ASCODI: An XAI-based interactive reasoning support system for justifiable medical diagnosing

Dominik Battefeld[1,*], Felix Liedeker[2], Philipp Cimiano[2] and Stefan Kopp[1]

[1]Social Cognitive Systems Group, CITEC, Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany
[2]Semantic Computing Group, CITEC, Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany

Abstract

Research has shown that approximately 10% of medical diagnoses are wrong. As a direct consequence, appropriate medical treatment may be delayed or even absent, leading to an increased burden on patients, increased costs for medication, or even harm and death. As cognitive biases contribute to roughly three-quarters of these diagnostic errors, a lot can be gained from increasing a physician's reasoning quality during the diagnostic process. Clinical decision support systems (CDSS), leveraging recent advances in artificial intelligence (AI) and insights from eXplainable AI (XAI), aim at providing accurate predictions and prognosis paired with corresponding *ex-post* explanations that make the reasoning of the system accessible to humans. Viewing explanations as involving the interactive construction of shared belief, we propose to move from diagnostic decision support to reasoning support which, in its true sense, needs to tailor the timing and content of generated explanations to the state of the reasoning process of physicians to meet their information needs and effectively mitigate the influence of cognitive biases. We claim that, given the uncertain and incomplete information inherent to medical diagnosis, the most effective way not to fall prey to cognitive reasoning errors is to establish and maintain proper justification for each decision throughout the diagnostic process. This paper contributes (1) a conceptual model and desiderata for AI-based interactive reasoning support that enhances reasoning quality through increased justification at every stage of the process, and (2) preliminary work on the development of the **as**sistive, **co**-constructive **di**fferential **di**agnosis system, **ASCODI**, which provides reactive as well as proactive reasoning support to improve the justification of actions taken and decisions made during and after medical diagnosing. We also present selected use cases of ASCODI concerning its application in supporting the diagnosis of transient loss of consciousness and highlight their connection back to the theoretical concepts established.

Keywords

Interactive CDSS, Reasoning Support, Differential Diagnosis, Diagnosis Justification

1. Introduction

Medical diagnoses are key to managing and curing diseases [1], and hinge on a solid categorization of a patient's symptoms. If done incorrectly, selected categorizations and subsequent ineffective treatment based on it may lead to systemic issues like unnecessary test procedures [2]

ECAI'24: Workshop on Multimodal, Affective and Interactive eXplainable AI (MAI-XAI), October 20, 2024, Santiago de Compostela, Spain

[*]Corresponding author.
dbattefeld@techfak.uni-bielefeld.de (D. Battefeld); fliedeker@techfak.uni-bielefeld.de (F. Liedeker); cimiano@techfak.uni-bielefeld.de (P. Cimiano); skopp@techfak.uni-bielefeld.de (S. Kopp)
0000-0002-5480-0594 (D. Battefeld); 0009-0006-2556-9430 (F. Liedeker); 0000-0002-4771-441X (P. Cimiano); 0000-0002-4047-9277 (S. Kopp)

or serious patient harm by missing so-called red-flag indications of severe diseases [3]. But given the overwhelming number of diseases to consider, diagnostic tests to conduct, and questions to ask [4], medical diagnostic reasoning over inherently uncertain information is a challenging and error-prone task. As many diagnostic errors arise from faulty information processing and synthesis [5], clinical decision support systems (CDSS) with their contemporary focus on machine learning and AI have the potential to increase diagnostic accuracy and thus patient safety [4]. But given the high stakes of diagnostic decisions, generating an explanation for uncertain predictions within CDSS should be mandatory [6], lifting the targeted *decision* support to actual *reasoning* support [7]. However, a systematic review by Antoniadi et al. found "an overall distinct lack of application of XAI in the context of CDSS and, in particular, a lack of user studies exploring the needs of clinicians" [6, p. 1]. This is, in turn, problematic, because cognitive factors contribute to roughly three-quarters of diagnostic errors [5] which is framed as "the challenge of cognitive science for medical diagnosis" [8]. Over the years, several de-biasing techniques have been studied to mitigate the influence of cognitive reasoning errors by enforcing metacognition [9], checklists [10] or rule of thumbs [11], but their effectiveness in empirical studies remains mixed [12].

This paper proposes to combine approaches to decision/reasoning support and de-biasing strategies in a bias-aware interactive reasoning support system. The system aims to increase the objective justification of each action taken or decision made during diagnosing by monitoring the reasoning process of its user in the background. Leveraging these insights enables the system to provide reactive as well as proactive reasoning support tailored to the information needs of its user at any given point within their reasoning process. In this paper, we present our conceptual understanding of interactive reasoning support through six theoretical desiderata and ongoing work on their implementation in the interactive reasoning support system ASCODI, designed for neurologists diagnosing patients with transient loss of consciousness.

2. Related Work

Even before the rise of XAI, the potential of CDSS to improve diagnosis was extensively discussed in the literature [13]. At the same time, however, the usage of CDSS has been limited and still is underwhelming, mainly due to challenges arising from the integration into everyday clinical practice [13, 4]. In some cases, the potential is completely misjudged and CDSS are even perceived as a threat to their job by physicians [14], making the development of co-constructive, "doctor-in-the-loop" [15, p. 2] support systems even more important.

In addition, it has been argued both in general [16] and for the special case of AI-based decision support [17] that inherently interpretable *white box* models (Bayesian Networks [18], Decision trees [19], etc.) are more preferable than *black box* models (Deep Neural Networks [20], etc.). A general need for explainability, as well as interpretability of machine learning applications, has furthermore been recognized [21]. Nonetheless, only a small number of CDSS have been developed with a focus on explainability [6]. Most of the current state-of-the-art systems rely on black box models with post-hoc explanations [22].

