The AAAI 2023 Workshop on Representation Learning for Responsible Human-Centric AI (R^2HCAI)

AI in Medical Diagnosis
AI Prediction vs Human Judgment

Dóra Göndöcs[1] – Viktor Dörfler[2]

dora.gondocs@gmail.com, viktor.dorfler@strath.ac.uk

Abstract

AI has long been regarded as a panacea for decision-making and many other aspects of knowledge work – something that will help humans get rid of their shortcomings. We believe that AI can become a useful tool to support decision-makers, but not that it can replace decision-makers. It is not just a matter of time and of the level of AI development. Decision-making makes use of algorithmic analysis, but it is not solely algorithmic analysis. It also involves other factors, many of which are very human, such as creativity, intuition, emotions, feelings, and value judgments. We have conducted semi-structured open-ended research interviews with 17 dermatologists to understand what they expect from an AI application to deliver to a medical diagnosis. We have found three aggregate dimensions along which the thinking of dermatologists can be described: responsibility, explainability of AI, and the new way of thinking (mindset) needed for working with AI. Furthermore, we have identified three distinct ways in which our participants chose to interact with AI. We believe that our findings will help physicians, who consider using AI in their diagnosis, to understand how to make the best use of AI. It will also be useful for AI vendors in improving their understanding of how medics want to use AI in diagnosis. Further research will be needed to explore if our findings have relevance in the wider medical field and beyond.

Introduction

The rapid advances in AI developments over the past few decades have resulted in increasingly available AI applications to support human experts in their work, including decision-making. In this paper we examine how dermatologists use or envisage using AI in their diagnostic work. We chose medicine, as it is one of the most developed AI application areas, there is already substantial experience in using AI, and the high quality of this use is critical – i.e. lives are at stake. We decided to choose one single area in medicine, in order to achieve high consistency. Dermatological diagnosis is particularly suitable area of study, as it makes use of image processing aspect of AI, which is particularly well developed. Specifically, we focus on the process of diagnosing melanoma; this provides useful basis of comparability for the participants' accounts. In addition, the lead author has access to the participants, which provides the benefits of the "insider view". Some of our research participants already had hands-on experience in using AI in their diagnostic work, others only thought about it, based on publications and conversations with colleagues, which adds the richness of diverse participant perspectives to our data.

As we see it, the tendency is not about replacing the human decision-makers with AI, it is about producing accurate algorithmic predictions, which are then supplemented with the (value) judgments by the human experts. The algorithmic predictive capability of AI is an input into the decision-making process and the human expert's final decision (judgment) remains critical. Thus our starting point is what can be legitimately called a "decision support" (Sharda, Delen and Turban 2020) and what is referred to more recently as "decision augmentation" (May, Utts and Spottiswoode 1995, 2019; Leyer and Schneider 2021).

The AI-generated predictions and the human judgments are both inputs and they can be strongly intertwined; but in the end, there will be a final (judgment) point and this is in the hands of the human. This means that we are not covering AI-enabled automation processes, we acknowledge that these work in some areas but we are interested in areas, such as medical diagnosis, in which human mastery plays a crucial role. Within this scope, we focus on the human side of the human-AI interaction. We investigate the human factors, expectations, and impressions, and highlight how an AI solution, if designed right, would affect their work – more precisely, how this is seen by our research participants.

We have designed an exploratory, qualitative empirical study, aimed at understanding how dermatologists think and feel about AI and about using AI, as well as how the use of AI altered or would alter their established diagnosis processes. During the completed semi-structured interviews, we learned about the way the process of mole checkups changes with the use AI and how the algorithm might influence the doctors who diagnose melanoma. Further, we have asked these doctors what information they believe is necessary for introducing AI in their diagnosis work. In other words, what it takes to get these diagnosticians AI-ready.

We found that the process of melanoma diagnosis consists of two components, a prediction and a judgment. In a human-only scenario, i.e. when no AI is used, the predictions are created and the judgment are made by the doctor, therefore the two components of the process build on each other and the process is linear. When AI is used, the two components are disentangled, contributed by separate entities, and thus they are arguably less intertwined. This, however, does not suggest a lower complexity, what constitutes the complexity changes as well – namely, the human-AI interaction becomes one of the sources of complexity.

In order to depict how dermatologists think and feel about using AI in their diagnosis, in what follows, we first provide a brief overview of the background knowledge on using AI in medicine. Then we outline our methodological considerations, explain our choices, and describe the scope of the study. Next, we present our findings, organized around three themes: responsibility, explainability, and the mindset needed to work with AI. Subsequently, we discuss the findings in the light of the extant literature, highlighting what is significant about our improved understanding, and also explore the implications of the findings. We finish with a final commentary rather than a conclusion to emphasize that this is the beginning of our inquiry rather the end of it.

Background Knowledge

In this section, we introduce the background literature that is directly relevant for this study. We do not cover the general AI literature, only the specific development and applications. Having said that, it is important to state what position we take on AI; for the purpose of this paper:

> "AI is loosely defined as machines that can accomplish tasks that humans would accomplish through thinking." (Dörfler 2020)

This definition does not say anything about AI accomplishing such tasks would do it in a way that resembles human thinking; we do not see anything in this definition that implies that AI would think in the human sense of the word. Importantly, AI as a field is not simply a study of the machines, it is as much the study of the human mind (Dörfler 2022a; for a more detailed description see e.g. Dörfler 2022b). Specifically in the area of decision-making, including medical diagnosis, we believe that Davenport's (2018: 44) description of AI as *"analytics on steroids"* is particularly expressive and that therefore AI cannot be said to make decisions but it can make our (human) decisions better informed. This is in line with what we have heard from our research participants.

AI in the Medical Field

There is a growing number of publications on AI in medical research over the last decade (Jiang et al. 2017; Yu, Beam and Kohane 2018; Ruiz, Wyszynska and Laudanski 2019; Guo et al. 2020; Rong et al. 2020). One of the most promising AI developments in medicine is in the field of machine learning (ML) in artificial neural networks (ANN), with a focus on predicting clinical events, such as improving the accuracy of diagnosis, defining new preventions or treatments, clinical decision support, postprocessing, and quality control (Choi et al.; Patil, Szolovits and Schwartz 1981, August; Rizzi 1994; Miller 2010; Bussone, Stumpf and O'Sullivan 2015; Semigran et al. 2015; Choy et al. 2018; Davenport and Glover 2018; Cai et al. 2019; Esteva et al. 2019; Fagherazzi and Ravaud 2019; Treasure-Jones et al. 2019; Richens, Lee and Johri 2020; Maassen et al. 2021).

Among medical AI solutions, image processing was the main AI tool to advance disease detection in radiology primarily by using deep learning (DL), which can be understood as ML in so-called deep neural networks (DNN), meaning that there is more than one hidden layer in the ANN (Choi et al.; LeCun, Bengio and Hinton 2015; Wang, Lin and Wong 2020; Cabitza, Campagner and Sconfienza 2021). Expectations towards AI advances are extremely high, with the goal to improve medical healthcare as seen by physicians (Yu, Beam and Kohane 2018; Rong et al. 2020). Thus, AI is viewed as changing the long-held status quo in healthcare, including the physicians' role, towards precision and personalized medicine (Fröhlich et al. 2018). However, it is assumed that AI will not fully replace but augment the work of physicians establishing a new kind of human-AI interaction, in line with the idea of Augmented Intelligence (Lai, Kankanhalli and Ong; Park et al.; Claburn 2016; Tschandl et al. 2020).

