Artificial Intelligence In Medicine 149 (2024) 102769

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Research paper

AI in medical diagnosis: AI prediction & human judgment

Dóra Göndöcs [a], Viktor Dörfler [b,*]

[a] Széchenyi István University, Hungary
[b] University of Strathclyde Business School, United Kingdom

A R T I C L E I N F O

Keywords:
Medical diagnosis
Melanoma
Human-computer interaction
Augmented intelligence
Explainability
Responsible AI

A B S T R A C T

AI has long been regarded as a panacea for decision-making and many other aspects of knowledge work; as something that will help humans get rid of their shortcomings. We believe that AI can be a useful asset to support decision-makers, but not that it should replace decision-makers. Decision-making uses 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 medical diagnosis. We have found four aggregate dimensions along which the thinking of dermatologists can be described: the ways in which our participants chose to interact with AI, responsibility, 'explainability', and the new way of thinking (mindset) needed for working with AI. We believe that our findings will help physicians who might consider using AI in their diagnosis to understand how to use AI beneficially. 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 examine if our findings have relevance in the wider medical field and beyond.

1. Introduction

The rapid advances in AI development 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 a 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 a useful basis for comparing the participants' accounts. In addition, the lead author has access to the participants, which provides the benefit of an "insider view". We have designed an exploratory, qualitative empirical study, aimed at understanding how dermatologists think and feel about AI and using AI, as well as how the use of AI would alter their established diagnostic processes.

This positions our work in the broad area of 'future of work', where the emergence of automation and AI transform human work [1–3]. The human-computer interaction (HCI), which is an important aspect of future work, raises the question of automation or augmentation [4]. Both automation and augmentation are about identifying and resolving employees' weaknesses and limitations, but the approaches and possible consequences for the future of work are quite different (see [5] for a review based on three recent books). Employees rather dislike automation, but they have a much more positive attitude towards augmentation: while automation threatens their jobs and salaries, augmentation increases their work quality and productivity by complementing their weaknesses and enhancing their capabilities [5,6]. Although many different terms are used – including human-AI symbiosis, teams, and collaborations – currently AI-based technologies complement and augment human capabilities rather than replace them, and this can be expected for the foreseeable future [7–11]. The biggest challenge is consistently getting right the integration of AI into the existing organizational processes [12,13].

As we see it, the objective is not to replace the human decision-makers with AI; it is to produce accurate algorithmic predictions, which are then supplemented with the (value) judgments by 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 standpoint is what can be legitimately called a "decision support" [14], and what is referred to more recently as

* Corresponding author.
E-mail address: viktor.dorfler@strath.ac.uk (V. Dörfler).

https://doi.org/10.1016/j.artmed.2024.102769
Received 20 June 2023; Received in revised form 2 December 2023; Accepted 14 January 2024
Available online 20 January 2024

"decision augmentation" [15]. 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 judgments are made by the doctor: therefore the two components of the process are intertwined and the process is nonlinear. When AI is used, the two components are disentangled – contributed by separate entities – and thus they are arguably less intertwined; rather, they build on each other in a fashion that resembles a linear process. This does not suggest a lower complexity however, as what constitutes the complexity changes as well – namely, the human-AI interaction becomes an additional source 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 four themes (aggregate dimensions): the role of AI, 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 exploring the implications of the findings. We finish with a final commentary in which we account for the lessons learned for designing AI to support physicians in a manner that suits them.

2. 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."
>
> [16]

This definition, which can be traced back to the 1950s, does not say anything about AI accomplishing such tasks 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 (for a more detailed description see e.g. [17,18]). Specifically in the area of decision-making, including medical diagnosis, we believe that Davenport's ([19], p. 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.

2.1. AI in the medical field

There have been a growing number of publications on AI in medical research over the last decade [20–24]. 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 [4,25–37].

Among medical AI solutions, image processing was the main AI tool to advance disease detection in radiology. It did this 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 [27,38–40]. Expectations towards AI advances are extremely high, with the goal to improve medical healthcare as seen by physicians [22,24,41–46]. Thus, AI is viewed as changing the long-held status quo in healthcare, including the physicians' role, towards precision and personalized medicine [47]. 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 [48–51].

Much of AI application in medicine relies heavily on big data analysis, particularly image and speech processing, available due to recording an astonishing amount of medical data in a structured way in medical databases [40,42,52–54]. Such medical big data analysis uses various ML techniques, including DL, shallow or convolutional neural networks (CNN), vector machines, or random forests [29,55,56]. Among these techniques, DL shows great potential where large datasets are available, especially in the field of images, language, and speech processing [29]. Where such large datasets are not as available for studying medical conditions, other ML techniques may be superior [55]. 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 [57–60], clinical decisions [61], or workflows in hospital management [62–64]. There are also (predominantly hybrid) AI applications in marginal fields that produce important results in basic research, such as the recent successes of AlphaFold [65].