Many current applications of XAI in the field of medical diagnosis revolve around medical image analysis, e.g. the detection of COVID-19 from X-ray images [23]. In this domain, feature attribution methods highlighting input features most relevant to the instance that should be explained, are particularly widespread [24]. Outside of image analysis, counterfactual explanations (CFs) are widely used because they are close to human-like approaches to explaining [25]. CFs provide a hypothetical, yet similar, counterexample that would change the prediction of a machine learning model, thereby explaining its decision. In practice, different CDSS have utilized CFs, e.g. [26]. A growing number of researchers recently outlined the importance of human factors in XAI design along with the development of human-centered XAI applications [27, 15, 28] - including CDSS [29]. In this context, there's also an increased effort to use interactive and "co-constructive" approaches to explanation processes [30], which has led to the development of multiple interactive XAI tools [31, 32, 33].

The human-centered side of XAI is also influenced by the cognitive biases of physicians that have repeatedly been shown to accompany medical diagnostic reasoning [9, 34, 35, 36, 11]. Here, an anchoring bias (clinging to a hypothesis despite contradictory information), an availability bias (generating hypotheses that readily come to mind), a confirmation bias (neglecting possibly contradictory information during information exploration), premature closure (ending information exploration too early), and overconfidence (perceiving evidential strength as higher than it is) are prominent and prevalent examples [34]. Despite these more commonly known instantiations, Croskerry identified a total of 50 cognitive biases associated with diagnosis and medical practice in general [37]. Graber et al. showed that these biases affect 74% of all pathologically confirmed diagnostic errors under review [5].

Today, minimizing the influence of cognitive errors is seen as a major lever to improve the quality of care [4, 8] and ease the burden on the medical system [38]. Multiple solutions have been proposed over the years from checklists during information exploration [10] to metacognitive approaches like the "dramatic big five", which formalize diagnostic options for acute thoracic pain [11], inducing critical thinking [12] or cognitive forcing strategies as algorithmic step-by-step guide [9], technological interventions like providing visual interpretations of statistics [12], and motivational interventions by holding people personally accountable [12]. Their promised observable benefits remain mixed, where checklists improve reasoning on difficult cases but worsen it on simpler ones [10], trying to reflect out of self-induced motivation fails to translate into a sustainable effect [39] and leads to an improvement in only 50% of cases [12].

We argue that de-biasing strategies will become more effective if they are designed to align with the already existing psychological reasoning process of physicians. Checklists and differential diagnosis generators may present as helpful external sources of information, but as many errors remain of cognitive origin, increasing the cognitive load on physicians by adding even more uncertain sources of information will not suffice. The ASCODI system is thus not designed to provide support *for a decision* of a physician but co-constructively form an explainable diagnosis *with* a physician, taking a collaborative part throughout the process to reduce instead of increase cognitive load.

3. Interactive Reasoning Support: Desiderata

We conceive of interactive reasoning support for medical diagnosing as an ongoing interaction of proactive and reactive actions from both the physician and the support system. To sharpen this concept theoretically, we propose six general desiderata as guiding principles: (1) synchronizing the observable reasoning state between physician and system on a shared dashboard, (2) monitoring and inferring the hidden cognitive reasoning state of the physician, (3) supporting an informed choice of reasoning direction by the physician based on their information needs, (4) providing feedback on the chosen reasoning direction, (5) suggesting alternative reasoning directions, (6) displaying warnings in case of potential pitfalls.

Desideratum 1: Shared Dashboard Physician and system should collaborate on a shared dashboard that summarizes the observable state of the reasoning process, i.e. which information about the patient is already known and how this information has already been integrated to form all currently active hypotheses. The idea behind the dashboard draws from the concept of extended cognition [40], where physical information-storing entities like written formulas or diaries can play an active part in a cognitive process, and aims to establish sufficient grounding [41] about the reasoning state between both interaction parties.

Desideratum 2: Monitored Reasoning The system should monitor the behavior of its user to detect the hidden state of the reasoning process, i.e. not explicitly stated assumptions or cognitive biases. This detection can be realized by building a computational cognitive process model [42] that formalizes how the dynamically changing mental state of the user leads to observable actions. Subsequently inverting the process model to predict mental states from observable information queries and hypothesis revisions via inverse planning [43] enables the detection of cognitive biases. This is a crucial capability of interactive reasoning support because the system can only tailor explanations towards the mitigation of specific reasoning errors if the presence of errors is acknowledged in the first place.

Desideratum 3: Information Needs Physicians should be able to express information needs targeting the diagnostic problem or the interactive system itself and receive an appropriate response from the system, i.e. "Which symptoms cannot be explained by our current hypotheses?" or "Why do you judge hypothesis *X* as less likely than me?". Multiple studies have found and categorized information needs within the diagnostic process, e.g. [44, 45]. Further information needs arise from the interaction with an AI system itself and the requirement to understand the system's actions and decisions [21]. Satisfaction of these needs through justified explanations that physicians can integrate into their reasoning will lead to increased problem understanding and this in turn will lead to increased reasoning quality [7].

Desideratum 4: Behavioral Feedback Physicians should receive feedback on their reasoning behavior, e.g. a visual marker to display (dis-)agreement with the current state of each hypothesis, whose validity can be justified on demand by the system. This leads to an active exchange of arguments where both the physician and the system align their assessment of the current situation to combine the capabilities of AI in statistical reasoning and those of physicians in constructing and contextualizing a "picture of the patient" [7].

