From clinical decision support to clinical reasoning support systems.

Sophie van Baalen[1], Mieke Boon[2], and Petra Verhoef[1]

[1]Rathenau Instituut
[2]University of Twente

September 11, 2020

Abstract

Despite the great promises that artificial intelligence (AI) holds for health care, the uptake of such technologies into medical practice is slow. In this paper, we focus on the epistemological issues arising from the development and implementation of a class of AI for clinical practice, namely clinical decision support systems (CDSS). We will first provide an overview of the epistemic tasks of medical professionals, and then analyse which of these tasks can be supported by CDSS, while also explaining why some of them should remain the territory of human experts. Clinical decision-making involves a reasoning process in which clinicians combine different types of information into a coherent and adequate 'picture of the patient' that enables them to draw explainable and justifiable conclusions for which they bear epistemological responsibility. Therefore, we suggest that it is more appropriate to think of a CDSS as clinical reasoning support systems (CRSS). Developing CRSS that support clinicians' reasoning process therefore requires that: 1) CRSSs are developed on the basis of relevant and well-processed data; and 2) the system facilitates an interaction with the clinician. Therefore, medical experts must collaborate closely with AI experts developing the CRSS. In addition, responsible use of an CRSS requires that the data generated by the CRSS is empirically justified through an empirical link with the individual patient. In practice, this means that the system indicates what factors contributed to arriving at an advice, allowing the user (clinician) to evaluate whether these factors are medically plausible and applicable to the patient. Finally, we defend that proper implementation of CRSS allows combining human and artificial intelligence into hybrid intelligence, were both perform clearly delineated and complementary empirical tasks. Whereas CDRSs can assist with statistical reasoning and finding patterns in complex data, it is the clinicians' task to interpret, integrate and contextualise.

Keywords:

Artificial Intelligence, Machine Learning, CDSS, Clinical Decision-Making, Clinical Reasoning, Epistemological Responsibility

Introduction

Artificial intelligence (AI) holds great promises for health care, according to developers, policy makers and medical professionals. It is expected to improve health care by alleviating workload of care workers, improving the quality of decision-making or improving the efficiency of health care. Hence, it is often presented as a solution to deal with the challenges faced by health care in the (near) future.11See for example the white paper issued by the European Commission in February 2020, which in its first sentence states that AI "will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases)" and healthcare is repeatedly mentioned as sector that will benefit greatly from AI. The introduction of AI systems to medical practice is one aspect of the increasing digitization of society. Responsible digitization of society and the medical domain requires that the consequences for specific practices and people are carefully considered and taken into account at an early stage of the development. Public values such as equity and equality, privacy, autonomy and human dignity must be safeguarded. In addition, citizens and practitioners must be enabled to develop the skills needed to deal with the new tasks and responsibilities associated with digital technologies.[1,2]Our paper focuses on this last point, namely the epistemological issues arising from the development and implementation of AI technologies (particularly clinical decision support systems, CDSS) in clinical diagnostic practices, and their implications for the epistemic tasks and responsibilities of health-care professionals.22Because the introduction of AI poses many ethical, regulatory, technological, medical, legal and organizational challenges for medical practice, the Dutch Rathenau institute has asked (through a series of blog posts) several relevant players in the field of Dutch health care and innovation (i.e. government, developers, entrepreneurs, lawyers and scientists) to share their view on the responsible innovation of AI for health care.: https://www.rathenau.nl/nl/maakbare-levens/kunstmatige-intelligentie-de-zorg-samen-beslissen-blijkt-de-crux. In addition to challenges related to the safe (i.e. taking into account the privacy and other fundamental right of patients) collection, sharing, saving and use of medical data they identify opportunities and challenges that concern the implementation of AI systems in health care practices, such as fitting AI into specific clinical situations, and training (future) medical professionals to critically reflect on their use of such technologies.

Although research in CDSS is developing rapidly, the uptake of such technologies into medical practice is slow.[3,4]Kelly et al. (2019) show that this is partly, due to the fact that clinical evaluation through randomized controlled trials (as the gold standard for evidence generation) through machine learning is not always appropriate or feasible. Furthermore, the metrics for *technical accuracy* used in machine learning studies often do not reflect metrics used in robust clinical evaluation, which essentially includes quality of care and patient outcomes.[3] Greenes et al. (2018) provide an overview of the factors that need to be considered to overcome challenges related to the implementation of computer-based CDSS, namely: how systems are integrated into the clinical workflow; how the output of a CDSS is represented to the user and (intended to be) used for cognitive support; how the systems can be implemented legally and institutionally; how the quality and the effectiveness of a systems can be evaluated; and how the cognitive tasks of medical professionals can be supported.[4] In this paper, we focus on one of these factors: what cognitive tasks can be supported by CDSS, and how? More specifically, our question is how CDSS impacts the epistemic activities of (a team of) medical professionals, who have the task of determining a diagnosis and a strategy for cure or care based on heterogeneous information (from different sources) about a patient. To answer this question, we will first provide an overview of the epistemic tasks of medical professionals in performing these clinical tasks. Then, we analyse which of the epistemic tasks can be supported by computer-based systems, while also explaining why some of them tasks should remain the territory of human experts.

