call-takers in meeting their skill requirements when handling emergency calls and sketch possible integrations into ODD-ES. Afterward, we design a mechanism to support call-takers in their judgment about how much to rely on AI methods when handling emergency calls.

5.1. Integration of AI Methods

As defined in section 3, handling medical or firefighting-related emergency calls requires a broad set of skills, ranging from domain-specific expertise and experience to creativity and empathy. Call-takers apply these skills to make decisions about various topics like a suspected diagnosis or an assessment of risks to patients or the population.

In the following, we will outline which AI methods could be integrated into ODD-ES to benefit call-takers in meeting their skill requirements. Thereby we build on a simplified visualization of ODD-ES shown in figure 4, in which functions are represented by large ellipses labeled with a rough description of their implementation. On the other hand, attributes and output individuals are represented by small ellipses with either dashed (attribute) or solid outlines (output). They are further labeled with the name of their domain-specific type and mapped to the left of a function if they are taken as input and to the right if they are returned as output. An interconnection of functions thus leads to a chain formation. Attributes to which outputs are returned are not shown for simplicity.

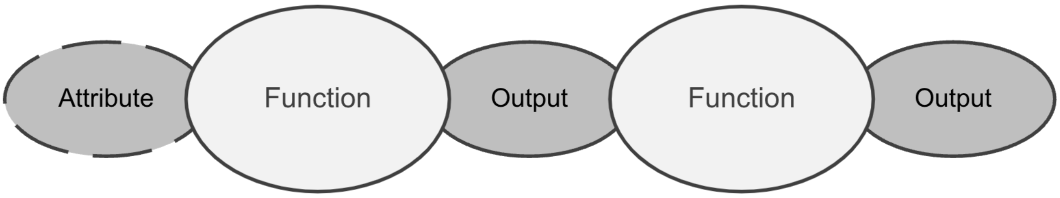

Figure 4: Simplified Visualization of ODD-ES Functions

An AI method that promises broad support for call-takers is rule-based reasoning. Rule-based reasoning essentially relies on if-clauses expressing a logical interconnection of parameters to derive an inference [17]. Rules can be based on different types of logic, whereby they can be either modeled manually or learned from example cases. Regarding emergency call handling, we have identified that modeled rules are particularly convenient to domain experts. They allow them to express and maintain their knowledge in an instruction-oriented, controllable, and understandable manner. Rules could address almost any topic of emergency call handling – from suspected events and risk assessment to required measures, materials, and resources. The integration of an example rule into ODD-ES is shown in figure 5. In this example, a patient's state of consciousness is linked to their breathing rate to suspect a cardiac arrest. Although this example rule seems comprehensible, a modeled rule may generally fail to detect a disease reliably in practice. In the context of such bad-performing rules, advanced call-takers would use their experience from other cases to identify when to contradict a rule-based inference. To

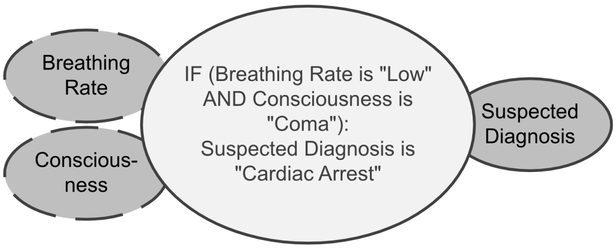

Figure 5: Integration of Rules into ODD-ES

allow novice call-takers to benefit from their colleagues' experience, the AI method of case-based reasoning provides a promising perspective. Case-based reasoning is a problem-solving methodology that reuses solutions from similar past experiences to solve the currently regarded problem [18]. Case-based reasoning has already been applied to improve inaccurate rule-based inferences [19]. With regard to establishing a suspected diagnosis, case-based reasoning could build a suspicion based on previous cases with similar symptomatology. Figure 6 sketches how the integration of case-based reasoning into ODD-ES could be realized. In this example, all data

Figure 6: Integration of Case-based Reasoning into ODD-ES

elements that could impact the suspected diagnosis are used as input for a function that draws conclusions based on similar cases and possibly regards an adaptation if necessary. Thereby, an application of case-based reasoning is not bound to suspected diagnoses, but we expect it can also be applied to other topics like required measures and possible risks.