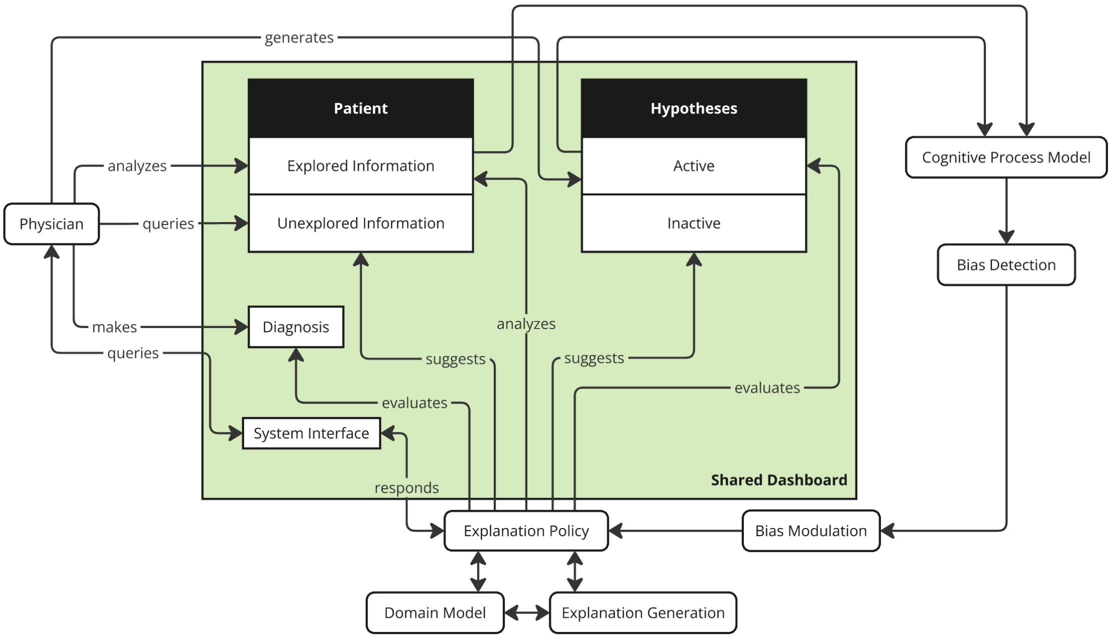


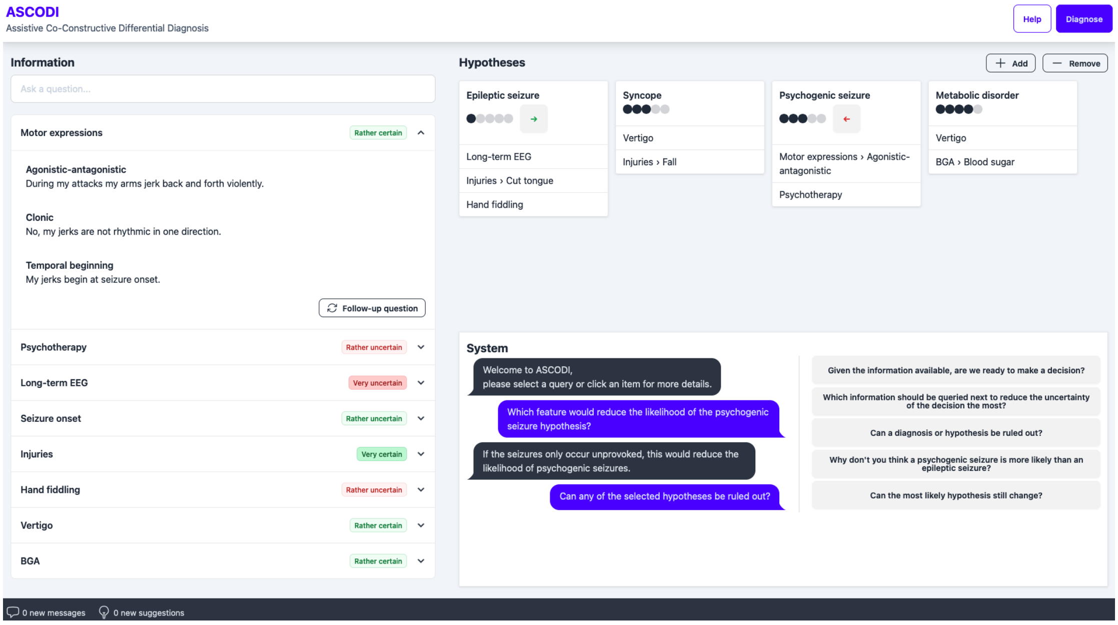
Figure 1: A schematic view on the interactive reasoning support for medical diagnosing within ASCODI.

Desideratum 5: Situational Suggestions Physicians should receive justified suggestions for their reasoning behavior, i.e. unknown information to explore or hypotheses to generate. Productivity in interaction thrives if both parties proactively participate [46] and by suggesting alternative reasoning directions, insufficiently justified restrictions of options as observable for a confirmation bias (neglecting possibly contradictory information) or an anchoring bias (clinging to a hypothesis despite contradictory information) can be effectively mitigated.

Desideratum 6: Red Lines Physicians should be warned clearly when crossing red lines within the reasoning process, e.g. trying to reject a viable diagnostic hypothesis or trying to commit to an insufficiently justified diagnosis. While it would be legally and ethically infeasible to set up red lines as impossible to overrule because physicians are the ones carrying the legal and epistemic responsibility for the outcome of the reasoning process and its implications, expressing the system's clear disagreement with the decision is expected to trigger reflection. By adding this extra reasoning step the decision - even if unchanged - is promoted to a product of consciously selected thought.

4. System Description: ASCODI

We present the assistive co-constructive differential diagnosis system (ASCODI) to implement the above-described general desiderata. A schematic overview of the architectural layout is shown in Figure 1. ASCODI provides interactive reasoning support for experienced neurologists during the diagnosis of patients suffering from transient loss of consciousness (TLOC) [47].



Agonistic-antagonistic
During my attacks my arms jerk back and forth violently.

Clonic
No, my jerks are not rhythmic in one direction.

Temporal beginning
My jerks begin at seizure onset.

Welcome to ASCODI,
please select a query or click an item for more details.

Which feature would reduce the likelihood of the psychogenic seizure hypothesis?

If the seizures only occur unprovoked, this would reduce the likelihood of psychogenic seizures.