Much of AI applications in medicine heavily relies on big data analysis, image and speech processing available due to recording an astonishing amount of medical data in a structured way in medical databases (Kumar, Vimala and Britto 2019; Wang, Lin and Wong 2020). Such medical big data analysis uses various ML techniques, including DL, shallow or convolutional neural networks (CNN), vector machines, or random forests (Chen et al. 2019; Esteva et al. 2019; of the Madrid et al. 2019). Among these techniques, DL shows great potentiality where large datasets are available, especially in the field of images, language, and speech processing (Esteva et al. 2019). Where such large datasets are not as much available for studying medical conditions, other ML techniques may be superior (Chen et al. 2019). Besides ML, the most commonly applied examples of AI in healthcare either support the process of the diagnosis by predicting the course of a disease (Montazeri et al. 2016; Hosny et al. 2018; Meiring et al. 2018; Kather et al. 2019), clinical decisions (Xu et al. 2020), or workflows in hospital management (Rush, Celi and Stone 2019; Baltruschat et al. 2021; Davenport and Bean 2022). There are also (predominantly hybrid) AI applications of marginal fields that produce important results in basic research, such as the recent successes of AlphaFold (Heaven 2020).

To conclude, the most critical precondition of the emerging AI developments in healthcare is the data availability needed to develop and train algorithms. Therefore, it is of paramount value to make anonymized and consolidated data available for research purposes by establishing freely accessible research databases such as MIMIC-III, the Medical Information Mart for Intensive Care (Johnson et al. 2016). Among these applied solutions, vendors in radiology and other areas of imaging have started integrating AI into their solutions as a final (at least current) stage of the technological evolution of radiology (Willemink and Noël 2019), or in other cases, the recorded healthcare data have been increasingly used to validate the amount of data for medical predictions (Brajer et al. 2020; Manz et al. 2020). Finally, we acknowledge the importance of the sociotechnical components necessary for successful AI implementations in clinical environments that Cabitza et al. (2020) calls "last mile gap" of AI bridging implementation and operation. Finally, besides the social and technical conditions, the regulatory and the human factors might also hinder AI implementations in medical health care (He et al. 2019). We do not suggest rushing towards more numerous medical implementations without necessary caution, we argue for careful advances primarily trusting experienced physicians to determine the pace of such advances.

Using AI

The first part of the literature review on AI in healthcare confirms that that the ongoing AI developments might bring one of the most significant potential benefits in the diagnostic process, even though the use of such AI tools is still relatively rare in real-life medical practice. Indeed, no FDA-approved AI-based medical device has been introduced into dermatological practice yet.

In this paper, we want to understand the human factors in the medical diagnosis involving human-AI interactions. More specifically, we explore how dermatologists think when AI is involved in the diagnostic process and how they make decisions (judgments) about melanoma by considering the AI-generated predictions (Montazeri et al. 2016).

Combining human and artificial intelligence, for instance, a dermatologist using CNNs specifically to distinguish melanoma (cancerous tissue) from non-malign skin tissue, achieved higher performance than either a dermatologist or AI on its own (Hekler et al. 2019). Moreover, an increasing number of studies confirmed that integrating human expertise with feedback from an AI system, could lead to a synergy that outperforms both the human and the AI (Bulten et al. 2021). However, as the process of diagnosis is significantly altered, using AI requires developing new knowledge, especially for medical trainees, during image interpretation perception, analysis, and synthesis (van der Gijp et al. 2014). The use of AI in melanoma diagnosis is not a unitary construct, Tschandl et al. (2020) suggest using different AI-based applications at different levels of mastery (Göndöcs and Dörfler 2022). We note that the scope of this problem is possibly far more general than the medical field, as those who studied the levels of mastery emphasize the qualitative changes in the nature of knowledge with the increase of mastery (Chase and Simon 1973b, 1973a; Larkin et al. 1980; Dreyfus and Dreyfus 1984, 1986, 1987; Ericsson and Smith 1991; Gobet and Simon 1996a, 1996b, 2000; Dörfler, Baracskai and Velencei 2009). Tschandl et al. (2020) also observe the significance of testing the performance of AI-based solutions under real-world conditions and by the intended users, rather than testing isolated AI application by programmers.

Applying Kasparov's law in the field of radiology, Cabitza et al. (2021) call for using good interaction protocols, as those can contribute to improved decision-making which may exceed the individual agents' performance. The same study (Cabitza, Campagner and Sconfienza 2021) shows that, in line with the second part of Kasparov's law, teams of weaker radiological readers supporting their judgment (decision) by "fit-for-use" protocols could outperform teams of stronger readers, supported by similar but not "fit-for-use" protocols.

Thinking along similar lines, Davenport and Glover (2018) emphasize the importance of choosing the right augmentation approach when medical knowledge workers interact with AI. Their framework consists of five approaches that can be used in healthcare decision-making as well by medical experts during their interaction with AI: step up, step aside, step in, step narrow, and step forward.

Narrowing our focusing to human-AI interaction in medicine, we identified three leading groups of studies, which do not form a taxonomy, but which signify what are the hot topics in the problem area. The first group of studies explores what information users need in order to rely on AI-generated predictions in the diagnostic process (Bussone, Stumpf and O'Sullivan; Poursabzi-Sangdeh et al.; Ribeiro, Singh and Guestrin; Luo et al. 2019). Studies in the second group focus on the principles of designing AI applications that can be seamlessly implemented into medical practice (Yang, Steinfeld and Zimmerman). Finally, we found one study that analyzes the onboarding process of medics who use an AI application for the first time and try to figure out what AI can do and how to work with it (Cai et al. 2019). This study underlines the need for defining appropriate mental models and the need for determining strategies when using AI in decision-making.

Overall, there is a generic agreement in the literature that AI cannot and should not replace human decision-making but only augment it. Nevertheless, the literature also emphasizes that using AI alters the decision-making process and therefore humans must learn how to think differently *in* and *about* decision-making to benefit from using AI.

Methods

The background literature reflects the rapidly increasing trend of medical AI studies, which discuss the impacts of using AI in medical diagnosis. We have designed an exploratory, interpretivist, phenomenon-driven, qualitative empirical study, aimed at understanding the human side of the human-AI interaction. We conducted semi-structured open-ended interviews with 17 dermatologists, inquiring about their expectations and experiences (if they had any) involving AI in the diagnostic process of melanoma. Our methodological choices and the research design are outlined in this section.

Philosophical and Theoretical Positioning

We loosely position our study within the interpretivist philosophical approach, specifically within the phenomenological tradition (Husserl 1913a, 1913b; Heidegger 1927; Husserl 1936; Schütz 1967, 1970; Heidegger 1975; Finlay 2009), as we are interested in our participants' lived experiences.

This is an early-stage exploratory study, the purpose of which is to achieve an initial understanding of the phenomenon of using AI in medical diagnosis, therefore, we did not aim for a large number of interviews, rather for spending more time on each interview trying to unpack what is in there. This means that we work with 'thin data', based on which we engage in theorizing (Furnari 2014; Bas, Sinclair and Dörfler 2022). Furthermore, for the same reason, we wanted to keep our options open, thus we do not commit to a particular 'theoretical lens' – as a lens always limits what the researcher can see. Instead, we engage in phenomenon-driven theorizing, letting the phenomenon take us wherever it goes (Ployhart and Bartunek 2019; Fisher, Mayer and Morris 2021; Langley 2021).