Further support for call takers to fulfill their skill requirements can be found when considering the application of artificial neural networks. Artificial neural networks are designed to imitate biological learning activities [20] and are widely applied in the clinical context of emergency medicine [21]. However, only a few approaches are known in emergency call handling. One application can suspect a cardiac arrest based on a classification of a textual representation of the call [9]. An example integration of such an approach into ODD-ES is sketched in figure 7. As this approach is faster than call-takers in identifying a disease but results in a significant amount of false alerts when applied in practice [10, 11], we expect them to be especially helpful for call-takers to get a quick hint towards a specific diagnosis that has to be investigated further through other AI methods. In this context, we have already sketched a mechanism that helps call-takers to prioritize questions that would generate the information required by other AI methods [3].

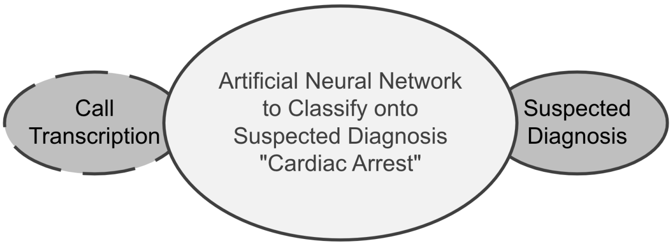

Figure 7: Integration of Artificial Neural Networks into ODD-ES

In order to identify the emotional state of a caller, call-takers have to apply their empathy during a sentiment analysis. Call-takers further have to identify the implications of an emotional state on emergency call handling. This is especially relevant with regard to the feasibility of creative solutions to a problem that may require that a caller is not panicking. In the context of automatic sentiment analysis, artificial neural networks are widely applied [22]. Figure 8 therefore sketches how artificial neural networks could be integrated into ODD-ES to support call-takers when using their empathy to detect the sentiment of a caller. Thereby, an artificial

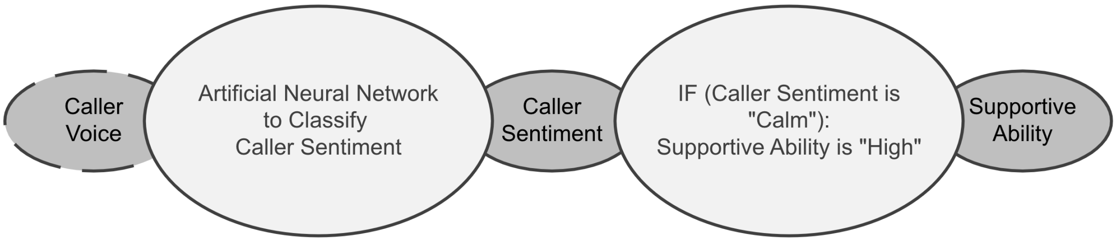

Figure 8: Integration of Artificial Neural Networks in Combination with Rules into ODD-ES

neural network performs a sentiment analysis based on the caller's voice. The output of this function represents the emotional state of the caller. To guide a call-takers application of creativity, this example further includes a rule to derive the extent to which this caller is capable of supporting the call-taker in quick out-of-the-box solutions for the reported problem. Therefore supportive abilities are inferred to be high if the caller's sentiment is detected to be calm.

Another area in which AI could support call-takers is the search for appropriate emergency resources. In this context, various dispatching policies can be applied to identify ideal emergency resources based on different tactical considerations [23, 24]. However, as none of the available search algorithms to implement these policies fits all situations, a system was designed that selects and applies an appropriate algorithm while further explaining its decisions [24]. Such a system could be integrated into ODD-ES as shown in figure 9. It would receive the required case data from the knowledge base, like the position of the emergency and a patient's condition to determine a proposal for an emergency resource. This proposal and its explanation would then be returned to ODD-ES as output and become part of the artificial mental picture.

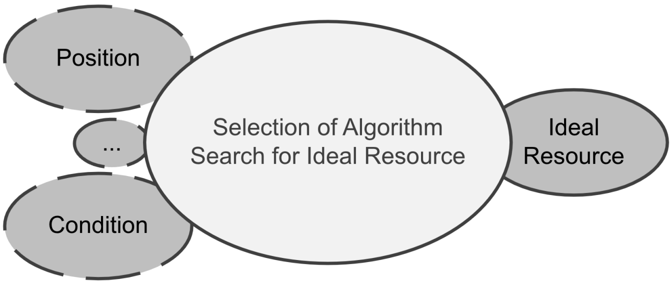

Figure 9: Integration of Search Algorithms into ODD-ES

In this section, we have outlined examples of how AI methods could be integrated into ODD-ES to support emergency call-takers concerning their skill requirements. Integrations were either stand-alone or in combination. If AI methods were combined, they either build consecutively on each other or were applied in parallel to the same topics in emergency call handling. When applied in parallel, AI-based inferences may diverge and possibly conflict. In the following, we design a mechanism that provides the basis for call-takers to resolve such conflicts.