Can any of the selected hypotheses be ruled out?

Given the information available, are we ready to make a decision?

Which information should be queried next to reduce the uncertainty of the decision the most?

Can a diagnosis or hypothesis be ruled out?

Why don't you think a psychogenic seizure is more likely than an epileptic seizure?

Can the most likely hypothesis still change?

Figure 2: The user interface of ASCODI with its three main components for information exploration (left), hypothesis revisions (top right), and system interaction (bottom right).

Diagnosis of TLOC places a high emphasis on the cognitive categorization and contextualization capabilities of physicians because the presence or absence of specific symptoms cannot warrant a decision towards or against any differential diagnosis and information is mainly obtained through subjective, personal dialogue rather than objective test results [48] - rendering it an ideal application area. Within the user interface of ASCODI (see Figure 2), physicians can work on one case at a time by exploring information about the patient, generating and refining hypotheses about the presence or absence of specific diseases, and receiving reactive as well as proactive reasoning support.

Clinical Vignettes Medical cases within ASCODI are stored as pre-defined clinical vignettes. A clinical vignette is a set of variable-value pairs for each queryable information ranging from biographic information and medical history to current complaints and results of medical examinations. All vignettes are based on publicly available patient data [49], published case reports [50, 51, 52], and confidential electronic health records of patients at the Ruhr-Epileptology in Bochum. Each piece of information is associated with a certainty on a 4-point Likert scale and a patient response that formulates the value of the variable in natural language. All vignettes consist of the same set of variables with values chosen to fit the particular medical case. The variables are ordered hierarchically in a tree-like graph structure, i.e. the patient experiencing pain is a parent node while the location and the duration of that pain are two child nodes. Information exploration thus means to traverse the patient graph.

Shared Dashboard Each diagnostic process within the ASCODI system is carried out on a shared dashboard (see Figure 2). The design of the dashboard is based on a monitoring tool to elicit step-by-step cognitive process trajectories of physicians for empirical data collection [53]. The dashboard summarizes the current state of the reasoning process, i.e. already explored information and the current state of all hypotheses. Every diagnosis starts with a short self-report of the patient about their sex, age, and initial complaints. Then, the physician can prompt the patient for biographic information, medical history, current complaints, and other typical anamnesis questions, conduct sophisticated examinations, and request lab reports by entering the name of a specific variable as stored in the clinical vignette. All of this information can be requested at any time and in any order. Explored information is displayed in semantic categories where child nodes of the root in the patient graph are top-level entries and detailed information (e.g. the location and duration of experienced pain) are grouped in the corresponding entry. Given each entry, physicians can issue follow-up questions on specific symptoms by querying child nodes of already explored variables to exhaustively explore their manifestation. Variables can also be queried by aliases to account for potential synonyms or slight differences between names of the same variable and suggestions are displayed based on the text entered to ensure that users find what they are looking for.

During information exploration, physicians can generate new hypotheses, adjust the certainty of existing ones, and reject unreasonable hypotheses. By default, all possible hypotheses are viewed as inactive until the user explicitly generates them. Each hypothesis is understood as an argument towards or against the presence of a disease. Thus, the user not only has to generate potential diagnoses but also connect the already explored information with each hypothesis via drag-and-drop, which creates isolated disease arguments visible to the user and the system.

Monitored Reasoning The ASCODI system constantly monitors information exploration and hypothesis revisions by the physician to infer the mental state of its user and detect potential cognitive biases. This inference is based on a computational cognitive process model that - given a clinical vignette and a cognitive bias as input - aims to replicate empirical reasoning trajectories of physicians, i.e. sequences of information queries and hypothesis updates. The dynamic problem of repeated exploration and opinion revision is formalized as a Markov decision process (MDP) [54] where the summed subjective strength of each hypothesis constitutes the reward signal from which the action policy is derived via Monte Carlo tree search (MCTS) [55]. This strength is defined as the Bayesian posterior probability of the hypothesis mediated by the amount of anticipated regret [56] that grows with the severity of the disease. Aligned to the dual process theory in which human thinking [57], including medical diagnostic reasoning [58], is divided into a fast, associative route (System I) and a slow, deliberate route (System II), the posterior computation is based on the observable patient information as hard evidence and assumed information that is intuitively derived from previous experience as soft evidence. Experience is formalized based on the MINERVA-DM model [59, 60] that captures past events (i.e. patients) as memory traces (i.e. vectors) whose similarity to the current observed event enables computations on assumed symptom likelihoods.

Once this process model is fully implemented, we aim to run it exhaustively beforehand, to collect a sufficient amount of labeled data to train a classifier that inverts the generation of reasoning trajectories given a clinical vignette and a cognitive bias to infer cognitive biases given the reasoning trajectory. We aim to model an anchoring bias (clinging to a hypothesis despite contradictory information), an availability bias (generating hypotheses that readily come to mind), a confirmation bias (neglecting possibly contradictory information during information exploration), premature closure (ending information exploration too early), and overconfidence (perceiving evidential strength as higher than it is), which are prominent examples repeatedly found in empirical research on cognitive biases in diagnostic reasoning [34, 5, 61, 62, 36]. Detected cognitive biases are then forwarded to the explanation policy that tailors the timing and content of explanations to the user.

Explanation Policy & Domain Model The explanation policy incorporated into the ASCODI system ensures the most appropriate explanation for a given situation is used. 'Most appropriate' explanation refers to the determination of (1) the type of the explanation, e.g. counterfactual (2) its form, e.g. written in the system chat, and (3) timing of the explanation-giving, e.g. instantly. The main purpose of fine-tuning an explanation is to mitigate detected biases that are fed into the explanation policy by the cognitive process model. In addition, the explanation policy is responsible for the tracking of previous user interactions. In particular, previously issued queries and the corresponding explanations given by the ASCODI system are stored to adapt the following explanations to the user based on the interaction history.