Furthermore, the research design of this study qualifies as insider ethnography, as the first-named author works at the same clinic as our research participants. This 'insiderness' brings the benefits of insight, but is also often criticized for researcher bias – we deal with this in the way of phenomenology, using bracketing (see later in this section).

Selecting Participants

17 dermatologists have been interviewed from various private and public healthcare institution, and 11 of them work at the same private clinic that specializes in dermatology as one of its core services. The first-named author, the interviewer, works at the clinic as an operational director, so she was able get access to the research participants and conduct the interviews as an insider. She asked directly these physicians about their thoughts, impressions, and feelings towards AI. The interviewer made the access easy and the participants were likely honest in their responses.

The interviewees were at different stages of their careers (see Appendix 1), and presumably therefore in their levels of mastery (Dörfler, Baracskai and Velencei 2009). Although the number of years in the profession does not automatically translate into mastery, it is often used as a proxy, and with highly specialized knowledge workers this proxy should be at least somewhat informative. Of the 17 participants 6 did not have any experience with AI, 11 had experience in research and laboratory, and none of them had clinical experience with AI. So the group is homogenous in terms of work area, they are all deeply engaged in the studied diagnostic process, but they represent variations in terms mastery and AI experience. This was the purposive sample we were aiming for.

Collecting Data

In total 17 dermetologists were interviewed in two rounds, nine in the first and eight in the second. We set up an outline interview protocol for the first round of interviews, focusing on how the participants use or could use AI-generated predictions when diagnosing melanoma and how that would influence their decision-making process (judgment) about melanoma. The second round of interviews commenced five months later, following the analysis of the interviews from the first round, therefore the interview protocol, albeit loosely, centered around the initial themes. In this second round we probed what we learned from the first round, aiming for high consistency, and digging deeper trying to unpack further richness.

The interviews were semi-structured, we formulated a small number of research themes to provide structure to the interviews. The idea was that these themes can help the participants focus on the changes in the process of diagnosing melanoma before, during, and after introducing AI:

- How could you work using AI in your diagnostic work? Up to what level would you trust and use the predictions as proposals provided by AI? How do you regard AI? How do you relate to it?
- What information you would need if you were considering whether to use AI in your diagnostic work?

What information would help you make the best use of AI?

- How do you think AI would affect other dermatologists' work in working out the final diagnosis? Would this be different by levels of mastery?
- How would you, as a dermatologist, design medical AI for diagnosis support? What are the critical parameters?
- Have you ever thought of an AI solution that can learn the level of mastery and adapt to it? So, it would provide different kinds of support at different levels of mastery.

In both rounds, we were collecting new data until the saturation point was reached, i.e. until we did not learn anything new from additional interviews. In this study, this meant 17 interviews in total. Of course, one can never be sure that the next interviewee or the one after that or one 10 interviews later would not say something new, but we felt that we have understood the phenomenon that we were interested in at this point (Pratt 2008, 2009; Saunders and Townsend 2016). The interviews were all conducted in the local language, which is the native language of the interviewer as well as the interviewees. The analysis were also conducted in this language, and only quotes that were included in the paper were translated.

Analyzing Data

To analyze the interviews, we used a variant of *thematic analysis* (Saunders, Lewis and Thornhill 2019), which allowed us to formulate some pre-established 'a priory' codes based on the literature but also allowed for emergent ' in vivo' codes. The coding was hierarchical, and we used Gioia's approach to visualizing our code/data structure (Gioia, Corley and Hamilton 2012). Analyzing the interviews with the 17 dermatologists led to 29 first-order concepts, which reflected the interviewees' points on a wide range of issues. Synthesizing these first-order concepts, we obtained 10 second-order themes that represent our (authors') understanding. Finally, the second-order themes have been aggregated into three aggregate dimensions at the highest level; we present the findings organized around these three aggregate dimensions.

Bracketing

Bracketing is the tool that phenomenology offers for dealing with the researchers' judgments and pre-understandings. This is an essential aspect of phenomenological research, which focuses on the lived experience of the research participants, particularly when the researcher is an insider (Olekanma, Dörfler and Shafti 2022). Importantly, in our interpretivist approach, the purpose of bracketing was to make use of pre-understandings and insider knowledge as a source of insights rather than affecting the research in unknown ways (Dörfler and Eden 2014; Stierand 2014; Stierand and Dörfler 2014; Dörfler and Stierand 2021).

During the data collection, the interviewer practiced bracketing though personal reflexivity, meaning, that she focused on what the participants had to say, refraining from making her own interpretation or judgement. During the analysis, we practiced bracketing through transpersonal reflexivity, meaning that the interviewer did all the coding, and the other researcher queried the interviewer's interpretation, as if metaphorically holding a mirror to the interviewer (Dörfler and Stierand 2021). Typical questions in this stage would be along the lines of "so how do you know what your interviewee meant by XYZ?"

Analysis and Findings

In the final step of the analysis, we synthesized the second-order themes into three aggregate dimensions (see Appendix 2). There was one other dimension that could have been aggregated, however, as it seems to feed into each of the three dimensions, we label this one the "0^{th}" dimension. Each of these dimensions is outlined below. Using Pratt's (2008) suggestion, we use a few "power quotes" in the main text and further "proof quotes" are available in Appendix 2.

We have also learned a great deal about the participants' attitude towards AI. Two participants did not expect much benefit from using AI; they were not harshly against it, they expressed resigned ambivalence. Three participants showed some interest but also voiced serious concerns, such as:

> "I'm ambivalent, there are possible benefits but serious risks too…" (Participant M)

They pointed out the dangers of dermatologists not being prepared for using AI:

> "I think if we don't learn to use AI properly, it may cause misdiagnosis…." (Participants L)

Similarly, it could be dangerous for junior dermatologists in the process of learning to diagnose:

> "AI can be beneficial but at the same time also risky for young professionals if they trust AI-generated predictions more on the prediction of AI and less inthan their own judgment." (Participants A)

12 participants were very keen on using AI in their diagnostic work. They emphasized data processing power and speed of AI concerns, like Participants B, saying:

> "I can see its clear benefit that compared to a human, AI can handle big volume of data, and if it could scan and analyze the whole body of a patient and point out that might have a risk for melanoma, that could be a great support and save time for us, physicians."

Time saving for physicians was a leightmotif, one suggestions was that AI could provide a kind of pre-screening:

> "AI could point out those that differ from the rules and may bring any risk of melanoma. With that, it could save time for the dermatologist and money for the patient." (Participant D)

AI's capacity to identify patterns over time, was also flagged as a postential souce of performance improvement:

> "It would be a great benefit if AI could track the changes of a mole via the images recorded by time passing and warn in case of negative changes. In that case, an AI can significantly augment the dermatologist's work and improve performance." (Participant C)

The number of respondents is far too low to warrant any statistical analysis, so we do not suggest that the ratios are representative of the population of dermatologists, it only signifies that our participants had diverse attitude towards AI. In the following subsection, we show in what ways the research participants envisaged using AI, and in each of the next three we elaborate an aggregate dimension, showing how our participants thought of *responsibility*, *explainability*, and the about the need for a *different mindset* to benefit from using AI. Interestingly, each of these demansions is widely discussed in the AI literature, but our participants made some unusual comments.