5.2. Mechanism to Calculate Reliability Values

In ODD-ES, Call-takers are seen as AI verifiers. Consequently, they decide how far to trust AI-based inferences in situations that may impact patient outcomes or population safety. To support their decisions, a thorough analysis of the internal state of an AI method could be applied to provide indications about their reliability. For example, it is common to derive certainty values for inferences of artificial neural networks by analyzing the degree of activation in their output layer [25]. In rule-based reasoning, conversely, certainties are often managed based on the certainty factor model [26]. However, as ODD-ES considers integrated AI methods as black boxes, less informed approaches promise a better fit. Therefore we subsequently sketch a mechanism to calculate indications about the reliability of a function based on its correctness in similar past situations. To comply with the black-box approach of ODD-ES, the mechanism primarily regards a function's current and past input and output parameters. As our proposed mechanism is intended to apply to all integrated AI methods, it holds the potential for comparable results. We expect this may promote human-AI interaction in ODD-ES – an aspect that we will discuss more in detail after introducing the mechanism. Further, if only one mechanism is applied to all integrated AI methods to calculate reliances, only a single mechanism is required to find appropriate explanations.

To derive the extent to which a given inference can be considered reliable, our approach needs feedback on the correctness of a function's past inference results. One option to gain such feedback is deriving it implicitly from the call-takers' interaction with the system or explicitly from surveys after handling emergency calls. Alternatively, rescue workers on site could also provide this feedback, who usually have more precise information to verify the correctness of an inference. For example, this could be when an emergency doctor on the scene assesses a patient to establish a diagnosis. This diagnosis could be used to revise a suspected diagnosis made during the emergency call.

In the following, we will outline the proposed mechanism to calculate indications for the reliability of a function in ODD-ES. Case-based reasoning combined with a k-nearest neighbor classification will provide a basis for this. We primarily rely on these approaches as their results are easy to explain. Further, as call-takers are used to grounding their decisions on experiences in similar situations, we expect these approaches to feel familiar to them, possibly leading to increased trust in the system's assessments. Past situations are subsequently called 'cases'. In contrast, the current situation is called 'query'. While cases are stored in a case base (CB) specific to a given function in ODD-ES, queries are used to search the case base for similar cases from which a value for reliability can be derived. Queries and cases consist of an attribute-value-based description of the inputs and outputs of the current function for which reliability is to be calculated (see 1, 2). A case extends this description by indicating whether the output of the function proved correct in the context of the situation in the past emergency call (2). This correctness is a boolean value (3).

$$Query = (Inputs_q, Outputs_q) \tag{1}$$

$$Case = (Inputs_c, Outputs_c, Correctness_c) \tag{2}$$

$$Correctness_c \in \{True, False\} \tag{3}$$

To determine the most similar cases to a given query, a similarity measure maps the similarity between the query and a case concerning their respective inputs onto a value between 0 and 1 (4). Based on this, cases with a similarity above a certain threshold are retrieved from the case base (5).

$$sim(Query, Case) = sim(Inputs_q, Inputs_c) \in [0, 1] \tag{4}$$

$$retrieved_q = \{Case_i | Case_i \in CB \land sim(Query, Case_i) > threshold\} \tag{5}$$

The next step is to calculate the extent to which the correctness of a retrieved case can be transferred to the current situation. The result of this calculation is the transferrable correctness (TCorr) (6). Correctness is only transferable if the function's outputs in the case and query match. If this is not the case, the function leads to different results in similar situations. The transferrable correctness in such cases is set to 0. Similarly, the transferrable correctness is set to 0 in cases where the inference result was judged to be incorrect. However, if there is a positive correctness based on the same outputs, the correctness can be transferred to the extent that both cases are similar.

$$TCorr(Query, Case) = \begin{cases} sim(Query, Case) & \textbf{if } Outputs_c = Outputs_q \\ & \quad \land Correctness_c = true \\ 0 & \textbf{if } Outputs_c \neq Outputs_q \\ & \quad \lor Correctness_c = false \end{cases} \tag{6}$$

The transferrable correctness represents the foundation to calculate an indication for the reliability of the function's inference results in a current situation. The reliability is thereby calculated as the arithmetic mean of the transferrable correctness from the most similar cases retrieved from the case base (7).

$$reliability = \frac{\sum_{i=1}^{i=|retrieved_q|} TCorr(Query, Case_i)}{|retrieved_q|} \tag{7}$$

The described reliability calculation mechanism allows for an experience-based assessment of the inference performance of AI methods in ODD-ES. If the mechanism is applied with a uniformly defined similarity threshold, comparable reliabilities are created, regardless of the AI method that is being evaluated. Subsequently, we will discuss the potential of this comparability in the context of a human-AI interaction in ODD-ES. However, to provide a foundation, we must first extend the ODD-ES metamodel to include reliabilities in artificial mental pictures. For this purpose, as shown in Figure 10, output individuals are linked to calculated reliabilities via a new property called "reliability" ($Output \xrightarrow{reliability} Literal$).