Applications of CDSS

CDSS is a class of computer and AI-based systems that is designed as a tool to support clinical decision-making by medical professionals or patients. More technically, CDSSs are 'active knowledge systems which use two or more items of patient data to generate case-specific advice'.[5] There are many different types of CDSS which provide different types of support to different kinds of decision-making processes in a variety of clinical situations, ranging from providing alerts or reminders for example while monitoring patients, emphasizing clinical guidelines during care, identify drug-drug interactions, or, advise on possible diagnosis or treatment plans.[6]Regarding diagnosis and treatment, CDSS can have many functions, such as predicting the outcome of a specific treatment, image interpretation (i.e. contouring, segmentation or pathology detection), prescribing (the dosage of) medication, and screening and prevention.[7]In performing these kinds of epistemic tasks, a CDSS uses artificial intelligence to 'reason' according to its algorithms about a specific patient by comparing that patient's data with the data in its system. CDSSs are primarily designed to mimic reasoning by medical professionals, but faster, less prone to human error or cheaper.[6] The rules that the CDSS follows to reason about a specific patient are either programmed by the developers (i.e. 'knowledge' or 'rule-based' expert systems), or inferred from a large amount of data about a group of patients, using statistical AI methods, such as machine learning or deep learning algorithms (i.e. ‘data-driven’).[8,9,10]

Preventing risks of CDSS by better understanding cognitive tasks

However, there are several potential risks associated with the introduction of CDSS in clinical practice, which were reviewed in a recent report.[6] Because the clinical decisions made by healthcare professionals have consequences for the wellbeing of patients, risks associated with the uses of CDSS are substantial and undesirable. These risks can be classified into: 1) risk related to the ‘datafication’ of medical information; 2) control that is transferred from humans to machines; 3) the lack of a human element, and 4) the changing division of labour.[6] An important aspect of each of these risks is that cognitive tasks, which are usually performed by medical professionals who bear the responsibility to perform these tasks to the best of their knowledge and ability,[11]are now delegated to machines. Therefore, to deal with the risks associated with the implementation of CDSS, it is crucial to understand how the use of a CDSS will impact the daily practice of medical professionals (i.e. clinicians) – more specifically, to understand the cognitive tasks involved in decision-making on diagnosis and treatment.

Overview

In this paper, we will argue that CDSS can potentially support clinical decision-making, but that this poses specific requirements on the CDSS as well as on the (training of) cognitive abilities of the professionals using the CDSS.

In Section 2, we will analyse the epistemic tasks in clinical decision-making and suggest that human and artificial intelligence each have different capacities to fulfil specific kinds of epistemic tasks. In order to achieve a high quality decision-making process for diagnosis and treatment of patients, human and artificial intelligence should complement each other in performing these epistemic tasks. For example,*knowledge-based CDSSs* , on the one hand, can function as an automated ‘handbook’ that efficiently supports searches by clinicians.*Data-driven CDSSs* , on the other hand, may identify patterns in data that are inaccessible to humans or detect similarity of data patterns among patients, thus providing a diagnosis and suggesting a possible treatment.[3,8,10,12,13] Clinicians, in turn, deal with individual patients, and will diagnose based on existing data and their experience. They will find the most suitable treatment taking into account both the diagnosis, the personal situation of the patient, and the local situation of the hospital. In arriving at a suitable treatment, they may also consult colleagues and deliberate with them. In other words, the CDSS makes a proposal for treatment based on the diagnosis only, i.e., without taking into account the specific context of the patient. We will conclude that, when using a computer-based CDSS, clinicians have an epistemological responsibility to *collect* ,*contextualize* and *integrate* all kinds of clinical data and medical information about an individual patient similar to when using*evidence based medicine.* [11,14]

Section 3 elaborates on what is needed for good use of computer-based CDSSs in clinical practice. We suggest that, since clinical decision-making involves a complex and demanding cognitive process for which they bear ultimate responsibility, it is more appropriate to think of a CDSSs as a clinical *reasoning support* system (CRSS) rather than a *decision support* system. Based on this analysis, some suggestions can be made on what this implies for the collaborations of clinicians and CRSSs. We will conclude that for CRSSs this means that: 1) CRSSs are developed on the basis of relevant and well-processed data, the preparations of which requires human expertise; 2) the system facilitates an interaction with the clinician, allowing the clinician to ask questions that a CRSS answers and thereby also providing some insight into how the answer is created; and 3) there is a clear empirical relationship between the data generated by the CRSS and the information of the individual patient, providing *empirical justification* for the use of the CRSS in reasoning about that patient. Conversely, clinicians must have cognitive skills to perform epistemic tasks that cannot performed by the CRSS (such as collecting, contextualizing and integrating data on individual patients) and to understand the (CRSS supported) clinical reasoning for each specific patient to the extent that they can still take responsibility for the outcome.

In Section 4, finally, we will defend that proper implementation of CRSS allows clinicians to combine their (human) intelligence with the artificial intelligence of the CRSS into *hybrid intelligence* , in which both have clearly delineated and complementary tasks. We will sketch out how the epistemic tasks can be divided between the clinician and the system, based on their respective capacities. CRSS, for example can assist in cognitive tasks that humans are notoriously bad at, such as the statistical reasoning, or finding patterns in complex data. The task of clinicians is to incorporate the outcomes of CRSS into medical reasoning, by asking questions that the machine (CRSS) can answer, and by interpreting, integrating and contextualizing the outcome of the system. We conclude that the configuration such a hybrid intelligence poses requirements on the side of the CRSS as well as the clinician.