What is the Role of AI?

We have noticed that the interviewed dermatologists do not think about AI in terms like the AI vendors do, i.e. whether it is embedded AI or only an image recognition software. Initially they used a larger number of terms, but through a deeper discussion three distinct roles crystallized, in which our participants would think of using AI in medical diagnosis, we describe these using three metaphors: (1) a tool, (2) an assistant, or (3) a 'colleague'.

Although the metaphors are anthropomorphic, we think of them more like use patterns. When AI is regarded a tool, all that matters is the sheer processing power, the role of AI would be to perform well-structured tasks:

> "I would think of it as a tool that works with image recognition that has seen thousands of images. Thus, it can provide a differential diagnosis for me and specific probabilities of melanoma." (Participant O)

AI as assistant should be taught of along the lines of smartphones and such, that "learn" the habits of the user and prepare things for them, often without prompt:

> "Sometimes it may help to set up a differential diagnosis that may or may not be accepted by the doctor." (Participant H)

AI as a colleague is primarily about having a discussion with someone in order to form an opinion, in this case, a diagnosis. If the physician comes up with a diagnosis and 'runs it by the AI colleague', AI could be very useful in determining if the opinion has some major flow, if the physician overlooked something, or the opinion can be easily refuted. This is particularly important for those who are not completely confident in their diagnosis:

> "The younger, less experienced dermatologists might think of AI as a peer colleague, while the most experienced ones said they could instead look at it as a resident supporting them." (Participant M)

It is important to note that melanoma diagnosis is matter of life and death, and therefore the action is heavily skewed towards the positive (i.e. cancer) judgement:

> "Indeed, if AI said it was a melanoma and I thought of it as a naevus or a basalioma, I would go for safety, and I would still cut it off." (Participant E)

Importantly, none of the interviewed dermatologists thought that AI, at least currently, thinks, and they did not engage in a fantasy world, they were very much focused on improving their diagnosis. This links closely to the first aggregate dimension, the notion of responsibility.

Who is Responsible?

Most interviewed physicians expressed a positive attitude toward an AI in medicine, but every single one of them confirmed that, at the end of the day, it is the physician who must take responsibility and make the final decision, based on a value judgment, about a diagnosis. Only one participant speculated that perhaps some day AI will be able to take responsibility, but the rest firmly rejected even a remote future possibility:

> "The AI system can assist but can never become the one who makes the final diagnosis." (Participant E)

This is not surprising, what we were really interested in was the reasoning behind it. We have found that they were not worried about their jobs, they were conscious of the life-and-death nature of the disgnosis:

> "We need to go for safe, and the final decision about a diagnosis will remain the responsibility of the physician." (Participant M)

If they were worried about something, it was their patients and their professional integrity:

> "I could hardly imagine that a patient would accept if I told him that the AI systems said this and that…" (Participant N)

They realize that medicine is not only about establishing the diagnosis but also about communicating it:

> "My patients want to talk and discuss every little detail…" (Participant J)

We also noticed, that the interviewed dermatologists made assumptions about their patients – they did not actually ask them if they would be happy with the explanation that an AI conjured the diagnosis. This raises the question if our interviewees really thought that their patients would be so reluctant to accept AI as a source of diagnosis, or it was them who needed to understand – this will be further unpacked in the next subsection.

Understanding how the users think about responsibility is not only important for us as researchers, but also to vendors: medical doctors are not looking to get rid of their responsibilities.

Can you Explain?

Unsurprisingly, most of our interviewees suggested that the future of diagnosis will be a mix of human mastery and artificial intelligence. In an attempt to understand how they envisage this mix, we tried to understand when the dermatologists would trust the AI predictions. It was hardly surprising to find that, just like between humans, coming to trust AI takes time:

> "Probably the longer I use such an AI tool and previously gave me good predictions, the more I could rely on that in the next cases." (Participant D)

The other aspect of trusting AI is also something be expected: explainability. However, our interviewees did not think about explainability in a trivial way. Before a widespread routine implementation of AI, these medical experts want to see scientific proof of its validity, and they all wanted to get a broad range of detailed information about the design, operation, learning, and adaptive capabilities of AI in their domain:

> "I doubt I could trust entirely and would use 100% of what the AI proposes, but if I knew how the AI tool has been designed and who did participate in the design, that could increase my trust." (Participant B)

Those participants who understood a bit more about how AI (specifically ML) worked, expressed more specific information requirements regarding AI design:

> "One key factor is knowing that the outcome of each diagnosis was looped back into the system, which could further train the AI system reliability." (Participant H)

We note that offering AI to medical experts (and presumably any expert in any field) brings explainability to a new level. They do not only want to understand how a specific prediction has been achieved, many of them realize that this may not be possible, as there is too much data processing. Instead, they want to understand how AI was set up, how it works. They have a good understanding of science, and they want to understand AI on scientific terms.

Thinking Differently with AI

Our final aggregate dimension reveals that using AI in medical diagnosis will require a new mental model, a new way of thinking, about the process of diagnosis. This new mental model needs to incorporate both AI predictions and human judgments, where both the dermatologists and the AI must learn and adapt to the each other (although, clearly, learning means different things for the physicians and for AI). Without involving any AI, the predictions and the judgments are all handled as one in the physician's mind; the physician does not distinguish between the preliminary-diagnosis and the final diagnosis.

We asked our interviewees to explain the current process and how they diagnose melanoma without AI. They all emphasized that a diagnostic procedure is complex, it is not just a search for specific patterns and application of rules but involves an understanding of the whole picture of a patient and translating that into a diagnosis. One of our participants, for instance, noted that even a patient's anxiety level might influence the final judgment of a dermatologist. They also admitted that there are personal preferences, different dermatologists diagnose differently:

> "I prefer to check all moles with the macroscopy, looking for that specific structure- and color-based characteristic of melanoma. I prefer to do this because it can cause surprises in both directions, and I might set up a different diagnosis if I check first without and then with dermatoscopy." (Participant D)

Many dermatologists, particularly those at the highest levels of mastery, start the examination with their eyes, they pick the suspicious moles, and they study these in more depth with a dermatoscop.

> "Some moles might cause surprises, and checking with my eyes or a dermatoscop might lead to a different diagnosis." (Participant C)

This is just one example where it can be seen that medics use their tacit knowledge, rooted in years of experience. They are also very much aware of using tacit knowledge, and of the value it may provide.

When introducing AI into the diagnostic process, not only the decision-making process of the diagnosis changes entirely, but together with that, it may partially or fully change the approach of the dermatologist. In other words, an augmented diagnosis process, featuring AI, will require new thinking, working methods, and procedures.

Discussion

There are two types of elaborations that we provide here: the first is concerned with how our findings fit with the extant literature, and the other is about exploring the implications. These two aspects of the discussion are intertwined in this section; the structure is the same as for the findings.