A three-layer Bayesian network (BN) [63] serves as the domain model for explanation generation within the ASCODI system. The BN is trained on data provided by Wardrope et al. [49] as well as annotated and anonymized outpatient letters from the Ruhr-Epileptology in Bochum. A major advantage of a BN is its inherent interpretability and the ability to directly model causal relationships between variables [63]. The feasibility of a similar domain model for explanation generation in a CDSS has been demonstrated in a previous prototype [64]. Additionally, the modular structure of the system renders the domain model interchangeable both in terms of the algorithmic approach (e.g. BNs vs. Neural Networks) and the disease domain (e.g. Neurology vs. Cardiology).

Explanations that are part of the ASCODI system can be divided into four groups based on the scope and the algorithms used to calculate the explanation: Suggestions of the next features to be queried by the physician are determined by maximizing mutual information [65]. Explanations for queries covering the likelihood of events (e.g. "What are likely diagnoses for a patient with symptoms u, v, w?") are calculated via Bayesian inference in our domain model. Queries posing *What if*-questions (e.g. "If the patient would have diagnosis X, what other symptoms would they have?") are answered by counterfactual explanations, and explanations that aim to justify (e.g. "Why do you judge X different from me?") rely on our definition of justification that is outlined below.

A formal definition of justification is not only relevant for answering user queries but crucial for providing meaningful feedback or suggestions to the physician and ensuring compliance with red lines. We propose a combination of a relevance measure of the explanation with a measure for the degree of explored information to be used as justification. The rationale behind this is, that for a decision to be justified, it must be relevant and at the same time it must be certain, i.e. only made if supported by sufficient evidence. The Most Relevant Explanation

Information Need	Example Query	Implementation
Diagnostic procedure	What information should be queried next?	Mutual information
Differential diagnosis	Could this patient have condition X? What are likely diagnoses for a patient with symptoms u, v, w?	Bayesian Inference
Disease complication	If the patient with symptoms u, v, w has disease X, what other signs of symptoms would they have?	Counterfactuals
System behavior	Why do you judge hypothesis X as less likely than me?	Justification

Table 1
Information needs with corresponding query examples and the underlying implementation in ASCODI. Note: The list of example queries is not exhaustive.

(MRE) [66] method is utilized as our relevance measure for generated explanations. MRE is the partial instantiation of all possible diagnoses which maximizes the generalized Bayes factor. The requirements regarding the necessary amount of explored information are preliminarily implemented as thresholds for the ratio of explored patient features that are highly correlated to the current hypotheses under consideration.

Information Needs The ASCODI system includes predefined user queries to satisfy the information needs that physicians encounter during the interactive reasoning support. Information needs defined in the literature [44, 67] as well as the most common questions asked by physicians [45, 68] can be grouped into three categories based on the source of the information need: Diagnosis, patient information, and treatment. Though treatment options including the prognosis of patients are beyond the scope of ASCODI and hence these information needs are not considered here. Information needs concerning patient information are not realized by explicit queries, but rather via the iterative exploration of the clinical vignette. However, the usage of the system itself results in further information needs [21]. Table 1 provides a summary of all information need categories that are part of ASCODI along with their operationalization as queries and their implementation. The queries are based on the prior theoretical work of Van Baalen et al. on diagnostic reasoning support systems [7].

System Responses The interaction between ASCODI and a physician is twofold: On the one hand, a reactive part acts as the direct response to user-triggered queries and, on the other hand, a proactive part can intervene at any point if the system detects a situation that requires action to support the reasoning process or prevent pitfalls.

Users can trigger queries not only via the system interface but also at the 'point of interaction' within the system. For example, the query "Why do you judge hypothesis X as less likely than me?" can be called via the system interface, but is also available in the dashboard by clicking on hypothesis X and requesting the explanation of why the system assigns hypothesis X a lower likelihood than the user does.

Whereas explanations to user queries are always answered in the system chat, interventions by the proactive system part are presented in different formats and at different locations within the shared dashboard to be able to draw the user's attention to certain areas and have a further lever for mitigating detected biases. Furthermore, the strength of system responses can be modulated, by using different forms of explanations and notifications, to attract the user's attention in different scenarios.

5. Reasoning Support Use Cases

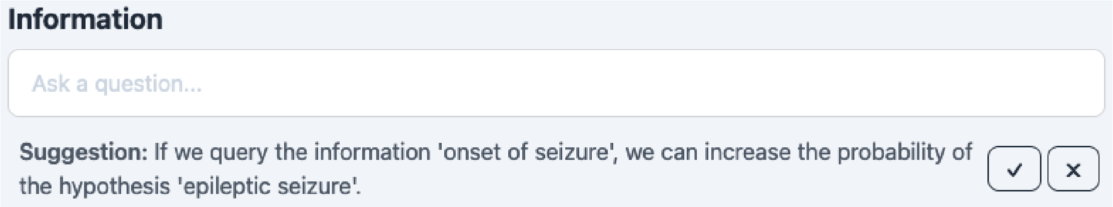
As described above in detail, ASCODI's reasoning support capabilities range from reactive explanations of user-triggered events to proactive advice to mitigate cognitive biases and reasoning pitfalls. To illustrate the system usage more vividly and highlight how it links back to the theoretical desiderata for interactive reasoning support, we present four selected use cases in which ASCODI (1) answers a user-triggered query to the system, (2) reacts to the user-triggered belief change in one hypothesis, (3) suggests an alternative direction during information exploration and (4) displays a warning for an unjustified diagnosis.

Query Response As outlined before, ASCODI is capable of providing explanations for user-triggered queries to satisfy the information needs arising in the diagnostic reasoning process (desideratum 3). An example of this user interaction can be seen in the bottom right of Figure 2. In the given example, the user selected the query "Which feature would reduce the likelihood of the psychogenic seizure hypothesis?" via the system interface and received the written explanation in the system chat.