Epistemic tasks in clinical decision-making

The goal of clinical decision-making is to compose a diagnosis and treatment plan that is suitable to the patients' personal situation, signs and symptoms and based on relevant and reliable evidence. Computer-based clinical decision support systems (CDSSs) are expected to improve clinical decision-making by making it faster, cheaper, less prone to human errors or more precise.[6,12] In practice, clinician and computer can complement each other, each having different capacities to perform crucial but different epistemic tasks that together add up to a diagnosis or treatment plan. In order for CDSS to support clinical decision-making, the capacities of human and artificial intelligence need to be maximally utilized and aligned to each other. First, we will analyse which epistemic tasks can be better done by CDSS, and which by clinicians.

Clinical decision support systems

CDSS makes use of artificial intelligence (AI) that is designed to mimic or improve clinical decision-making. Two broad categories of AI uses in CDSS are usually distinguished:[6,8,10]'knowledge-based' AI (also called rules-based expert systems[9]) and data-driven AI. Knowledge-based AI systems have been in use since the late 1970's, and aim to replicate human decision-making by programming the rules experts employ when they make decisions in their field in computational terms.[10] As such, a knowledge-based system can best be thought of as a database of 'best-practice' rules that can be employed to find the most suitable procedure (e.g. examination or treatment) for an individual patient.[9] The 'logic' employed by the system can be represented as formal rules, such as "when a patient with disease X also has symptom Y, use medication Z." As such, the 'reasoning' employed by the system to arrive at a specific advice, can easily be backtracked and evaluated.

The *data-driven* use of AI has developed significantly over the last decade, and employs statistical machine learning algorithms to abstract patterns from large amounts of data. In the so-called supervised machine learning to develop a CDSS, the machine learning system is fed with a large amount of data about a group of patients labelled with the clinical diagnosis by medical professionals, the so-called 'training dataset'. In this learning-phase, the CDSS learns to 'recognize' the patterns (represented by a 'model') in the training-set that fit best with the correct diagnoses. When a new case is entered into the system, it will use the patterns that it has inferred from the 'training-set' to make a prediction about an individual case.[10] The 'logic' employed by this type of CDSS is (rather than rule-following as in knowledge-based CDSS), based on comparisons between cases, such as "other patients with disease X and symptom Y have benefited from using medication Z".[9] Because data-driven CDSSs are often trained using data from thousands of cases or more, a multitude of the amount of cases that a physician sees in a lifetime, these systems are able to detect very subtle and complex patterns in the data (e.g. Savage 2020[12]). However, unlike knowledge-based AI, the decision made in a data-driven CDSS cannot easily be explained,[6] which leads to critical questions about the robustness, explainability, reliability and accountability of these types of systems.[10]

Epistemic tasks by CDSSs: statistical reasoning and pattern recognition

Knowledge-based systems can be thought of as a database of best practice in terms of rules, such as evidence-based guidelines. The advantage of an automated system is that it can use the patient's individual characteristics to find the most suitable guidelines and procedures. Data-driven systems do not use this type of rule-following, but have other capacities. Boon (2020) has analysed the epistemic tasks that machine-learning algorithms are capable of doing. According to her categorisation of epistemic tasks, machine-learning algorithms can *match* input data (e.g., an image or a set of data points such a clinical signs and symptoms) with similar cases in their database; *interpret* input data as belonging to a specific category, defined by humans or by a machine-learning algorithm;*diagnose* a set of input data as probably belonging to a certain class and from that infer other properties of the target;*structure* large amounts of data to find patterns, correlations and causal relations; *calculate* in a way that outperforms humans; and *simulate* complex dynamic process.[15] In short, computers outperform humans when it comes to deductive and inductive reasoning, and are also rapidly improving at recognizing patterns and images. As such, the medical field in which CDSS has been most successful is radiology (and also other types of visual data, e.g., electrocardiograms), detecting conditions such as tumours and other lesions in large amounts of imaging data in short amounts of time.[3,12,16] Furthermore, as humans are notoriously bad at statistical reasoning (for example, estimating odds based on quantitative information, see e.g. Kahneman 2011[17]), CDSS can provide a valuable contribution to the process of clinical decision-making by comparing the information clinicians do have about a patient with the information about other (groups of) patients in the database of the CDSS. And, based on similarities with other cases, use this to make suggestions about the diagnosis and predictions of possible outcomes of a certain treatment.

However, as Boon contends, in most professional fields, the goal of performing epistemic tasks such as those listed above, is not (only) to*identify* the most refined classification, or the most perfectly fitting class. Rather, the epistemic purpose is knowing how to control or interact with the targeted phenomenon (e.g., the symptoms or illness of a patient), which requires relevant understanding to begin with. Translated to clinical practice, the goal of performing epistemic tasks is to device interventions that contribute to making the correct diagnoses or actions that alleviate the patient's symptoms or benefit the health of patients. This requires human intelligence, for example to collect, review and process data before it can be entered into the CDSS, to judge which information is relevant, and to evaluate the outcomes. In the next section, we will therefore elaborate on the epistemic tasks of clinicians.