To conclude, the most critical precondition of emerging AI developments in healthcare is the data availability needed to develop and train algorithms. Therefore, it is of a 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 [66]. Among these applied solutions, vendors in radiology and other areas of imaging have started integrating AI into their products as a final (at least current) stage of the technological evolution in radiology [67]. In other cases, the recorded healthcare data has been increasingly used to validate the amount of data for medical predictions [68,69]. We also acknowledge the importance of the sociotechnical components necessary for successful AI implementations in clinical environments that Cabitza et al. [70] calls the "last mile gap" of AI bridging implementation and operation. Furthermore, besides the social and technical conditions, regulatory [71] and human factors might also hinder AI implementations in medical healthcare [72]. 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.

2.2. Use of AI in medical diagnosis

The first part of the literature review on AI in healthcare confirms that 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 focus on 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 [60].

Combining human expertise and AI, 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 [73]. 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 [74]. 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 [75]. The use of AI in melanoma diagnosis is not a unitary construct; Tschandl et al. [51] suggest using different AI-based applications at different levels of mastery [76]. 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 [77–87]. Tschandl et al. [51] also argue for the significance of testing the performance of AI-based solutions under real-world conditions and by the intended users, rather than testing isolated AI applications by programmers.

Applying Kasparov's law in the field of radiology, Cabitza et al. [38] call for using good interaction protocols, as those can contribute to improved decision-making, which may exceed the individual agents' performance. The same study [38] 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 [4] emphasize the importance of choosing the right augmentation approach when medical knowledge workers interact with AI. They offer a framework of five approaches that can be used in healthcare decision-making, by medical experts during their interaction with AI: step up, step aside, step in, step narrow, and step forward.

Narrowing our focus to human-AI interaction in medicine, we identified three leading groups of studies, which do not form a taxonomy, but which signify 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 [25,88–90]. Studies in the second group focus on the principles of designing AI applications that can be seamlessly implemented in medical practice [91]. 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 [26]. This study underlines the need for defining appropriate mental models, and for determining strategies when using AI in decision-making.

Overall, the literature 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. Next, we look into issues of AI ethics, such as explainability and responsibility.

2.3. Explainability and responsibility in AI

The landscape of AI ethics is vast, in many ways replicating the complexity of the generic ethics field, which is here accompanied with the sense of urgency (see e.g. [92]). Specifically within medicine, AI ethics issues are particularly burning; there is some excellent work done providing guidelines (e.g. [93]). In this paper we cannot take on the whole field, but we were not surprised that ethics concepts, particularly responsibility and explainability, play a significant role.

Responsible AI is a rather unfortunate term, as it appears to suggest an AI solution that is taking responsibility. This is, however, not the case; according to Trocin et al. [94] responsible AI is a field concerned specifically with establishing ethical principles and human values in order to reduce biases and promote fairness, facilitate interpretability and ensure robustness and security (see also [95–98]).

Explainable AI, in the simplest terms, means that the user should understand the outcome (e.g. recommendation or prediction) produced by AI. Barredo-Arrieta et al. [95] describe explainability with reference to a given audience, to whom the functioning of AI should be clear or easy to understand. This involves describing the causal relationships underlying the outcome either in specific terms (why this is the outcome in this case) or in general terms (how is AI in general arriving to recommendations). Therefore, causality plays a critical role in explainable AI [99]. However, determining what is understandable for humans remains an open question in this context.

AI is bridging the fields of cognitive science and computer science [100] by encompassing cognitive functions and real-world problem-solving to build systems that may deliver a performance similar to people [101]. Explainability has always been a crucial aspect of AI, as represented already by some early systems like the Advice Taker [102] that proposed reasoning. as a crucial aspect of AI.

In the early stages of AI, reasoning methods were logical and symbolic [103] where these symbolic expert systems were successful but with limited domain. Within these limitations, such early AI systems provided transparency of reasoning, thus supporting explainability [104–106].

In line with the development of ANN, developing causal models in AI systems to support explanation and understanding beyond simple pattern recognition became important [107]. Today's ML relying on statistical learning algorithms, large datasets, and available computational capacity [108], which should, in principle, enable evidence-based decision-making [109] across various domains by replicating statistical frequencies from previous data and improving it based on new data [110]. However, such ML faces challenges in explainability including sense-making, consideration of context, and decision-making under uncertainty, making it necessary to incorporate human expertise for usable intelligence [111–113].

Deep Learning (DL), as a subset of ML, is based on deep convolutional neural networks [114], has gained popularity due to its remarkable performance, often exceeding human-level performance in chosen limited domains [39]. DL has demonstrated success in various medical fields, such as skin cancer classification [29,115] and diabetic retinopathy identification [116]. However, recent trends in AI research emphasize the need for 'usable intelligence', which requires not only learning from prior data, extracting knowledge, and generalizing but the necessity of disentangling explanatory factors and understanding the context within application domains [117], and with that recognizing the indispensable role of human expertise alongside automated approaches [118].