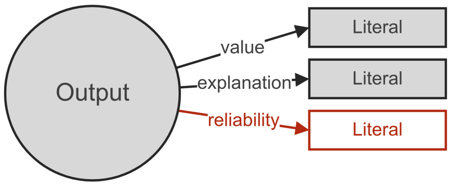

Figure 10: Extension of the ODD-ES Metamodel to Represent Reliabilities

The usefulness of comparable reliabilities becomes particularly clear in the context of a parallel application of different AI methods to the same inferencing issue. The previous section showed that applying multiple AI methods to single inferencing issues could benefit call-takers, for example, when establishing a suspected diagnosis. Thereby it could be possible that different methods come to diverging results. In such situations, it is up to the call-taker to resolve this issue. Comparable reliance values could here be beneficial to quickly identify a possible solution based on the degree to which an inference has been proven to be correct in the past. In addition, reliability thresholds could help resolve such a conflict automatically. If reliability is below such a threshold, the inference will not be displayed to the call taker or only with a low priority.

Although the proposed mechanism could assist in resolving conflicting inferences, a cold-start problem limits its supportive capabilities: Suppose there is a function with moderate performance and a new one that could perform significantly better, but there is no experience yet. In this case, a call-taker is likely to choose the mediocre function because the consequences of a possible poor performance of the new function could be severe. Therefore, using experiences as the only criterion for trust could block a positive development for patients and the population.

6. Future Work

The illustrated cold start problem of the proposed mechanism is a possible area for future work. Our exemplary integration of AI methods may provide the basis for a possible solution in this regard, as the respective methods were integrated with a complementary focus. While the artificial neural networks in our case could provide quick but maybe inaccurate suspected diagnoses, case-based reasoning was primarily integrated for correction. In this context, conflicting inferences could also be resolved based on this distribution of roles between AI methods. Another area for future work is the automation of decisions to resolve conflicts. Here we have only roughly outlined what effects threshold values could have. However, we have left open how exactly these threshold values are stored and how a call-taker will be informed about automation. Future work in this regard could deal with the design of user interfaces and the extension of the metamodels.

7. Conclusion

In this paper, we have taken further steps towards a hybrid intelligent support system for emergency call handling based on ODD-ES. We have defined exemplary integrations of AI methods and a mechanism to support call-takers in making decisions about their reliance on AI. We found that call-takers could benefit from both stand-alone and combined integration of AI methods concerning their skill requirements. In the case of a combined integration, we expect AI-based inferences may diverge and conflict as they are meant to correct ill-fitting inferences. In this context, the proposed mechanism seems to help resolve these differences, as it can lead to comparable reliance values, which probably allows call-takers to choose between inferences. However, the proposed mechanism suffers from a cold-start issue that will be the focus of our future work.

Acknowledgments

This work is funded by the Federal Ministry for Economic Affairs and Climate Action under grant No. 22973 SPELL.

References

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

[2] D. Dellermann, A. Calma, N. Lipusch, T. Weber, S. Weigel, P. Ebel, 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, ScholarSpace, 2019.

[3] C. Maletzki, E. Rietzke, R. Bergmann, 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, volume 3242 of *CEUR Workshop Proceedings*, 2022, pp. 79–89.

[4] F. N. Ford, Decision Support Systems and Expert Systems: A Comparison, Information & Management 8 (1985) 21–26. doi:10.1016/0378-7206(85)90066-7.

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

[6] A. Baumann, Leitlinien oder Protokolle zur Notrufabfrage, in: A. Hackstein, H. Sudowe (Eds.), Handbuch Leitstelle, volume 2, 2017, pp. 189–193.

[7] E. Rietzke, C. Maletzki, R. Bergmann, N. Kuhn, Execution of Knowledge-Intensive Processes by Utilizing Ontology-Based Reasoning: ODD-BP: An Ontology- and Data-Driven Business Process Model, Journal on Data Semantics 10 (2021) 3–18. doi:10.1007/s13740-021-00127-w.

[8] E. Rietzke, Modellierung, Visualisierung und Ausführung wissensintensiver Prozesse unter Verwendung Semantischer Technologien, Doctoral thesis, Universität Trier, 2021. doi:10.25353/ubtr-xxxx-ebc1-ec41.