Epistemic tasks by clinicians: constructing a 'picture of a patient'

In an earlier paper, we have argued that good quality decision-making involves highly complex and refined ways of clinical reasoning, of which several examples can be given.[11] First, while considering the available information, clinicians continuously deduce and verify options – this is because they understand, for instance, that one effect can have multiple causes and one cause can have multiple effects. Second, in addition to algorithmic and deductive, rule-based reasoning, "creative" thinking and nuanced styles of reasoning are an important part of good clinical decision-making. For example, clinicians make use of case reports, descriptions of individuals or small groups with 'surprising' or 'problematic' symptoms[18] to come up with a possible diagnosis. Or they use narrative techniques to logically integrate all available information.[19]Third, an understanding of the mechanisms of a disease is necessary to translate general statistical information to the situation of individual patients.[20,21] Finally, Khushf (1999) argues that the diagnostic process involves both *determinative judgement*(bringing a particular instance under a general concept) and*reflective judgement* (beginning with a particular and seeking out a concept). When a patient visits a medical professional, this expert develops an initial insight into what is the matter with that patient (a set of possible diagnoses based on integration of the patient's specific signs and symptoms), thus providing a reflective judgment. A diagnosis is then established by a determinative judgment, i.e. by determining under which diagnosis the observed (but usually incomplete) signs and symptoms fit best.[22] These epistemic tasks (i.e., making these judgments) cannot be outsourced to a machine learning system because it concerns reasoning which is not algorithmic or statistical. It is therefore important that clinicians have developed *expertise*, which includes *tacit knowledge* and *cognitive skills*, enabling them to draw up a diagnosis or treatment plan, despite incomplete information and uncertainty.[14] In addition, clinical decisions are often based on the integration of pieces of evidence generated by medical professionals with different expertise. Interpreting and adjusting the pieces of evidence into a coherent diagnosis takes place in interaction between different experts. This requires specific skills to enable the (social and epistemic) interaction between experts, i.e. opening up and explaining their deliberation to others and justifying to others how they come to a certain interpretation, while being sensitive to deliberations and interpretations from others.[23]

Epistemological responsibility

In the previous sections we have analysed which epistemic tasks concerning clinical decision-making CDSS are well-equipped to perform, and which epistemic tasks require human intelligence. Additionally, we need to explain why clinicians remain responsible for the decisions made in clinical practice, for which we give epistemological reasons. Earlier, we have pointed out that clinicians have the epistemic task to develop a 'picture' of a patient that is logically coherent and consistent with contextual and personal information as well as general, scientific and statistical knowledge.[11] Clinicians together with the patient, and usually in collaboration with other medical experts, use this 'picture' in their clinical *reasoning* about the diagnosis and treatment of the patient. Usually this involves a process in which the clinician, based on the formed picture so far, forms hypotheses about the illness and asks new questions. This leads to additional diagnostic tests and searches in medical literature, which in turn produces new information that is added to the picture, leading to new hypotheses and questions, etc. In other words, the clinician enters into a search process (exploration and investigation) in which new information is *adapted* and *integrated* with the existing information. In this process, clinicians continually update the 'picture' they have of their patients, and use it to direct the next step in the search process.

Collecting, interpreting, adapting and integrating the data into a coherent picture involves a considerable amount of choice, deliberation and justification by clinicians, for example about the relevance and quality of the information. Clinicians are epistemologically responsible for these choices and deliberations, although CDSS can help by providing information in ways suggested above. As a consequence of this epistemological context, clinicians are responsible for the way they construct and use the 'picture' of the patient. This also means that clinicians need to be able to explain and justify their decision-making. We have therefore argued that clinicians should consider themselves *epistemologically responsible* to produce good quality knowledge about their patients.[11] The idea of epistemological responsibility is based on Lorraine Code's (1984) insight that cognitive agents (such as doctors) have an important degree of freedom when it comes to reasoning (e.g., in deciding which information is relevant and which not in their argument; and how to interpret specific information) and that they are accountable for how they deal with this freedom.[24] Therefore, in contrast to passive information processors (such as CDSS or other algorithms) that are at best reliable and fast, clinicians, as cognitive agents, should be evaluated in terms of responsibility. With the notion of epistemological responsibility we aim to grasp the specific epistemic challenges faced by clinicians to perform epistemic activities involved in clinical decision-making concerning diagnosis and treatment. As CDSSs outperform clinicians in some specific, well-defined tasks, their applications may still comply with the epistemological responsibility of clinicians. This requires, however, that the CDSS is fitted into the *clinical reasoning process*, and that the clinician is still able to take responsibility for this process. In Section 3 we will analyse what this means for the development of CDSSs, the required properties of a CDSS, the required skills of the clinicians and the role that a CDSS can play in clinical reasoning.

CDSS as clinical reasoning support systems

Above, we have argued that coming up with a diagnosis and treatment plan involves a search process (exploration and investigation) that is directed by clinical experts. Specific to the reasoning of clinical experts in this search process is, for example, to ask relevant and sensible questions about the case, to decide which parameters (clinical data and other) about a patient are relevant to include and which not, to formulate possible explanations for the symptoms, and to see similarities with other cases. In this epistemological context, CDSS must support this process by answering questions asked by the clinician. For example: 1. What are likely diagnoses for a patient with symptoms x,y,z? 2. What treatments have been found effective for patients with diagnosis A, from age group B, with comorbidities C,D and E? 3. What are the chances that a patient with symptoms x,y,z has disease A? Or disease B? 4. How likely is it that treatment T will be effective for a patient with symptoms x,y,z? 5. If the patient with symptoms x,y,z has disease D, what other signs of symptoms would they have? 6. What if, instead of symptom x, the patient would have symptom w?