The concepts of responsible and explainable AI have been introduced in a minimalistic way here, in order to provide sufficient grounding for our findings, but this will be further detailed when discussing those findings in Section 5.

3. 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 human-AI interaction in the context of medical diagnosis. We conducted semi-structured open-ended interviews with 17 dermatologists (see Table 1 for descriptive details), 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, following Saunders et al.'s [119] model known as the "research onion" [119, p. 130], starting with the most external layer of philosophical positioning, ending with the specific methods of data collection and analysis.

3.1. Philosophical and theoretical positioning

We loosely position our study within the interpretivist philosophical approach, specifically within the phenomenological tradition [120–127], 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, instead 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 [128,129]. For the same reason, we wanted to keep our options open, and 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 [130–132].

Furthermore, the research design of this study qualifies as an insider

Table 1
Descriptors of research participants.

Interviewee	Year of experience	Dermatology examination	Highest degree	Tried/used already AI	Openness/interest	Use AI as
A	20	Yes	PhD	In laboratory	Mid	Tool
B	10	Yes	PhD	In research	High	Tool
C	2	Yes	MD	In laboratory	High	Assistant
D	5	Yes	MD	None	High	Tool
E	5	Yes	PhD	In research	High	Tool
F	2	Resident	MD	None	High	Colleague
G	20	Yes	MD	None	High	Tool
H	3	Yes	MD	In laboratory	Mid	Colleague
I	1	Resident	PhD (ABD)	In research	High	Colleague
J	14	Yes	PhD	None	High	Tool
K	17	Yes	MD	None	High	Tool
L	20	Yes	MD	None	Low	Tool
M	18	Yes	PhD	Observed	Low	Tool
N	20	Yes	MD	None	Mid	Tool
O	10	Yes	MD	None	High	Assistant
P	4	Yes	MD	None	High	Assistant
Q	7	Yes	PhD	None	High	Tool
Aggregate	10.5	2 resident	10 MD	10 none	12 high	11 tool
		15 yes	7 PhD	6 yes[a]	5 mid/low	6 advanced[b]

[a] Yes includes both those who used AI in laboratory and those who used AI in research.
[b] Advanced includes "assistant" and "colleague" use in addition to the tool.

ethnography, as the first-named author works at the same clinic as most of 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).

3.2. Research participants

17 dermatologists have been interviewed, from various private and public healthcare institutions, and 11 of them work at the same private clinic, which 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 'in-house' access to half of the research participants and conduct the interviews as an insider. She asked these physicians directly about their thoughts, impressions, and feelings towards AI. The insiderness made the access easy, and the participants were likely honest in their responses.

The interviewees were at different stages of their careers (see Table 1), and presumably therefore in their levels of mastery [79]. 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. 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.

3.3. Collecting data

In total 17 dermatologists were interviewed in two rounds; nine in the first and eight in the second. An outline interview protocol was set up 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, 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 feel that we have understood the phenomenon that we were interested in at this point [133–135]. The interviews were all conducted in the local language, which is the native language of the interviewer as well as the interviewees. The analysis was also conducted in this language, and only quotes that were included in the paper were translated.

3.4. Analyzing data

To analyze the interviews, we used a variant of *thematic analysis*, which is a foundational method for qualitative analysis [136,137], has an established history in health research [138], the purpose of which is to search for patterns across our data set [119]. Thematic analysis is particularly flexible, which allowed us to formulate some pre-established 'a priori' themes based on the literature but also allowed for emergent 'in vivo' themes. An example of the former is the modified work process and an example of the latter is the roles that AI can play in the diagnostic process. Importantly, the a priori and in vivo themes not fully separable; for instance, we did expect to hear about explainable AI from the participants, but what we hear shifted our understanding of this a priori theme. The coding process was hierarchical, and we used Gioia's [139] approach to visualizing our code/data structure, which can be seen on Fig. 1.

The first named author, who conducted the interviews, also undertook the coding process. The a priori themes have been noted in advance, so that the codes can feed into these. Interesting thoughts of the participants were identified, and these were assigned codes that express that thought. For example, one participant said: "I prefer to check all spots with dermatoscopy, looking for that specific structure- and color-based characteristic of melanoma. I prefer to do it this way because it can cause surprises in both directions if I check first without and then with dermatoscopy; I see it quite differently." – this sentence received the code "Code 8: Preferential diagnostic methods (e.g. dermatoscopy)". The initial codes (26 in total) have been merged together, creating first-order concepts (13 in total). The first-order codes and the first-order concepts both represent the participants' viewpoints. The shift to the researchers' viewpoint happens when the second-order themes (11 in total) are created by merging first-order concepts. Finally, the second-order themes are merged into aggregate dimensions (4 in total), according to which we present our findings. This hierarchical coding process can be followed on Fig. 1. In the next subsection we explain the notion of bracketing, the role of which is to ensure the robustness of the research process and thus the reliability of the results.