Behavioral Feedback ASCODI constantly monitors how the user explores information and ranks hypotheses. According to the available and explored information, the belief of the domain model is updated and the system itself assigns probabilities to all hypotheses. A possible discrepancy between the user's and the system's ranking of hypotheses is displayed in the dashboard according to the explanation policy and the current state of the user. The 'weakest' form of this notification is realized by green (red) arrows within the hypotheses panel indicating a higher (lower) ranking by the system as shown in Figure 2.

If the explanation policy determines that the user is in a state that does require a more vehement notification, e.g. because previously made suggestions were ignored, further notifications can be sent to the system chat. A further notification in the system chat includes an explanation of why the system ranks a hypothesis higher than the user. An example of such an explanation of why the system would rank a hypothesis differently is shown in the system chat in Figure 2.

Situational Suggestion The ASCODI system can proactively suggest information queries at any point within the reasoning trajectory (see Figure 3). Suggestions are attached to an explanation, that motivates why a certain query is deemed reasonable at this point (e.g. to increase the certainty of a hypothesis). Suggestions and explanations align to potentially detected cognitive biases (e.g. suggesting falsifying information for a hypothesis if a potential confirmation bias is detected). The same mechanism is employed to propose overlooked yet feasible hypotheses or already falsified alternatives. This capability links back to three desiderata, where the system first observes the current state of the reasoning process through the shared



Information

Suggestion: If we query the information 'onset of seizure', we can increase the probability of the hypothesis 'epileptic seizure'.

Figure 3: A proactively suggested query during information exploration. Suggestions can be accepted or rejected by the user.

dashboard, then infers the possibility of a confirmation bias through the monitored reasoning, and then provides an alternative reasoning direction by suggesting unexplored information. The idea behind this suggestion is to not tell a biased person that they're biased but to provide reasonable information that induces reflection.

Diagnosis Safeguard Once a physician attempts to submit a final diagnosis, ASCODI checks its validity and will display a warning if the decision is unjustified, e.g. if there is a viable alternative hypothesis not even considered. By explicitly mentioning that the seizure of a patient could also resemble a psychogenic instead of an epileptic seizure and by pointing out that it would be worthwhile to explore the psychological history of the patient more closely to increase the strength of differentiation between these two diagnostic options, ASCODI implements four desiderata: observing the end of the diagnostic process on the dashboard, detecting overconfidence within the user, displaying a warning as a red flag to show major disagreement and suggesting an alternative reasoning direction by specifying which options are deemed insufficiently explored. The motivation behind this behavior is that it takes more cognitive effort to insist on a decision although someone else explicitly told you they would have decided differently which again is hypothesized to induce reflection.

6. Conclusion

This paper presented a conceptual understanding of interactive reasoning support for medical diagnosis in the form of six desiderata as well as **ASCODI**, an **as**sistive, **co**-constructive **di**fferential **di**agnosis system, whose architecture draws from this theoretical foundation to build a system that (1) keeps track of its user's reasoning process to detect potential cognitive biases, (2) tailors the timing and content of its explainable responses to the current external and cognitive state of the reasoning process and (3) proactively takes part in knowledge exploration and hypothesis revision by providing explainable suggestions and safeguards. To become an active part of its user's reasoning process, ASCODI utilizes XAI-based explanation generation techniques to derive justifiable opinions and responses [7]. While the implementation of ASCODI is still ongoing and subsequent empirical validation is lacking, we believe that support systems like ASCODI are the logical next step to enhance the reasoning quality of physicians through justifiable decisions along each path chosen. In a time when diagnostic errors are mainly caused by faulty information processing and synthesis [5], such systems present themselves as a promising direction for future work.

Acknowledgments

Funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation): TRR 318/1 2021 – 438445824.

References

[1] N. Donner-Banzhoff, Die ärztliche Diagnose: Erfahrung - Evidenz - Ritual, Programmbereich Medizin, 1. auflage ed., Hogrefe, Bern, 2022.

[2] E. S. Berner, M. L. Graber, Overconfidence as a Cause of Diagnostic Error in Medicine, The American Journal of Medicine 121 (2008) S2–S23.

[3] H. C. Sox, M. C. Higgins, D. K. Owens, Medical decision making, 2nd ed ed., John Wiley & Sons, Chichester, West Sussex, UK : Hoboken, New Jersey, 2013. Medium: electronic resource.

[4] M. L. Graber, Reaching 95%: decision support tools are the surest way to improve diagnosis now, BMJ Quality & Safety 31 (2022) 415–418. Publisher: BMJ Publishing Group Ltd Section: Editorial.

[5] M. L. Graber, N. Franklin, R. Gordon, Diagnostic Error in Internal Medicine, Archives of Internal Medicine 165 (2005) 1493.

[6] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, C. Mooney, Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review, Applied Sciences 11 (2021) 5088. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[7] S. Van Baalen, M. Boon, P. Verhoef, From clinical decision support to clinical reasoning support systems, Journal of Evaluation in Clinical Practice 27 (2021) 520–528.

[8] P. Croskerry, S. G. Campbell, D. A. Petrie, The challenge of cognitive science for medical diagnosis, Cognitive Research: Principles and Implications 8 (2023) 13.

[9] P. Croskerry, The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them, Academic Medicine 78 (2003) 775–780.

[10] T. Shimizu, K. Matsumoto, Y. Tokuda, Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis, Medical Teacher 35 (2013) e1218–e1229.

[11] M. Gäbler, Denkfehler bei diagnostischen Entscheidungen, Wiener Medizinische Wochenschrift 167 (2017) 333–342.

[12] R. Ludolph, P. J. Schulz, Debiasing Health-Related Judgments and Decision Making: A Systematic Review, Medical Decision Making 38 (2018) 3–13.