Our participants apparently do not need to be convinced to give a chance to AI in their diagnostic work – from what we gather, this is because they are open to anyting that improves the diagnosis, that saves lives. It is fairly obvious, that ML advances can improve diagnostic radiology imaging (Choy et al. 2018; Cabitza, Campagner and Sconfienza 2021). Furthermore, a study in Nature found that diagnosis can be particularly improved using causal ML for rather-rare or very-rare diseases, where the possible errors of diagnostics are typically more common and more serious (Richens, Lee and Johri 2020). On the other hand, machine learning methods might fail when incorporating causal reasoning (Patil, Szolovits and Schwartz; Rizzi 1994). Also AI appeared as complimentary to human doctors in several studies in the literature, for instance, AI performs better on vignette surveys (as opposed to medical records and claims) where doctors struggle, while they excel in highly contextual diagnosis where AI does not deliver. (Veloski et al. 2005; Semigran et al. 2015) Further research will be needed to figure out a more precise delineation of suitable tasks (Hoffman and Johnson 2019; Shneiderman 2020a, 2020b).

On the Role(s) of AI

Using AI as a tool, getting its services as an AI-assistant, and consulting it for a second opinion are widely diverse requirements, and they are unlikely to be delivered by the same AI solution. The various forms of AI to address different problem types is a subject for future research.

Furthermore, the literature suggest that different levels of mastery may need different type of AI support (Tschandl et al. 2020). We have found a bit of controversy here: on the one hand, less adept diagnosticians would benefit the most from AI, on the other hand, the higher the mastery, the better the judgement of the input from AI. Further research will be necessary to understand the relationship between the levels of master and the suitable type of AI.

This, combined with the understanding of how significantly the process of diagnosis is changing with the use of AI, together with Tschandl et al., we suggest that AI development must involve the actual users and testing needs to happen in the real-world context of the application. Only then it is reasonable to expect human+AI to outperform both human and AI (Bulten et al. 2021).

On Responsibility

Nowadays there are great debates on whether AI can have agency and what this means for responsibility – for instance, can AI be responsible? Although this problem appears significant both as a philosophical (Moor 1985; Dennett 1998; Floridi and Sanders 2004; Anderson 2011; Coeckelbergh 2020) and as a practical one (Čerka, Grigienė and Sirbikytė 2015; Daly et al. 2020; Moser, den Hond and Lindebaum 2021; Balasubramanian, Ye and Xu 2022), in our case, it seems that it can be simply resolved: medical doctors want to take responsibility. And, based on our data, we believe that this is not because they are worried about their jobs – they genuinely believe that this is the right thing to do. Additional implications of the concept of responsibility relate to AI design, specifically collaborative AI design, we address this in the final part of the discussion.

On Explainability

Explainability in AI is usually understood as the possibility to understand how a particular decision has been arrived at (Samek, Wiegand and Müller 2017; Samek and Müller 2019). As in many other areas, there is a high interest in explainable AI in the medical field. As a minimum, clients expect transparency and traceability of black-box ML/DL models (Holzinger et al. 2019). However, others suggest that one must go beyond explainable AI, because explainable medicine requires causality, where causality encompasses measurements for the quality of explanations (Patil, Szolovits and Schwartz 1981, August; Rizzi 1994; Holzinger et al. 2019; Richens, Lee and Johri 2020), it is important for human-AI interaction (Bologna and Hayashi 2017), and medical education, research, and clinical decision-making (Holzinger 2018; Holzinger et al. 2019). Our study, however, suggests that there is a whole other level of AI explainability that medics may be interested in: they want to understand the AI that they use. Not only the specific process it performs, but what it is like and how it generally does what it does. They want the science behind the AI implementation explained.

On Thinking Differently with AI

When professionals at a high level of mastery need to use a new tool, they usually only need a crash-course, online training, or other short and to the point training that is all about the tool. However, our findings reveal that using AI in the diagnosis process is far more complex. The reason, we believe, is that the decision process itself changes significantly, and this means that medical doctors (in our case) need to unlearn and relearn a highly complex process (cf Schön 1975; Argyris 1982, 2005). This, in a sense, complements the previously noted idea that actual users need to test AI solutions in real-live application contexts. Now, however, we can also see that the users will change as the consequence of this process, and the users' real-life experiences should be 'looped back' and considered in collaborative AI design. We believe that in supporting knowledgeable users with AI, this will become the criterion of the minimum viable AI product (cf Davenport and Seseri 2020).

Final Commentary

Our findings help understand what dermatologists, as a specific group of knowledge workers, working in a field where AI implementations already exist, expect from AI, how they would like to use it. They all agree that the final decision is the physician who takes responsibility. In order to be comfortable taking this responsibility, they want to understand what AI is, what it does and how it does what it does. They also understand that the current processes are not designed to incorporate AI; we need new processes. The future of AI design is collaborative, it will involve medics collaborating with AI developers, an essential part of which will be looping back the experience of using AI into the process design.

References

Anderson, S. L. 2011. Philosophical Concerns with Machine Ethics In *Machine Ethics*, edited by Anderson, M., Anderson, S. L., 162-167. Cambridge, UK: Cambridge University Press. doi.org/10.1017/CBO9780511978036.014

Argyris, C. 1982. The Executive Mind and Double-Loop Learning. *Organizational Dynamics*, 11(2): 5-22. doi.org/10.1016/0090-2616(82)90002-X

Argyris, C. 2005. Double-Loop Learning in Organizations: A Theory of Action Perspective In *Great Minds in Management: The Process of Theory Development*, edited by Smith, K. G., Hitt, M. A., 261-279. Oxford, UK: Oxford University Press.

Balasubramanian, N.; Ye, Y.; Xu, M. 2022. Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning. *Academy of Management Review*, 47(3): 448-465. doi.org/10.5465/amr.2019.0470

Baltruschat, I.; Steinmeister, L.; Nickisch, H.; Saalbach, A.; Grass, M.; Adam, G.; Knopp, T.; Ittrich, H. 2021. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. *European radiology*, 31(6): 3837-3845.

Bas, A.; Sinclair, M.; Dörfler, V. 2022. Sensing: The Elephant in the Room of Management Learning. *Management Learning*. doi.org/10.1177/13505076221077226

Bologna, G., Hayashi, Y. 2017. Characterization of symbolic rules embedded in deep DIMLP networks: a challenge to transparency of deep learning. *Journal of Artificial Intelligence and Soft Computing Research*, 7.

Brajer, N.; Cozzi, B.; Gao, M.; Nichols, M.; Revoir, M.; Balu, S.; Futoma, J.; Bae, J.; Setji, N.; Hernandez, A. 2020. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. *JAMA network open*, 3(2): e1920733-e1920733.

Bulten, W.; Balkenhol, M.; Belinga, J.-J. A.; Brilhante, A.; Çakır, A.; Egevad, L.; Eklund, M.; Farré, X.; Geronatsiou, K.; Molinié, V. 2021. Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Modern Pathology*, 34(3): 660-671.

Bussone, A.; Stumpf, S.; O'Sullivan, D. The role of explanations on trust and reliance in clinical decision support systems, *The role of explanations on trust and reliance in clinical decision support systems*, 2015: 160-169.

Bussone, A.; Stumpf, S.; O'Sullivan, D. (2015) The role of explanations on trust and reliance in clinical decision support systems, *The role of explanations on trust and reliance in clinical decision support systems*, 21-23 Oct. 2015: 160-169.

Cabitza, F.; Campagner, A.; Balsano, C. 2020. Bridging the "last mile" gap between AI implementation and operation:"data awareness" that matters. *Annals of translational medicine*, 8(7).