3.5. 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 [140]. Importantly, in our interpretivist approach, the purpose of bracketing was to make use of pre-understandings and insider knowledge as a source of insight, rather than affecting the research in unknown ways [141–144].

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 judgment. 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 [142]. Typical questions in this stage would be along the lines of "so how do you know what your interviewee meant by XYZ?"

4. Analysis and findings

In the final step of the analysis, we synthesized the second-order themes into four aggregate dimensions (see Fig. 1). Each of these dimensions is outlined below. Following Pratt's [133] suggestion, we use "power quotes" in the main text.

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, but 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…."
> (Participant 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 than their own judgment."
> (Participant A)

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

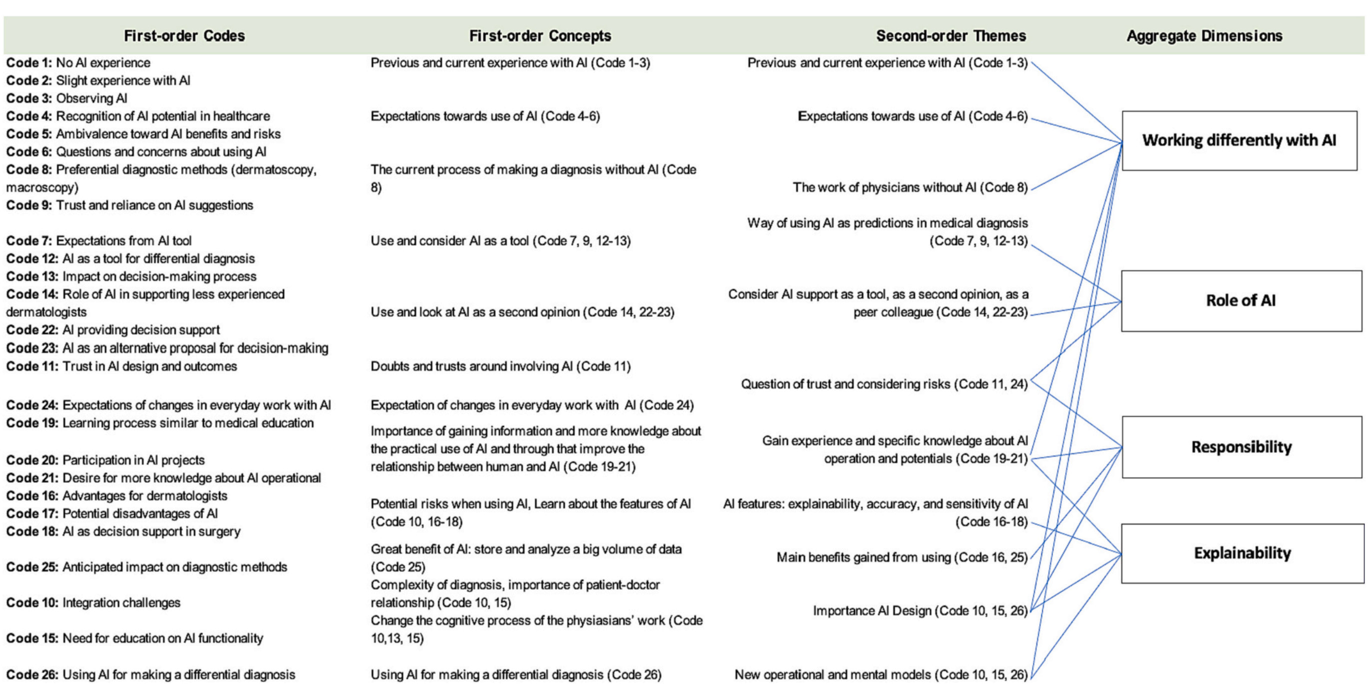

Fig. 1. Coding structure.

> "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 leitmotif, one suggestion 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 potential source 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, they only signify 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 another aggregate dimension, showing how our participants thought of *responsibility*, *explainability*, and about the need for a *different mindset* to benefit from using AI.

4.1. What is the role of AI?

We noticed that the interviewed dermatologists do not think about AI in the same way as AI vendors do, i.e. whether it is embedded AI or only 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 an assistant should be taught of along the lines of smartphones and such, which '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 if 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) judgment:

> "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 next aggregate dimension: the notion of responsibility.

4.2. Who is responsible?

Most of the interviewed physicians expressed a positive attitude towards 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 about a diagnosis, based on a value judgment. Only one participant speculated that perhaps AI will be able to take responsibility someday, 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, but 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 diagnosis:

> *"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 the patient 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 if it was them who needed to understand – this is further unpacked in the next subsection.

4.3. Can you explain?

Unsurprisingly, most of our interviewees suggested that the future of diagnosis will be a mix of human mastery and AI. 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 we 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 the AI was set up; how it works. They have a good understanding of science, and they want to understand AI in scientific terms.