[9] S. N. Blomberg, F. Folke, A. K. Ersbøll, H. C. Christensen, C. Torp-Pedersen, M. R. Sayre, C. R. Counts, F. K. Lippert, Machine learning as a supportive tool to recognize cardiac arrest in emergency calls, Resuscitation 138 (2019) 322–329. doi:10.1016/j.resuscitation.2019.01.015.

[10] F. Byrsell, A. Claesson, M. Ringh, L. Svensson, M. Jonsson, P. Nordberg, S. Forsberg, J. Hollenberg, A. Nord, Machine learning can support dispatchers to better and faster recognize out-of-hospital cardiac arrest during emergency calls: A retrospective study, Resuscitation 162 (2021) 218–226. doi:10.1016/j.resuscitation.2021.02.041.

[11] S. N. Blomberg, H. C. Christensen, F. Lippert, A. K. Ersbøll, C. Torp-Petersen, M. R. Sayre, P. J. Kudenchuk, F. Folke, Effect of Machine Learning on Dispatcher Recognition of Out-of-Hospital Cardiac Arrest during Calls to Emergency Medical Services: A Randomized Clinical Trial, JAMA Network Open 4 (2021). doi:10.1001/jamanetworkopen.2020.32320.

[12] W. M. P. van der Aalst, Hybrid Intelligence: to automate or not to automate, that is the question, International Journal of Information Systems and Project Management 9 (2021) 5–20. doi:10.12821/ijispm090201.

[13] D. Dellermann, P. Ebel, M. Söllner, J. M. Leimeister, Hybrid intelligence, Business & Information Systems Engineering 61 (2019) 637–643. doi:10.1007/s12599-019-00595-2.

[14] P. Hemmer, M. Schemmer, M. Vössing, N. Kühl, 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, 2021.

[15] M. Schemmer, P. Hemmer, N. Kühl, C. Benz, G. Satzger, Should I Follow AI-based Advice? Measuring Appropriate Reliance in Human-AI Decision-Making (2022). doi:10.48550/arXiv.2204.06916.
[16] O. Thomas, M. Fellmann, Semantic Process Modeling – Design and Implementation of an Ontology-based Representation of Business Processes, Business & Information Systems Engineering 1 (2009) 438–451. doi:10.1007/S12599-009-0078-8.
[17] C. Grosan, A. Abraham, Intelligent Systems: A Modern Approach, Springer Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-21004-4.
[18] A. Aamodt, E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, AI Communications 7 (1994) 39–59. doi:10.3233/AIC-1994-7104.
[19] A. R. Golding, P. S. Rosenbloom, Improving Rule-Based Systems through Case-Based Reasoning, in: T. L. Dean, K. R. McKeown (Eds.), Proceedings of the 9th National Conference on Artificial Intelligence, Anaheim, CA, USA, July 14-19, 1991, Volume 1, AAAI Press / The MIT Press, 1991, pp. 22–27.
[20] C. C. Aggarwal, Neural Networks and Deep Learning: A Textbook, Springer Cham, 2018. doi:10.1007/978-3-319-94463-0.
[21] B. Mueller, T. Kinoshita, A. Peebles, M. A. Graber, S. Lee, Artificial intelligence and machine learning in emergency medicine: a narrative review, Acute Medicine & Surgery 9 (2022). doi:10.1002/AMS2.740.
[22] A. Ligthart, C. Catal, B. Tekinerdogan, Systematic reviews in sentiment analysis: a tertiary study, Artificial Intelligence Review 54 (2021) 4997–5053. doi:10.1007/S10462-021-09973-3.
[23] C. S. Lim, R. Mamat, T. Bräunl, Impact of Ambulance Dispatch Policies on Performance of Emergency Medical Services, IEEE Transactions on Intelligent Transportation Systems 12 (2011) 624–632. doi:10.1109/TITS.2010.2101063.
[24] M. Reuter-Oppermann, S. Morana, P. Hottum, Towards Designing an Assistant for Semi-Automatic EMS Dispatching, in: T. Bui (Ed.), Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS), Waikoloa, HI, USA, January 4-7, 2017, ScholarSpace / AIS Electronic Library (AISeL), 2017, pp. 3556–3565.
[25] A. Mandelbaum, D. Weinshall, Distance-based Confidence Score for Neural Network Classifiers (2017). doi:10.48550/ARXIV.1709.09844.
[26] D. Heckerman, The Certainty-Factor Model, in: Encyclopedia of Artificial Intelligence, Second Edition, Wiley, New York, 1992, pp. 131–138.