Cabitza, F.; Campagner, A.; Sconfienza, L. M. 2021. Studying human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading. *Health Information Science and Systems*, 9(1): 8. doi.org/10.1007/s13755-021-00138-8

Cai, C. J.; Winter, S.; Steiner, D.; Wilcox, L.; Terry, M. 2019. " Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-computer Interaction*, 3(CSCW): 1-24.

Čerka, P.; Grigienė, J.; Sirbikytė, G. 2015. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, 31(3): 376-389. doi.org/10.1016/j.clsr.2015.03.008

Chase, W. G., Simon, H. A. 1973a. The Mind's Eye in Chess In *Visual Information Processing*, edited by Chase, W. G., 215-281. New York, NY: Academic Press.

Chase, W. G., Simon, H. A. 1973b. Perception in Chess. *Cognitive Psychology*, 4(1): 55-81. doi.org/10.1016/0010-0285(73)90004-2

Chen, D.; Liu, S.; Kingsbury, P.; Sohn, S.; Storlie, C. B.; Habermann, E. B.; Naessens, J. M.; Larson, D. W.; Liu, H. 2019. Deep learning and alternative learning strategies for retrospective real-world clinical data. *NPJ digital medicine*, 2(1): 1-5.

Choi, E.; Bahadori, M. T.; Schuetz, A.; Stewart, W. F.; Sun, J. Doctor ai: Predicting clinical events via recurrent neural networks, *Doctor ai: Predicting clinical events via recurrent neural networks*, 2016: 301-318.

Choy, G.; Khalilzadeh, O.; Michalski, M.; Do, S.; Samir, A. E.; Pianykh, O. S.; Geis, J. R.; Pandharipande, P. V.; Brink, J. A.; Dreyer, K. J. 2018. Current applications and future impact of machine learning in radiology. *Radiology*, 288(2): 318-328.

Claburn, T. (2016, 4th August) IBM: AI Should Stand for Augmented Intelligence, *InformationWeek,* 1-19.

Coeckelbergh, M. 2020. Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots, with Implications for Thinking About Animals and Humans. *Minds and Machines*, 31(3): 337-360. doi.org/10.1007/s11023-020-09554-3

Daly, A.; Hagendorff, T.; Li, H.; Mann, M.; Vidushi, M.; Wagner, B.; Wang, W. W. 2020. AI, Governance and Ethics: Global Perspectives In *Constitutional Challenges in the Algorithmic Society*, edited by Pollicino, O., de Gregorio, G. Cambridge, UK: Cambridge University Press.

Davenport, T. H. 2018. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. Cambridge, MA: MIT Press.

Davenport, T. H., Bean, R. (2022, 11th April 2022) Clinical AI Gets the Headlines, but Administrative AI May Be a Better Bet, *Sloan Management Review*.

Davenport, T. H., Glover, W. J. 2018. Artificial intelligence and the augmentation of health care decision-making. *NEJM Catalyst*, 4(3).

Davenport, T. H., Seseri, R. (2020, 15th December 2020) What Is a Minimum Viable AI Product?, *MIT Sloan Management Review*.

Dennett, D. C. 1998. When HAL Kills, Who's to Blame? Computer Ethics In *HAL's Legacy: 2001's Computer as Dream and Reality*, edited by Stork, D. G., 351-365. Cambridge, MA: MIT Press. doi.org/10.7551/mitpress/3404.003.0018

van der Gijp, A.; Schaaf, M. F.; Schaaf, I.; Huige, J.; Ravesloot, C.; Schaik, J.; ten Cate, O. 2014. Interpretation of radiological images: Towards a framework of knowledge and skills. *Advances in health sciences education : theory and practice*, 19. doi.org/10.1007/s10459-013-9488-y

Dörfler, V. 2020. Artificial Intelligence In *Encyclopedia of Creativity* (3rd edition), edited by Runco, M. A., Pritzker, S. R., 57-64. Oxford, UK: Academic Press. doi.org/10.1016/B978-0-12-809324-5.23863-7

Dörfler, V. 2022a. Artificial Intelligence In *The SAGE Encyclopedia of Theory in Science, Technology, Engineering, and Mathematics* (Vol. 1), edited by Mattingly, J., 36-41. Thousand Oaks, CA: Sage.

Dörfler, V. 2022b. *What Every CEO Should Know About AI*. Cambridge, UK: Cambridge University Press. doi.org/10.1017/9781009037853

Dörfler, V.; Baracskai, Z.; Velencei, J. (2009) Knowledge Levels: 3-D Model of the Levels of Expertise, *AoM 2009: 69th Annual Meeting of the Academy of Management*, 7-11 August 2009, Chicago, IL.

Dörfler, V., Eden, C. 2014. Research on Intuition using Intuition In *Handbook of Research Methods on Intuition*, edited by Sinclair, M., 264-276. Cheltenham, UK: Edward Elgar Publishing. doi.org/10.4337/9781782545996.00031

Dörfler, V., Stierand, M. 2021. Bracketing: A Phenomenological Theory Applied Through Transpersonal Reflexivity. *Journal of Organizational Change Management*, 34(4): 778-793. doi.org/10.1108/JOCM-12-2019-0393

Dreyfus, H. L., Dreyfus, S. E. 1984. From Socrates to expert systems : The limits of calculative rationality. *Technology in Society*, 6(3): 217-233.

Dreyfus, H. L., Dreyfus, S. E. 1986/2000. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York, NY: The Free Press.

Dreyfus, H. L., Dreyfus, S. E. 1987. The Mistaken Psychological Assumptions Underlying Belief in Expert Systems In *Cognitive Psychology in Question*, edited by Costall, A., Still, A., 17-31. New York, NY: St. Martin's Press.

Ericsson, K. A., Smith, J. (Eds.) (1991) *Toward a General Theory of Expertise: Prospects and Limits*, Cambridge University Press, Cambridge, UK.

Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. 2019. A guide to deep learning in healthcare. *Nature medicine*, 25(1): 24-29.

Fagherazzi, G., Ravaud, P. 2019. Digital diabetes: Perspectives for diabetes prevention, management and research. *Diabetes & metabolism*, 45(4): 322-329.

Finlay, L. 2009. Debating Phenomenological Research Methods. *Phenomenology & Practice*, 3(1): 6-25. doi.org/10.29173/pandpr19818

Fisher, G.; Mayer, K.; Morris, S. 2021. From the Editors - Phenomenon-Based Theorizing. *Academy of Management Review*, 46(4): 631-639. doi.org/10.5465/amr.2021.0320

Floridi, L., Sanders, J. W. 2004. On the Morality of Artificial Agents. *Minds and Machines*, 14(3): 349-379. doi.org/10.1023/B:MIND.0000035461.63578.9d

Fröhlich, H.; Balling, R.; Beerenwinkel, N.; Kohlbacher, O.; Kumar, S.; Lengauer, T.; Maathuis, M. H.; Moreau, Y.; Murphy, S. A.; Przytycka, T. M. 2018. From hype to reality: data science enabling personalized medicine. *BMC medicine*, 16(1): 1-15.