4.4. *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 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 (prediction) 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 dermatoscopy, 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 these in more depth with a dermatoscope.

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

This is just one example that shows how 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 does the decision-making process of the diagnosis change entirely, but 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.

5. 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 anything that improves the diagnosis; that saves lives. It is fairly obvious that ML advances can improve diagnostic radiology imaging [28,38]. 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 [34]. On the other hand, machine learning methods might fail when incorporating causal reasoning [33,35]. AI also appeared as complementary 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 doctors excel in highly contextual diagnosis where AI does not deliver [36,145]. Further research will be needed to determine a more precise delineation of suitable tasks [146–148].

5.1. *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 are a subject for future research.

Furthermore, the literature suggests that different levels of mastery may need different type of AI support [51]. 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 judgment of the input from AI. Further research will be necessary to understand the relationship between the levels of mastery and the suitable type of AI.

Combining this with the understanding of how significantly the process of diagnosis is changing with the use of AI suggests 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 humans and AI individually [74].

5.2. *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? This poses a significant ethical problem both in philosophy [149–153] as well as practice [93,154–157]. However, 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.

5.3. *On explainability*

Explainability in AI is usually understood as the possibility to understand how a particular decision has been reached [158,159]. 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 [160]. However, others suggest that one must go beyond explainable AI, because explainable medicine requires causality and, in turn, causality encompasses measurements for the quality of explanations [33–35,160]. Therefore explainability is important for human-AI interaction [161], and medical education, research, and clinical decision-making [111,160]. Our study suggests that there is a whole other level of AI explainability that medics may be interested in, however: 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 works; they want the science behind the AI implementation explained.