[13] K. K. Hall, S. Shoemaker-Hunt, L. Hoffman, S. Richard, E. Gall, E. Schoyer, D. Costar, B. Gale, G. Schiff, K. Miller, T. Earl, N. Katapodis, C. Sheedy, B. Wyant, O. Bacon, A. Hassol, S. Schneiderman, M. Woo, L. LeRoy, E. Fitall, A. Long, A. Holmes, J. Riggs, A. Lim, Making Healthcare Safer III: A Critical Analysis of Existing and Emerging Patient Safety Practices, Agency for Healthcare Research and Quality (US), Rockville (MD), 2020.

[14] C. Krittanawong, The rise of artificial intelligence and the uncertain future for physicians, European Journal of Internal Medicine 48 (2018) e13–e14.

[15] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller, Causability and explainability of artificial intelligence in medicine, WIREs Data Mining and Knowledge Discovery 9 (2019) e1312.

[16] C. Rudin, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nature machine intelligence 1 (2019) 206–215.

[17] R. L. Pierce, W. Van Biesen, D. Van Cauwenberge, J. Decruyenaere, S. Sterckx, Explainability in medicine in an era of AI-based clinical decision support systems, Frontiers in Genetics 13 (2022) 903600.

[18] J. Pearl, Causality, Cambridge University Press, 2000.

[19] J. R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106.

[20] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444. Publisher: Nature Publishing Group.

[21] F. Doshi-Velez, B. Kim, Towards A Rigorous Science of Interpretable Machine Learning, 2017. ArXiv:1702.08608 [cs, stat] version: 2.

[22] V. Arya, R. K. E. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović, S. Mourad, P. Pedemonte, R. Raghavendra, J. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, Y. Zhang, One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques, 2019. ArXiv:1909.03012 [cs, stat].

[23] S. Rajpal, N. Lakhyani, A. K. Singh, R. Kohli, N. Kumar, Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images, Chaos, Solitons & Fractals 145 (2021) 110749.

[24] H. Singh, A. N. D. Meyer, E. J. Thomas, The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations, BMJ Quality & Safety 23 (2014) 727–731.

[25] S. Wachter, B. Mittelstadt, C. Russell, Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR, 2018. ArXiv:1711.00399 [cs].

[26] Z. Wang, I. Samsten, P. Papapetrou, Counterfactual Explanations for Survival Prediction of Cardiovascular ICU Patients, in: A. Tucker, P. Henriques Abreu, J. Cardoso, P. Pereira Rodrigues, D. Riaño (Eds.), Artificial Intelligence in Medicine, volume 12721, Springer International Publishing, Cham, 2021, pp. 338–348. Series Title: Lecture Notes in Computer Science.

[27] M. Ribera, A. Lapedriza, Can we do better explanations? A proposal of User-Centered Explainable AI, Joint Proceedings of the ACM IUI 2019 Workshops (2019).

[28] K. Sokol, P. Flach, One Explanation Does Not Fit All: The Promise of Interactive Explanations for Machine Learning Transparency, KI - Künstliche Intelligenz 34 (2020) 235–250.

[29] T. A. J. Schoonderwoerd, W. Jorritsma, M. A. Neerincx, K. van den Bosch, Human-centered XAI: Developing design patterns for explanations of clinical decision support systems, International Journal of Human-Computer Studies 154 (2021) 102684.

[30] K. J. Rohlfing, P. Cimiano, I. Scharlau, T. Matzner, H. M. Buhl, H. Buschmeier, E. Esposito, A. Grimminger, B. Hammer, R. Häb-Umbach, I. Horwath, E. Hüllermeier, F. Kern, S. Kopp, K. Thommes, A.-C. Ngonga Ngomo, C. Schulte, H. Wachsmuth, P. Wagner, B. Wrede, Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design

of AI Systems, IEEE Transactions on Cognitive and Developmental Systems 13 (2021) 717–728. Conference Name: IEEE Transactions on Cognitive and Developmental Systems.

[31] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viegas, J. Wilson, The What-If Tool: Interactive Probing of Machine Learning Models, IEEE Transactions on Visualization and Computer Graphics (2019) 1–1. ArXiv:1907.04135 [cs, stat].

[32] T. Spinner, U. Schlegel, H. Schäfer, M. El-Assady, explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning (2019).

[33] H. Baniecki, D. Parzych, P. Biecek, The grammar of interactive explanatory model analysis, Data Mining and Knowledge Discovery (2023).

[34] G. Saposnik, D. Redelmeier, C. C. Ruff, P. N. Tobler, Cognitive biases associated with medical decisions: a systematic review, BMC Medical Informatics and Decision Making 16 (2016) 138.

[35] Z. I. Vally, R. A. Khammissa, G. Feller, R. Ballyram, M. Beetge, L. Feller, Errors in clinical diagnosis: a narrative review, Journal of International Medical Research 51 (2023) 03000605231162798.

[36] J. S. Blumenthal-Barby, H. Krieger, Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy, Medical Decision Making 35 (2015) 539–557.

[37] P. Croskerry, 50 Cognitive and Affective Biases in Medicine, https://sjrhem.ca/wp-content/uploads/2015/11/CriticaThinking-Listof50-biases.pdf, 2015. Accessed: 2024-09-06.

[38] Hardeep Singh, Gordon D Schiff, Mark L Graber, Igho Onakpoya, Matthew J Thompson, The global burden of diagnostic errors in primary care, BMJ Quality &amp; Safety 26 (2017) 484.

[39] D. M. Berwick, Not again!, BMJ 322 (2001) 247–248.

[40] A. Clark, D. Chalmers, The Extended Mind, Analysis 58 (1998).

[41] H. H. Clark, S. E. Brennan, Grounding in communication., in: Perspectives on socially shared cognition., American Psychological Association, Washington, 1991, pp. 127–149.