Furnari, S. 2014. Interstitial Spaces: Microinteraction Settings and the Genesis of New Practices Between Institutional Fields. *Academy of Management Review*, 39(4): 439-462. doi.org/10.5465/amr.2012.0045

Gioia, D. A.; Corley, K. G.; Hamilton, A. L. 2012. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, 16(1): 15-31. doi.org/10.1177/1094428112452151

Gobet, F., Simon, H. A. 1996a. Recall of Random and Distorted Chess Positions: Implications for the Theory of Expertise. *Memory & Cognition*, 24(4): 493-503. doi.org/10.3758/BF03200937

Gobet, F., Simon, H. A. 1996b. Templates in Chess Memory: Mechanism for Re-calling Several Boards. *Cognitive Psychology*, 31(1): 1-40. doi.org/10.1006/cogp.1996.0011

Gobet, F., Simon, H. A. 2000. Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, 24(4): 651-682. doi.org/10.1016/S0364-0213(00)00031-8

Göndöcs, D., Dörfler, V. (2022) AI-enabled Organizational Learning Strategy, *EGOS 2022: 38th Colloquium of the European Group for Organization Studies*, 7-9 July 2022, Vienna, Austria.

Guo, Y.; Hao, Z.; Zhao, S.; Gong, J.; Yang, F. 2020. Artificial intelligence in health care: bibliometric analysis. *Journal of Medical Internet Research*, 22(7): e18228.

He, J.; Baxter, S. L.; Xu, J.; Xu, J.; Zhou, X.; Zhang, K. 2019. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1): 30-36.

Heaven, W. D. (2020, 30th November 2020) DeepMind’s protein-folding AI has solved a 50-year-old grand challenge of biology, *MIT Technology Review*.

Heidegger, M. 1927/1996. *Being and Time*. New York: Harper.

Heidegger, M. 1975/1988. *The Basic Problems of Phenomenology*. Bloomington, IN: Indiana University Press.

Hekler, A.; Utikal, J. S.; Enk, A. H.; Hauschild, A.; Weichenthal, M.; Maron, R. C.; Berking, C.; Haferkamp, S.; Klode, J.; Schadendorf, D. 2019. Superior skin cancer classification by the combination of human and artificial intelligence. *European Journal of Cancer*, 120: 114-121.

Hoffman, R. R., Johnson, M. 2019. The Quest for Alternatives to “Levels of Automation” and “Task Allocation” In *Human performance in automated and autonomous systems*, 43-68CRC Press.

Holzinger, A. (2018) From Machine Learning to Explainable AI, *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, 23-25 August 2018: 55-66.

Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. 2019. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4): e1312.

Hosny, A.; Parmar, C.; Coroller, T. P.; Grossmann, P.; Zeleznik, R.; Kumar, A.; Bussink, J.; Gillies, R. J.; Mak, R. H.; Aerts, H. J. W. L. 2018. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS medicine*, 15(11): e1002711.

Husserl, E. 1913a/1983. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Husserl, E. 1913b/1990. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: Second Book: Studies in the Phenomenology of Constitution*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Husserl, E. 1936/1970. *Crisis of European Sciences and Transcendental Phenomenology*. Evanston, Il: Northwestern University Press.

Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4).

Johnson, A. E. W.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1): 1-9.

Kather, J. N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.-A.; Gaiser, T.; Marx, A.; Valous, N. A.; Ferber, D. 2019. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1): e1002730.

Kumar, M. A.; Vimala, R.; Britto, K. R. A. 2019. A cognitive technology based healthcare monitoring system and medical data transmission. *Measurement*, 146: 322-332.

Lai, Y.; Kankanhalli, A.; Ong, D. Human-AI Collaboration in Healthcare: A Review and Research Agenda, *Human-AI Collaboration in Healthcare: A Review and Research Agenda*, 2021: 390.

Langley, A. 2021. What Is “This” a Case of? Generative Theorizing for Disruptive Times. *Journal of Management Inquiry*, 30(3): 251-258. doi.org/10.1177/10564926211016545

Larkin, J. H.; McDermott, J.; Simon, D. P.; Simon, H. A. 1980. Models of competence in solving physics problems. *Cognitive Science*, 4(4): 317-345. doi.org/10.1016/S0364-0213(80)80008-5

LeCun, Y.; Bengio, Y.; Hinton, G. 2015. Deep learning. *Nature*, 521(7553): 436-444. doi.org/10.1038/nature14539

Leyer, M., Schneider, S. 2021. Decision augmentation and automation with artificial intelligence: Threat or opportunity for managers? *Business Horizons*, 64(5): 711-724.

Luo, Y.; Tseng, H.-H.; Cui, S.; Wei, L.; Ten Haken, R. K.; El Naqa, I. 2019. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR| Open*, 1(1): 20190021.

Maassen, O.; Fritsch, S.; Palm, J.; Deffge, S.; Kunze, J.; Marx, G.; Riedel, M.; Schuppert, A.; Bickenbach, J. 2021. Future medical artificial intelligence application requirements and expectations of physicians in German University Hospitals: web-based survey. *J Med Internet Res*, 23(3): e26646.

Manz, C. R.; Chen, J.; Liu, M.; Chivers, C.; Regli, S. H.; Braun, J.; Draugelis, M.; Hanson, C. W.; Shulman, L. N.; Schuchter, L. M. 2020. Validation of a machine learning algorithm to predict 180-day mortality for outpatients with cancer. *JAMA oncology*, 6(11): 1723-1730.

May, E. C.; Utts, J. M.; Spottiswoode, S. J. P. 1995. Decision augmentation theory: Applications to the random number generator database. *Journal of Scientific Exploration*, 9(4): 453.

May, E. C.; Utts, J. M.; Spottiswoode, S. J. P. 2019. Decision Augmentation eory: Toward a Model of Anomalous Mental Phenomena1. *The Star Gate Archives: Reports of the United States Government Sponsored Psi Program, 1972-1995. Volume 3: Psychokinesis*: 332.

Meiring, C.; Dixit, A.; Harris, S.; MacCallum, N. S.; Brealey, D. A.; Watkinson, P. J.; Jones, A.; Ashworth, S.; Beale, R.; Brett, S. J. 2018. Optimal intensive care outcome prediction over time using machine learning. *PloS one*, 13(11): e0206862.

Miller, R. A. 2010. A history of the INTERNIST-1 and Quick Medical Reference (QMR) computer-assisted diagnosis projects, with lessons learned. *Yearbook of medical informatics*, 19(01): 121-136.

Montazeri, M.; Montazeri, M.; Montazeri, M.; Beigzadeh, A. 2016. Machine learning models in breast cancer survival prediction. *Technology and Health Care*, 24(1): 31-42.

Moor, J. H. 1985. What Is Computer Ethics? *Metaphilosophy*, 16(4): 266-275. doi.org/10.1111/j.1467-9973.1985.tb00173.x

Moser, C.; den Hond, F.; Lindebaum, D. 2021. Morality in the Age of Artificially Intelligent Algorithms. *Academy of Management Learning & Education*, 0(ja): null. doi.org/10.5465/amle.2020.0287

of the Madrid, O. C.; Reiz, A. N.; Sagasti, F. M.; González, M. Á.; Malpica, A. B.; Benítez, J. C. M.; Cabrera, M. N.; del Pino Ramírez, Á.; Perdomo, J. M. G.; Alonso, J. P. 2019. Big data and machine learning in critical care: Opportunities for collaborative research. *Medicina intensiva*, 43(1): 52-57.