5.4. On thinking differently with AI

When professionals at a high level of mastery need to use a new tool,

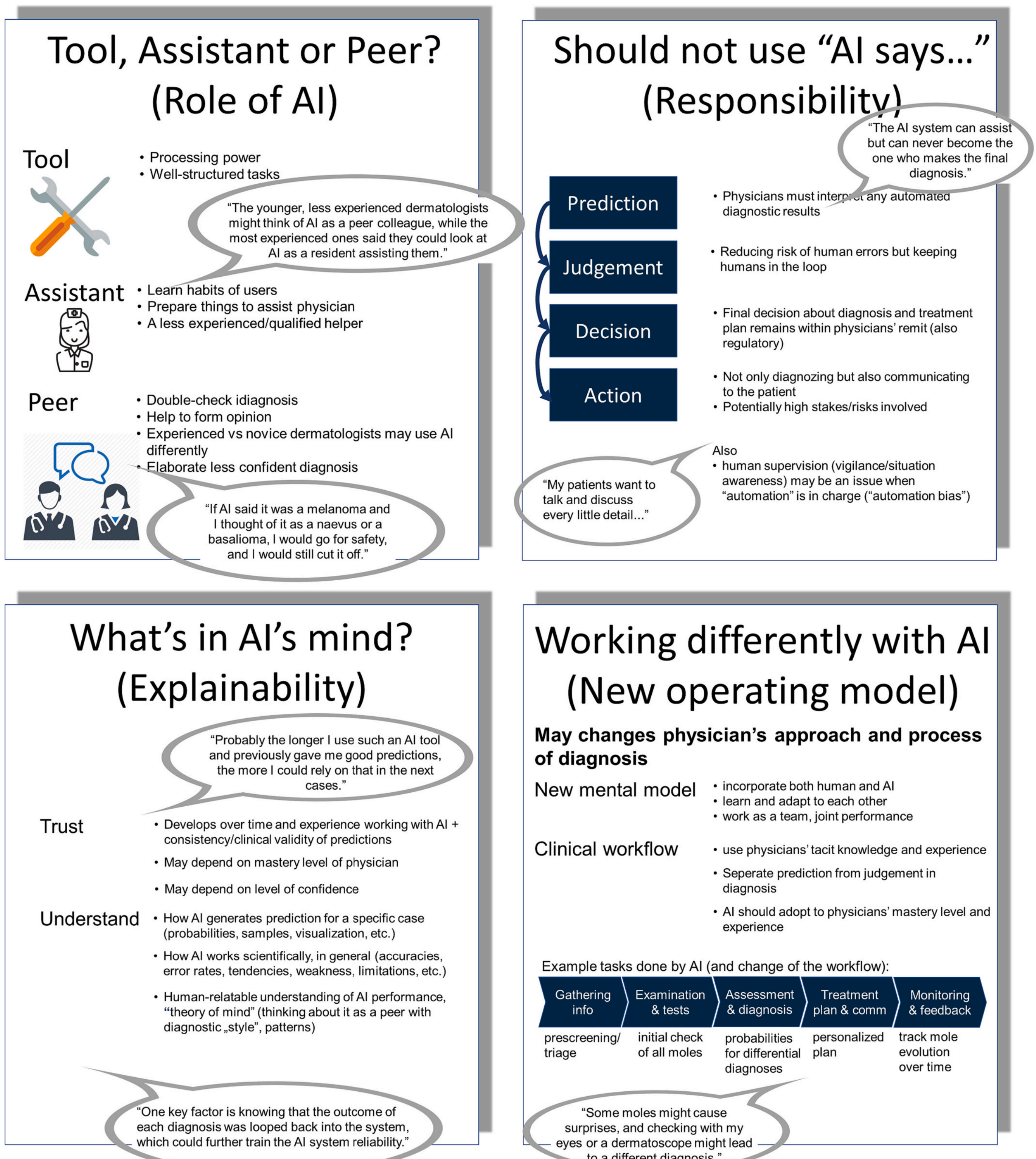

Fig. 2. The aggregate dimensions of the findings with practical examples.

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. We believe the reason 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 [162–164]). 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 [165]). Our participants showed awareness of the need for a new diagnostic process that incorporates explicitly both human and AI – but nobody knows yet what that process will look like, this will take a number of future studies.

6. Final commentary

To conclude our discussion, we take a step back, and look at our findings as 'lessons learned' for AI design in order to support physicians with technology that they are happy with. On Fig. 2 our findings are grouped on four panes, each illustrated by some of the previous used participant quotes that we find particularly illuminating. The first pane shows the modes of using AI: (a) the strength of the tool is the processing power, and can accomplish pre-defined well-structured tasks; (b) an assistant learns the physician's habits and prepares things before even asked; and (c) a 'colleague-AI' can help form opinion or double-check a diagnosis or even provide prompts for further elaborating the diagnosis – of course, it is likely that physicians at different levels of mastery use AI in substantially different ways. The second pane focuses on responsibility, where AI can help mitigate the risks of human errors, but humans also improve AI predictions through interpretation. As this is literally about life-and-death situations, figuring out how to minimize the risk is paramount. Importantly, the final responsibility is always with the physician, whose decision includes how to rely on AI. The third pane depicts explainability, which for these physicians means a deep scientific understanding of what the particular AI solution works, how it generates predictions, and they find a way to relate to AI, physicians may, gradually, develop a trust in AI. As an extra benefit, this may lead to an increased confidence and to a better understanding of their own diagnostic processes. The fourth pane, building on all the previous, foreshadows the necessity of a new mental model and by implication a new way of diagnosing, benefitting from the physicians' tacit knowledge as well as the AI's processing power to maximize the joint performance. Some aspects of this new diagnostic process are already visible, for instance the prediction and judgment will need to be decoupled, however, a new interaction between human and AI is added that increases complexity. These four lessons learned imply that AI vendors also need to create new processes, as they need either different AI solutions for the different roles, or one solution with different manifestations. Most importantly, however, the AI design with need to include the medics in their natural context, and they are ready and eager to participate, as it means saving lives.