In addition, CDSS could also be helpful in effectively searching the patient's medical records, for example to answer questions such as:

7. How often has the patient suffered from similar attacks? 8. What other drugs does the patient take, and might they interact? 9. What other examinations have been performed on this patient, and what was the outcome?

In short, the CDSS can provide information on the patient's records and statistical (numerical) information about illnesses and treatments in similar cases, and with that support all types of reasoning (deductive, inferential, hypothetical, counterfactual, analogical, etc.) employed by clinicians about their patients. Moreover, based on the data of a patient that is fed to the CDSS, the system could come up with suggestions itself (hypotheses). But still it is the clinical expert's epistemic task to: 1) come up with relevant questions and 2) judge the answers. Concerning the latter, the criteria employed by a CDSS to evaluate the answers are different from the criteria employed by the clinician. Whereas the CDSS uses a very limited set of epistemic criteria (such as technical and statistical accuracy, cf. Kelly et al. 2019), a clinician's judgement must meet a more extensive set of both epistemic criteria (such as adequacy, plausibility, coherency, intelligibility) and pragmatic criteria to assess the relevance and usefulness of the knowledge for the specific situation.

In short, we have argued that clinical decision-making is a complex and sophisticated reasoning process, and that a clinician is epistemologically responsible for this process. Instead of thinking of CDSS as a system that answers the question "what is the diagnosis for patient A with symptoms x,y,z" and, subsequently "what is the best treatment for this patient", it is better to think of the system as answering the numerous intermediate questions raised by a clinician in the clinical reasoning process. By answering these questions with the help of statistical information based on a large amount of reliable data, the clinician's reasoning process can be supported, substantiated and refined. Therefore, we propose that it is more suitable to think of CDSS as *clinical reasoning support systems (CRSS).* In the following paragraphs, we will further elaborate on what is needed for good use of a CRSS in clinical practice. We will defend that the designer of the system and the clinicians who will use it, already need to collaborate from early on in the development of the CRSS.

The epistemological role of experts in developing CRSS

Above, we explained that the epistemological role of clinicians in the diagnosis and treatment of individual patients is crucial, even though CRSS can provide important support. Here we will explain that the epistemological role of clinical and AI experts is also crucial in the development of a CRSS, and that these experts need to collaborate.

In a very simple schema, the development of a CRSS consists of three phases, the input, throughput and output. Human intelligence plays a crucial role in each phase.

The *input* in the development of a CRSS is existing medical knowledge (for knowledge-based AI-systems) and available data (for data-driven systems). In the development of *knowledge-based* CRSS all clinical, epidemiological and theoretical knowledge in the medical literature can be used. However, medical experts must indicate which knowledge is relevant for which purpose, and which knowledge belongs together, and also how reliable that knowledge is. In the development of*data-driven* CRSS, reliably labelled data are needed to train the system, while relevant, reliable unlabelled data are need for the system to find patterns and correlations. Knowledge from clinical experts is needed to generate the training set (such as labelled images), and to select sets relevant and reliable unlabelled data. In all these cases, knowledge of clinical experts plays a role in choosing appropriate categorizations, adequate labelling, and in the organization of data storage in order to make the system searchable and expandable for clinical practice.[25]

The *throughput* in the development of a CRSS is the machine-learning process in which the machine-learning algorithm searches for a 'model' (i.e., another algorithm) that connects the labelled data in the training set in a statistically correct way (i.e., supervised learning), or detects statistically relevant correlations in unlabelled data (i.e., unsupervised learning). The design, development and implementation of this machine-learning process requires AI experts rather than clinical experts. However, there will be overlap between the development of the input (the labelled or unlabelled data fed into the process) and the machine learning process, which implies that some collaboration is necessary in this phase.

The output (or result) of the mentioned steps in the development of a CRSS is a 'model' (an algorithm). This model is implemented in the CRSS to be used in clinical practice. But before implementation, the model must be checked by human experts for relevance and correctness, since its statistical correctness does not automatically mean that it is adequate and relevant for the CRSS.[3] 11For example, Kelly et al. (2019) describe a study in which "an algorithm was more likely to classify a skin lesion as malignant if an image had a ruler in it because the presence of a ruler correlated with an increased likelihood of a cancerous lesions" (ibid, 4). This is because the data is under-determined, which means that in principle many statically correct models (algorithms) can be found (cf. McAllister 2011[26]) to (i) connect between labelled data and their labels (in the case of supervised learning), or (ii) find statistically relevant correlations in unlabelled data (in the case of unsupervised learning). In order to be able to do this, clinical experts must, for example, know which parameters play a role in the model and then assess on the basis of their medical expertise whether this is medically/biologically/physically plausible. In short, here as well the contribution of human intelligence is crucial, since medical experts, in collaboration with AI experts must determine whether the resulting model is reliable and relevant.