Olekanma, O.; Dörfler, V.; Shafti, F. 2022. Stepping into the Participants' Shoes: The Trans-Positional Cognition Approach (TPCA). *International Journal of Qualitative Methods*, 21: 1-15. doi.org/10.1177/16094069211072413

Park, S. Y.; Kuo, P.-Y.; Barbarin, A.; Kaziunas, E.; Chow, A.; Singh, K.; Wilcox, L.; Lasecki, W. S. Identifying challenges and opportunities in human-AI collaboration in healthcare, *Identifying challenges and opportunities in human-AI collaboration in healthcare*, 2019: 506-510.

Patil, R. S.; Szolovits, P.; Schwartz, W. B. Causal understanding of patient illness in medical diagnosis, *Causal understanding of patient illness in medical diagnosis*, 1981, Vol. 81: 893-899.

Patil, R. S.; Szolovits, P.; Schwartz, W. B. (1981, August) Causal understanding of patient illness in medical diagnosis, *Causal understanding of patient illness in medical diagnosis*, 1981, Vol. Vol. 81: pp. 893-899.

Ployhart, R. E., Bartunek, J. M. 2019. Editors' Comments: There Is Nothing So Theoretical As Good Practice – A Call for Phenomenal Theory. *Academy of Management Review*, 44(3): 493-497. doi.org/10.5465/amr.2019.0087

Poursabzi-Sangdeh, F.; Goldstein, D. G.; Hofman, J. M.; Wortman Vaughan, J. W.; Wallach, H. Manipulating and measuring model interpretability, *Manipulating and measuring model interpretability*, 2021: 1-52.

Pratt, M. G. 2008. Fitting Oval Pegs Into Round Holes: Tensions in Evaluating and Publishing Qualitative Research in Top-Tier North American Journals. *Organizational Research Methods*, 11(3): 481-509. doi.org/10.1177/1094428107303349

Pratt, M. G. 2009. From the Editors: For the Lack of a Boilerplate: Tips on Writing Up (and Reviewing) Qualitative Research. *Academy of Management Journal*, 52(5): 856-862. doi.org/10.5465/amj.2009.44632557

Ribeiro, M. T.; Singh, S.; Guestrin, C. " Why should i trust you?" Explaining the predictions of any classifier, *" Why should i trust you?" Explaining the predictions of any classifier*, 2016: 1135-1144.

Richens, J. G.; Lee, C. M.; Johri, S. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(1): 3923. doi.org/10.1038/s41467-020-17419-7

Rizzi, D. A. 1994. Causal reasoning and the diagnostic process. *Theoretical medicine*, 15(3): 315-333.

Rong, G.; Mendez, A.; Assi, E. B.; Zhao, B.; Sawan, M. 2020. Artificial intelligence in healthcare: review and prediction case studies. *Engineering*, 6(3): 291-301.

Ruiz, A. A.; Wyszynska, P. K.; Laudanski, K. 2019. Narrative review of decision-making processes in critical care. *Anesthesia & Analgesia*, 128(5): 962-970.

Rush, B.; Celi, L. A.; Stone, D. J. 2019. Applying machine learning to continuously monitored physiological data. *Journal of clinical monitoring and computing*, 33(5): 887-893.

Samek, W., Müller, K.-R. 2019. Towards Explainable Artificial Intelligence In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, edited by Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L. K.; Müller, K.-R., 5-22. Cham: Springer International Publishing. doi.org/10.1007/978-3-030-28954-6_1

Samek, W.; Wiegand, T.; Müller, K.-R. 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*. doi.org/10.48550/arXiv.1708.08296

Saunders, M. N.; Lewis, P.; Thornhill, A. 2019. *Research Methods for Business Students* (8th edition),Pearson Education Limited.

Saunders, M. N., Townsend, K. 2016. Reporting and Justifying the Number of Interview Participants in Organization and Workplace Research. *British Journal of Management*, 27(4): 836-852. doi.org/10.1111/1467-8551.12182

Schön, D. A. 1975. Deutero-Learning in Organizations: Learning for. *Organizational Dynamics*, 4(1): 2-16. doi.org/10.1016/0090-2616(75)90001-7

Schütz, A. 1967. *The Phenomenology of the Social World*. Evanston, IL: Northwestern University Press.

Schütz, A. 1970. *On Phenomenology and Social Relations: Selected Writings*. Chicago, IL: University of Chicago Press.

Semigran, H. L.; Linder, J. A.; Gidengil, C.; Mehrotra, A. 2015. Evaluation of symptom checkers for self diagnosis and triage: audit study. *bmj*, 351.

Sharda, R.; Delen, D.; Turban, E. 2020. *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support* (11th edition),Pearson.

Shneiderman, B. 2020a. Design lessons from AI's two grand goals: human emulation and useful applications. *IEEE Transactions on Technology and Society*, 1(2): 73-82.

Shneiderman, B. 2020b. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6): 495-504.

Stierand, M. 2014. Developing creativity in practice: Explorations with world-renowned chefs. *Management Learning*, 46(5): 598-617. doi.org/10.1177/1350507614560302

Stierand, M., Dörfler, V. 2014. Researching Intuition in Personal Creativity In *Handbook of Research Methods on Intuition*, edited by Sinclair, M., 249-263. Cheltenham, UK: Edward Elgar Publishing. doi.org/10.4337/9781782545996.00030

Treasure-Jones, T.; Sarigianni, C.; Maier, R.; Santos, P.; Dewey, R. 2019. Scaffolded contributions, active meetings and scaled engagement: How technology shapes informal learning practices in healthcare SME networks. *Computers in Human Behavior*, 95: 1-13. doi.org/10.1016/j.chb.2018.12.039

Tschandl, P.; Rinner, C.; Apalla, Z.; Argenziano, G.; Codella, N.; Halpern, A.; Janda, M.; Lallas, A.; Longo, C.; Malvehy, J.; Paoli, J.; Puig, S.; Rosendahl, C.; Soyer, H. P.; Zalaudek, I.; Kittler, H. 2020. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8): 1229-1234. doi.org/10.1038/s41591-020-0942-0

Veloski, J.; Tai, S.; Evans, A. S.; Nash, D. B. 2005. Clinical vignette-based surveys: a tool for assessing physician practice variation. *American Journal of Medical Quality*, 20(3): 151-157.

Wang, L.; Lin, Z. Q.; Wong, A. 2020. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1): 1-12.

Willemink, M. J., Noël, P. B. 2019. The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. *European radiology*, 29(5): 2185-2195.

Xu, F.; Sepúlveda, M.-J.; Jiang, Z.; Wang, H.; Li, J.; Liu, Z.; Yin, Y.; Roebuck, M. C.; Shortliffe, E. H.; Yan, M. 2020. Effect of an artificial intelligence clinical decision support system on treatment decisions for complex breast cancer. *JCO Clinical Cancer Informatics*, 4: 824-838.

Yang, Q.; Steinfeld, A.; Zimmerman, J. Unremarkable ai: Fitting intelligent decision support into critical, clinical decision-making processes, *Unremarkable ai: Fitting intelligent decision support into critical, clinical decision-making processes*, 2019: 1-11.

Yu, K.-H.; Beam, A. L.; Kohane, I. S. 2018. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10): 719-731.