CRediT authorship contribution statement

Dóra Göndöcs: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Viktor Dörfler:** Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Acemoglu D, Restrepo P. Artificial intelligence, automation, and work. In: Agrawal A, Gans J, Goldfarb A, editors. The economics of artificial intelligence: an agenda. University of Chicago Press; 2019. p. 197–236. http://www.nber.org/chapters/c14027.
[2] Brynjolfsson E, McAfee A. The second machine age: work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company; 2014.
[3] Ford S, Despeisse M. Additive manufacturing and sustainability: an exploratory study of the advantages and challenges. J Clean Prod 2016;137:1573–87. https://doi.org/10.1016/j.jclepro.2016.04.150.
[4] Davenport TH, Glover WJ. Artificial intelligence and the augmentation of health care decision-making. NEJM Catal 2018;4(3). https://catalyst.nejm.org/ai-technologies-augmentation-healthcare-decisions/.
[5] Raisch S, Krakowski S. Artificial intelligence and management: the automation-augmentation paradox. Acad Manage Rev 2021;46(1):192–210. https://doi.org/10.5465/amr.2018.0072.
[6] Wilson HJ, Daugherty PR. Collaborative intelligence: humans and AI are joining forces [article]. Harv Bus Rev 2018;96(4):114–23. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.
[7] Daugherty PR, Wilson HJ. Human + machine: reimagining work in the age of AI. Harvard Business Review Press; 2018.
[8] Davenport TH. From analytics to artificial intelligence. J Bus Anal 2018;1(2):73–80. https://doi.org/10.1080/2573234X.2018.1543535.
[9] Davenport TH, Harris JG. Competing on analytics: the new science of winning. 2nd ed. Harvard Business School Press; 2017.
[10] Davenport TH, Kirby J. Beyond automation. Harv Bus Rev 2015;93(6):58–65. http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/.
[11] Davenport TH, Miller SM. Working with AI: real stories of human-machine collaboration. MIT Press; 2022.
[12] Davenport TH, Euchner J. The rise of human-machine collaboration. Res Technol Manag 2023;66(1):11–5. https://doi.org/10.1080/08956308.2023.2142435.
[13] Davenport TH, Kirby J. Only humans need apply: winners and losers in the age of smart machines. HarperCollins; 2016.
[14] Sharda R, Delen D, Turban E. Analytics, data science, & artificial intelligence: systems for decision support. 11th ed. Pearson; 2020.
[15] Leyer M, Schneider S. Decision augmentation and automation with artificial intelligence: threat or opportunity for managers? Bus Horiz 2021;64(5):711–24. https://doi.org/10.1016/j.bushor.2021.02.026.
[16] Dörfler V. Artificial intelligence. In: Runco MA, Pritzker SR, editors. Encyclopedia of creativity. 3rd ed. Academic Press; 2020. p. 57–64. https://doi.org/10.1016/B978-0-12-809324-5.23863-7.
[17] Dörfler V. What every CEO should know about AI. Cambridge University Press; 2022.
[18] Dörfler V. Artificial Intelligence. In: Mattingly J, editor. The SAGE encyclopedia of theory in science, technology, engineering, and mathematics. vol. 1. Sage; 2023. p. 37–41. https://doi.org/10.4135/9781071872383.n15.
[19] Davenport TH. The AI advantage: how to put the artificial intelligence revolution to work. MIT Press; 2018.
[20] Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020;22(7):e18228. 10.2196/18228.
[21] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4). https://doi.org/10.1136/svn-2017-000101.
[22] Rong G, Mendez A, Assi EB, Zhao B, Sawan M. Artificial intelligence in healthcare: review and prediction case studies. Engineering 2020;6(3):291–301. https://doi.org/10.1016/j.eng.2019.08.015.
[23] Ruiz AA, Wyszynska PK, Laudanski K. Narrative review of decision-making processes in critical care. Anesth Anal 2019;128(5):962–70. https://doi.org/10.1213/ane.0000000000003683.
[24] Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2(10):719–31. https://doi.org/10.1038/s41551-018-0305-z.
[25] Bussone A, Stumpf S, O'Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. In: 2015 international conference on healthcare informatics, Dallas, TX; 2015. https://doi.org/10.1109/ICHI.2015.26. 21–23 Oct.
[26] Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. In: *Proceedings of the ACM on human-computer interaction, 3*(CSCW); 2019. p. 1–24. https://doi.org/10.1145/3359206.
[27] Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. In: MLHC 2016: 1st machine learning and healthcare conference, Los Angeles, CA; 2016. https://doi.org/10.48550/arXiv.1511.05942.
[28] Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current applications and future impact of machine learning in radiology. Radiology 2018;288(2):318–28. https://doi.org/10.1148/radiol.2018171820.