[42] J. B. Jarecki, J. H. Tan, M. A. Jenny, A framework for building cognitive process models, Psychonomic Bulletin & Review 27 (2020) 1218–1229.

[43] C. Baker, R. Saxe, J. Tenenbaum, Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution, Proceedings of the Annual Meeting of the Cognitive Science Society (2011).

[44] R. N. Jerome, N. B. Giuse, K. Wilder Gish, N. A. Sathe, M. S. Dietrich, Information needs of clinical teams: analysis of questions received by the Clinical Informatics Consult Service, Bulletin of the Medical Library Association 89 (2001) 177–185.

[45] J. W. Ely, J. A. Osheroff, P. N. Gorman, M. H. Ebell, M. L. Chambliss, E. A. Pifer, P. Z. Stavri, A taxonomy of generic clinical questions: classification study, BMJ (Clinical research ed.) 321 (2000) 429–432.

[46] R. E. v. Geffen, Proactivity in concert: an interactive perspective on employee proactivity, Library of the University of Amsterdam, 2018. OCLC: 1049936204.

[47] T. Baumgartner, R. Surges, Synkope, epileptischer oder psychogener Anfall? Der Weg zur richtigen Diagnose, DMW - Deutsche Medizinische Wochenschrift 144 (2019) 835–841.

[48] K. Malmgren, M. Reuber, R. Appleton, Differential diagnosis of epilepsy, Oxford textbook of epilepsy and epileptic seizures (2012) 81–94.

[49] A. Wardrope, J. Jamnadas-Khoda, M. Broadhurst, R. A. Grünewald, T. J. Heaton, S. J.

Howell, M. Koepp, S. W. Parry, S. Sisodiya, M. C. Walker, M. Reuber, Machine learning as a diagnostic decision aid for patients with transient loss of consciousness, Neurology: Clinical Practice 10 (2020) 96–105.

[50] S. A. Haji Seyed Javadi, F. Hajiali, M. Nassiri Asl, Zolpidem Dependency and Withdrawal Seizure: A Case Report Study, Iranian Red Crescent Medical Journal 16 (2014).

[51] B. Hellmich (Ed.), Fallbuch Innere Medizin, 6 ed., Georg Thieme Verlag, Stuttgart, 2020. Pages: b-007-170975.

[52] R. Gerlach, A. Bickel, Fallbuch Neurologie, Fallbuch, 5., unveränderte auflage ed., Georg Thieme Verlag, Stuttgart New York, 2021.

[53] D. Battefeld, S. Mues, T. Wehner, P. House, C. Kellinghaus, J. Wellmer, S. Kopp, Revealing the Dynamics of Medical Diagnostic Reasoning as Step-by-Step Cognitive Process Trajectories, in: Proceedings of the Annual Meeting of the Cognitive Science Society, Rotterdam, The Netherlands, 2024.

[54] R. S. Sutton, A. Barto, Reinforcement learning: an introduction, Adaptive computation and machine learning, second edition ed., The MIT Press, Cambridge, Massachusetts London, England, 2020.

[55] M. Świechowski, K. Godlewski, B. Sawicki, J. Mańdziuk, Monte Carlo Tree Search: a review of recent modifications and applications, Artificial Intelligence Review 56 (2023) 2497–2562.

[56] M. Zeelenberg, R. Pieters, A Theory of Regret Regulation 1.0, Journal of Consumer Psychology 17 (2007) 3–18.

[57] K. Frankish, Dual-Process and Dual-System Theories of Reasoning, Philosophy Compass 5 (2010) 914–926.

[58] P. Croskerry, A Universal Model of Diagnostic Reasoning, Academic Medicine 84 (2009) 1022–1028.

[59] M. R. P. Dougherty, C. F. Gettys, E. E. Ogden, MINERVA-DM: A memory processes model for judgments of likelihood., Psychological Review 106 (1999) 180–209.

[60] R. P. Thomas, M. R. Dougherty, A. M. Sprenger, J. I. Harbison, Diagnostic hypothesis generation and human judgment., Psychological Review 115 (2008) 155–185.

[61] T. Watari, A. Gupta, Y. Amano, Y. Tokuda, Japanese Internists' Most Memorable Diagnostic Error Cases: A self-reflection Survey, Internal Medicine (2023) 1494–22.

[62] M. F. Loncharich, R. C. Robbins, S. J. Durning, M. Soh, J. Merkebu, Cognitive biases in internal medicine: a scoping review, Diagnosis 0 (2023).

[63] J. G. Richens, C. M. Lee, S. Johri, Improving the accuracy of medical diagnosis with causal machine learning, Nature Communications 11 (2020) 3923.

[64] F. Liedeker, P. Cimiano, A Prototype of an Interactive Clinical Decision Support System with Counterfactual Explanations, in: Proceedings of the xAI-2023 Late-breaking Work, Demos and Doctoral Consortium co-located with the 1st World Conference on eXplainable Artificial Intelligence (xAI-2023), 2023.

[65] F. Liedeker, P. Cimiano, Dynamic Feature Selection in AI-based Diagnostic Decision Support for Epilepsy, 2023. Poster presented at the 1st International Conference on Artificial Intelligence in Epilepsy and Neurological Disorders, Breckenridge, CO, USA.

[66] C. Yuan, H. Lim, T.-C. Lu, Most Relevant Explanation in Bayesian Networks, J. Artif. Intell. Res. (JAIR) 42 (2011) 309–352.

[67] J. A. Osheroff, D. E. Forsythe, B. G. Buchanan, R. A. Bankowitz, B. H. Blumenfeld, R. A. Miller, Physicians' information needs: analysis of questions posed during clinical teaching, Annals of Internal Medicine 114 (1991) 576–581.

[68] Y.-H. Seol, D. R. Kaufman, E. A. Mendonça, J. J. Cimino, S. B. Johnson, Scenario-based assessment of physicians' information needs, Studies in Health Technology and Informatics 107 (2004) 306–310.