Explainable and accountable CRSS to facilitate interaction with the clinician

To use a CRSS as a clinical reasoning support system in the manner we suggest above, it is necessary that a CRSS facilitates this. This requires22Another requirement is that a CRSS is equipped with a suitable interface that allows clinicians to enter their questions, possibly even by speaking. And the algorithm should be designed such that it can deal with various questions posed by clinicians. This kind of flexibility might be challenging to implement, it goes beyond the scope of this paper to address these challenges. that a CRSS should facilitate that a clinician can evaluate its answer and judge its accuracy and relevance for the specific patient. A well-known objection to AI for clinical practice is the opacity of the algorithm: how it establishes an outcome based on the input is 'black-boxed'. This, of course, obscures the users' ability to judge the accuracy and relevance of the outcome. Chin-Yee and Upshur (2019), for example, argue that because of the black-box nature of CRSS, using these systems conflicts with clinicians' ethical and epistemic obligation to the patient. According to them, this is one of central philosophical challenges confronting big data and machine learning in medicine.[27]

Similarly, in their 'Barcelona declaration for the proper development and usage of artificial intelligence in Europe' Sloane and Silva (2020) argue that decisions made by machine learning AI are often opaque due to the black box nature of the patterns derived by these techniques. This can lead to unacceptable bias.[9] Therefore, they state that "When an AI system makes a decision, humans affected by these decisions should be able to get an explanation why the decision is made in terms of language they can understand and they should be able to challenge the decision with reasoned arguments" (ibid, 489).

These requirements for the use of AI systems are indicated by the developers of machine learning developers by the concept of*explainable AI.* The idea of explainable AI is that humans can understand how a CSRS has produced an outcome, for example by developing algorithms that are understandable by the users. This, however, might limit the level of complexity of the algorithm, and with that negate the possible benefits of using AI. In case of clinical use it might not be necessary to understand the exact intricacies of the algorithm, but rather to have some insight into factors that are important or decisive to come up with a specific prediction or advice. What machine learning algorithms do is learn to assign weights to features in the data, in order to make optimal predictions based on that data. For clinicians, it is important to know which features are considered relevant by the algorithm and how much weight is assigned to this feature. Having that information, a clinician can judge whether the features that a CRSS picks out are indeed relevant or not (i.e. an artefact in an image, or an unreliable measurement). In the optimal configuration, a clinician can also enter feedback into the system, allowing the algorithm to come up with an alternative prediction, and to learn for future cases.

An advantage of using an *explainable* AI algorithm, assuming that CRSS should be considered as a *clinical reasoning support system*rather than a *decision* system, is that it aids clinicians to explicate their reasoning process. Important in this context is that medical expertise involves a lot of tacit knowledge that can easily remain hidden in the clinical reasoning of these experts. We have argued that epistemological responsibility entails elucidating knowledge and reasoning that otherwise remains implicit.[14] However, for clinicians this can be quite challenging. Using a system that formalises aspects of the reasoning process and explicates the factors that are combined, and with what weight, will support clinicians in developing their ability to articulate and justify their own reasoning process. This explicit understanding, in turn, can contribute to the communication between the clinician and the patient. The explanation enables patients to understand their clinician's reasoning process and add to it, thus empowering them to take part in the decision-making process concerning their own medical care.

Establishing a link between the CRSS and the individual patient

Sullivan (2020) argues that it is not necessarily the complexity or black-box nature that limits how much understanding a machine learning algorithm can provide.[28] If an algorithm is to aid understanding of the target phenomenon by its user (such as a scientist or a clinician) it is more important to establish how key features of the algorithm map onto features of the real-world phenomenon. This is called *empirical justification* . Sullivan calls a lack of this type of justification *link uncertainty* . Link uncertainty can be reduced by collecting evidence that supports the connection between "the causes or dependencies that the model uncover to those causes or dependencies operating in the target phenomenon" (ibid, 6).

Consider, for example, an algorithm that is used to classify cases of skin melanoma[29] (Esteva et al. 2017, as referred to by Sullivan), which is developed by a machine learning algorithm using large amounts of images from healthy moles and melanoma. Because there is extensive background knowledge linking the appearance of moles to instances of melanoma, for example explaining why possible interventions are effective for lesions that look a certain way, "the model can help physicians gain understanding about why certain medical interventions are relevant, and using the model can help explain medical interventions to patients" (ibid, 23). This background knowledge links the mechanisms that are uncovered by the AI algorithm (i.e. predicting which treatments will be effective for which cases) to relevant mechanisms in the target phenomenon (i.e. skin lesion that does or doesn't require treatment). Because of this link, empirical justification is established, and clinicians can use the algorithm to answer why-questions about skin lesions.

Concerning the transparency of algorithms, Sullivan contends that our understanding is quite limited if we know nothing whatsoever about the algorithms. She argues that having some insight in the weighing used by the algorithm is needed. Therefore, as long as the model is not opaque at the highest level, that is to say that there is some understanding of how the system is able to identify patterns within the data, it is possible to use a complex algorithm for understanding. What is needed is "some indication that the model is picking out the real*difference makers* (i.e., factors that matter) for identifying a given disease and not proxies, general rules of thumb, or artefacts within a particular dataset" (ibid, 21).

In our view, Sullivan identifies an important condition for the use of CRSS in clinical practice. Based on her analysis, we infer that it is important to ensure that the algorithm used by a CRSS (which was developed by data-driven AI) is linked to the target phenomenon, by empirical (preferably scientifically supported) evidence. Sullivan has more general links in mind: that the algorithm can generally be used to understand the mechanisms of a target phenomenon. For clinical practice we would add another important link: a link between the algorithm and the individual patient that the clinician intends to diagnose and treat. To establish this link and use a CRSS to better understand the individual patient, clinicians need to ensure/verify that 1) the type of outcome (i.e. the disease category) produced by the CRSS is consistent with the 'picture' of the patient that the clinician has constructed so far; 2) the data used to train the CRSS is relevant to the patient; and 3) that the input required by the CRSS is available to the patient in question and of good quality.

Clinician and CDSS as Hybrid Intelligence

In our approach to clinical decision-making, we contend that clinical decision-making is, in practice, a complex and intricate reasoning process. We argue that CRSS can play a role in or even improve this reasoning process as a *clinical reasoning support system,*provided that the system is reliable, that its outcomes are explainable in relevant respects, and that an empirical link can be established between the algorithm and the individual patient. If these requirements are met, clinicians can combine their human intelligence with the artificial intelligence of a CRSS into a hybrid intelligence,[30,] 11As rules-based systems and data-driven have different capacities, Steels and Lopez de Mantaraz (2018) suggest that "The full potential of AI will only be realized with a combination of these two approaches, meaning a form of hybrid AI." (ibid. 488) This is a different type of hybrid than we have in mind here. in which both have clearly delineated and complementary tasks. To achieve this, CRSSs must be given highly standardized and trainable epistemic tasks. For example, CRSSs can provide accurate and precise classifications based on their ability to detect patterns that are not discernible by humans. Or they can help search the database for the most suitable procedure, supported by the most up to date scientific evidence. Machine learning algorithms make use of large amounts of data, and therefore are able to establish similarities and correlations between (sub)groups of patients and rare cases. The task of clinicians is to incorporate the outcomes of CRSS into medical reasoning, first of all by hypothesizing about possible causes of the patient's signs and symptoms, and secondly by selecting the appropriate test to confirm or reject this hypothesis. In addition, clinicians are tasked with determining what data is relevant, collecting that data and entering it into the CRSS, such that the system can use it. In short, the task of a clinician is to ask questions that the CRSS can answer. Moreover, clinicians are tasked with interpreting, integrating and contextualizing the outcome of the CRSS, in order to utilize it for empirical tasks in practice.

Additionally, we have defended that clinical experts need to be closely involved in the development of AI systems. Developing a CRSS that facilitates clinical reasoning in practice means that clinicians, as future users, need to be involved at an early stage of development. They need to ensure that the system is designed to answer questions that are relevant to the clinical reasoning process, and that the data that is collected and used as training data is relevant and suitable to their patient population (e.g., that a CRSS to diagnose skin cancer is not just trained on using data of patients with white skin[31] and that the outcomes generated by the CRSS are interpretable. This requires that, along with an advice, the system indicates what factors contributed to arriving at that advice, allowing the user to evaluate whether these factors are indeed medically plausible and applicable to the current patient. In addition, medical education should prepare clinicians to perform the new epistemic tasks required to use CRSS in clinical practice. For example by teaching students how data is collected and processed and by teaching them how to evaluate whether the context in which the data is collected is relevant to the intended application.

In conclusion, a CRSS can aid clinical decision-making and possibly improve it, if clinicians use it in an epistemologically responsible manner. Both the system and their users need to be equipped for this. Clinicians need to develop new cognitive skills necessary to perform specific epistemic tasks related to the use of CRSS. For example, establishing an empirical link between the model and the individual patient, asking appropriate questions (that can be answered by the system), collecting and assessing the required data and evaluating the outcome. CRSS must not only be reliable, in the sense that the performance is scientifically proven to be as good or better than that of medical experts. It must also provide the information necessary to enable the clinician to perform the necessary epistemic activities, in a way that supports the performance of these activities. This entails, for example, to provide insight in the data set that is used to train the algorithm (e.g., which characteristics of patients were included in the data), as well as a precise description of the task that the algorithm is trained to perform (e.g., to use images of skin lesions for identifying melanoma); to give information about the reliability of the outcome (such as confidence intervals) and; to give information about the procedure with which the algorithm arrives at the outcome (i.e. the weightings of the different pieces of information).

If developers and users succeed in meeting the requirements that allow them to combine human and artificial intelligence into hybrid intelligence, CRSS holds great promises for health care by improving the accuracy, speed and consistency of clinical decision-making.

References

1. Niezen MGH, Edelenbosch R. Van Bodegom L. Verhoef P. *Health at the centre - Responsible data sharing in the digital society.* The Hague: Rathenau Instituut. 2019
2. Kool L, Timmer L, Royakkers L, Van Est R. *Urgent Upgrade - Protect public values in our digitized society.* The Hague: Rathenau Instituut. 2017
3. Kelly C J, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC med.* 2019;17(1). doi: 10.1186/s12916-019-1426-2
4. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. *J. biomed. inform* . 2018;78:134-143. doi: 10.1016/j.jbi.2017.12.005.
5. Wyatt J, Spiegelhalter D. *Field trials of medical decision-aids: potential problems and solutions.* In: Clayton P, ed. Proc. 15th Annu. Symp. on Comput. Appl. Med. Care. 1991. Washington.
6. Sikma T, Edelenbosch R, Verhoef P. *The use of AI in healthcare: A focus on clinical decision support system.* 2020 [RECIPES project: https://recipes-project.eu/]
7. Mahadevaiah G, Prasad RV, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and quality assurance. *Med. Phys.* 2020;47(5): e228-e235. Doi: https://doi.org/10.1002/mp.13562
8. Montani S, Striani M. Artificial Intelligence in Clinical Decision Support: a Focused Literature Survey. *Yearb. of Med. Inform.*2019;28(1):120-127. Doi: 10.1055/s-0039-1677911.
9. Sloane EB , Silva RJ. Artificial intelligence in medical devices and clinical decision support systems. In: Iadanza E, ed. *Clinical Engineering Handbook* . Academic Press. 2020.: 556-568 Doi: https://doi.org/10.1016/B978-0-12-813467-2.00084-5
10. Steels L, Lopez de Mantaras R. The Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe.*AI Comm.* 2018;31(6): 485 – 494. DOI**:** 10.3233/AIC-180607
11. Van Baalen S, Boon M. An epistemological shift: from evidence-based medicine to epistemological responsibility. *J Eval Clin Pract* , 2015;21(3):433-439. DOI: 10.1111/jep.12282.
12. Savage N. Another set of eyes for cancer diagnostics. *Nature*2020;579:S14-S16. doi 41586-020-00847-2
13. Dagliati A, Tibollo V, Sacchi L. et al. Big Data as a Driver for Clinical Decision Support Systems: A Learning Health Systems Perspective. *Frontiers in Digital Humanities* 2020;5 https://doi.org/10.3389/fdigh.2018.00008
14. Van Baalen S, Boon M. Evidence-based medicine versus expertise – knowledge, skills and epistemic actions. In: Bluhm R, ed.*Knowing and Acting in Medicine.* Rowman & Littlefield; 2017:21-38. ISBN: 978-178348810.

15. Boon M. (2020) How scientists are brought back into science - The error of empiricism. In: Bertolaso M, Sterpetti F, eds. *A critical Reflection on Automated Science - Will Science Remain Human.*Springer Series Human Perspectives in Health Sciences and Technologie. Dordrecht: Springer. 2020:43-66. DOI 978-3-030-25001-0_4
16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 2019;25:44-56. doi: https://doi.org/10.1038/s41591-018-0300-7
17. Kahneman D. *Thinking fast and slow.* New York: Farrar, Straus and Giroux. 2011
18. Ankeny R A. Using cases to establish novel diagnoses: Creating generic facts by making particular facts travel together. In: Howletts P, Morgan MS, eds. *How Well Do Facts Travel? The Dissemination of Reliable Knowledge.* New York: Cambridge University Press. 2011:252-272.
19. Solomon M. Epistemological reflections on the art of medicine and narrative medicine. *Perspectives in Biology and Medicine,*2008;51(3):406-417.
20. Russo R, Williamson J. Interpreting Causality in the Health Sciences.*Int. Stud. Philos. Sci.* 2007;21. pp. 157-170, https://doi.org/10.1080/02698590701498084
21. Parkkinen V-P, Wallmann C, Wilde M, et al. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedure.* 2018; SpringerOpen*https://doi.org/10.1007/978-3-319-94610-8*
22. Khushf G. ‘The Aesthetics of Clinical Judgment: Exploring the Link between Diagnostic Elegance and Effective Resource Utilization’,*Med Health Care Philos.* 1999; 2(2):141-59 DOI:10.1023/a:1009941101276.
23. Van Baalen S, Carusi A, Sabroe I, Kiely DG. A social-technological epistemology of clinical decision-making as mediated by imaging.*J Eval Clin Pract* , 2016;23(5):949-958. DOI: 10.1111/jep.12637.
24. Code L. (1984), ‘Toward a ‘Responsibilist’ Epistemology’,*Philos. Phenomenol. Res.* 1984;45(1):29-50. DOI: 10.2307/2107325.
25. Leonelli S, Tempini N, eds. *Data Journeys in the Sciences* . Berlin: Springer. 2020
26. McAllister, J.W. (2011). What do Patterns in Empirical Data Tell Us About the Structure of the World? *Synthese* 182 (1): 73–87. https://doi.org/10.1007/s11229-009-9613-x.
27. Chin-Yee B, Upshur R. Three problem with big data and artificial intelligence in medicine. *Perspect. Biol. and Med.* 2019;62(2): 237-256 DOI:*https://doi.org/10.1353/pbm.2019.0012*
28. Sullivan E. Understanding from Machine Learning Models. *Brit. J. Philos. Sci.* 2020;axz035,*https://doi.org/10.1093/bjps/axz035*
29. Esteva A, Kuprel B, Nova R, Ko J, Swetter S, Blau H, and Thrun S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature.* 2017;542:115–8.
30. Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid Intelligence.*Bus. Inform. Syst. Eng+* 2019;61:637-643. Doi:https://doi.org/10.1007/s12599-019-00595-2
31. Adamson AS, Smith A. Machine Learning and Health Care Disparities in Dermatology. *JAMA Dermatol.* 2018;154(11):1247-1248 DOI: 10.1001/jamadermatol.2018.2348

Acknowledgements

Conflict of interest statement

All authors declare not conflicts of interest.