[29] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019;25(1):24–9. https://doi.org/10.1038/s41591-018-0316-z.
[30] Fagherazzi G, Ravaud P. Digital diabetes: perspectives for diabetes prevention, management and research. Diabetes Metab 2019;45(4):322–9. https://doi.org/10.1016/j.diabet.2018.08.012.
[31] Maassen O, Fritsch S, Palm J, Deffge S, Kunze J, Marx G, et al. Future medical artificial intelligence application requirements and expectations of physicians in German University Hospitals: web-based survey. J Med Internet Res 2021;23(3):e26646. https://doi.org/10.2196/26646.
[32] Miller RA. A history of the INTERNIST-1 and quick medical reference (QMR) computer-assisted diagnosis projects, with lessons learned. Yearb Med Inform 2010;19(1):121–36. https://doi.org/10.1055/s-0038-1638702.
[33] Patil RS, Szolovits P, Schwartz WB. Causal understanding of patient illness in medical diagnosis. In: IJCAI'81: 7th international joint conference on artificial intelligence, Vancouver, BC, Canada; 1981. https://doi.org/10.1007/978-1-4612-5108-8_16. 24–28 August 1981.
[34] Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. Nat Commun 2020;11(1):3923. https://doi.org/10.1038/s41467-020-17419-7.
[35] Rizzi DA. Causal reasoning and the diagnostic process. Theor Med 1994;15(3):315–33. https://doi.org/10.1007/BF01313345.
[36] Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. bmj 2015;351. https://doi.org/10.1136/bmj.h3480.
[37] Treasure-Jones T, Sarigianni C, Maier R, Santos P, Dewey R. Scaffolded contributions, active meetings and scaled engagement: how technology shapes informal learning practices in healthcare SME networks. Comput Hum Behav 2019;95:1–13. https://doi.org/10.1016/j.chb.2018.12.039.
[38] Cabitza F, Campagner A, Sconfienza LM. Studying human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading. Health Inf Sci Syst 2021;9(1):8. https://doi.org/10.1007/s13755-021-00138-8.
[39] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.
[40] Wang L, Lin ZQ, Wong A. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images. Sci Rep 2020;10:19549. https://doi.org/10.1038/s41598-020-76550-z.
[41] Davis R. Diagnostic reasoning based on structure and behavior. Artif Intell 1984;24(1):347–410. https://doi.org/10.1016/0004-3702(84)90042-0.
[42] De Kleer J, Williams BC. Diagnosing multiple faults. Artif Intell 1987;32(1):97–130. https://doi.org/10.1016/0004-3702(87)90063-4.
[43] De Kleer J, Mackworth AK, Reiter R. Characterizing diagnoses and systems. Artif Intell 1992;56(2–3):197–222. https://doi.org/10.1016/0004-3702(92)90027-U.
[44] Genesereth MR. The use of design descriptions in automated diagnosis. Artif Intell 1984;24(1):411–36. https://doi.org/10.1016/0004-3702(84)90043-2.
[45] Lucas PJF. Analysis of notions of diagnosis. Artif Intell 1998;105(1–2):295–343. https://doi.org/10.1016/S0004-3702(98)00081-2.
[46] Reiter R. A theory of diagnosis from first principles. Artif Intell 1987;32(1):57–95. https://doi.org/10.1016/0004-3702(87)90062-2.
[47] Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med 2018;16:150. https://doi.org/10.1186/s12916-018-1122-7.
[48] Claburn T. IBM: AI should stand for augmented intelligence. InformationWeek; 2016. p. 1–19. 4th August, https://www.informationweek.com/d/d-id/1326496.
[49] Lai Y, Kankanhalli A, Ong D. Human-AI collaboration in healthcare: a review and research agenda. In: HICSS 2021: 54th Hawaii international conference on system sciences, Honolulu, HI; 2021. http://hdl.handle.net/10125/70657.
[50] Park SY, Kuo P-Y, Barbarin A, Kaziunas E, Chow A, Singh K, et al. Identifying challenges and opportunities in human-AI collaboration in healthcare. In: CSCW '19: 2019 conference on computer supported cooperative work and social computing, Austin, TX; 2019. https://doi.org/10.1145/3311957.3359433. 9-13 November 2019.
[51] Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. Nat Med 2020;26(8):1229–34. https://doi.org/10.1038/s41591-020-0942-0.
[52] Buchanan BG, Sutherland GL, Feigenbaum EA. Heuristic DENDRAL: a program for generating explanatory hypotheses in organic chemistry. In: Meltzer B, Michie D, editors. Machine intelligence. Oliver & Boyd; 1969. p. 209–54.
[53] Davis R, Buchanan B, Shortliffe E. Production rules as a representation for a knowledge-based consultation program. Artif Intell 1977;8(1):15–45. https://doi.org/10.1016/0004-3702(77)90003-0.
[54] Kumar MA, Vimala R, Britto KRA. A cognitive technology based healthcare monitoring system and medical data transmission. Measurement 2019;146:322–32. https://doi.org/10.1016/j.measurement.2019.03.017.
[55] Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. NPJ Digit Med 2019;2(1):1–5. https://doi.org/10.1038/s41746-019-0122-0.
[56] Madrid OCot, Reiz AN, Sagasti FM, González MÁ, Malpica AB, Benítez JCM, et al. Big data and machine learning in critical care: opportunities for collaborative research. Med Intensiva 2019;43(1):52–7. https://doi.org/10.1016/j.medin.2018.06.002.
[57] Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med 2018;15(11):e1002711. https://doi.org/10.1371/journal.pmed.1002711.
[58] Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med 2019;16(1):e1002730. https://doi.org/10.1371/journal.pmed.1002730.
[59] Meiring C, Dixit A, Harris S, MacCallum NS, Brealey DA, Watkinson PJ, et al. Optimal intensive care outcome prediction over time using machine learning. PloS One 2018;13(11):e0206862. https://doi.org/10.1371/journal.pone.0206862.
[60] Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care 2016;24(1):31–42. https://doi.org/10.3233/THC-151071.
[61] Xu F, Sepúlveda M-J, Jiang Z, Wang H, Li J, Liu Z, et al. Effect of an artificial intelligence clinical decision support system on treatment decisions for complex breast cancer. JCO Clin Cancer Inform 2020;4:824–38. https://doi.org/10.1200/CCI.20.00018.
[62] Baltruschat I, Steinmeister L, Nickisch H, Saalbach A, Grass M, Adam G, et al. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. Eur Radiol 2021;31(6):3837–45. https://doi.org/10.1007/s00330-020-07480-7.
[63] Davenport TH, Bean R. Clinical AI gets the headlines, but administrative AI may be a better bet. Sloan Manage Rev 2022. 11th April 2022, https://sloanreview.mit.edu/article/clinical-ai-gets-the-headlines-but-administrative-ai-may-be-a-better-bet/amp.
[64] Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data. J Clin Monit Comput 2019;33(5):887–93. https://doi.org/10.1007/s10877-018-0219-z.
[65] Heaven WD. DeepMind's protein-folding AI has solved a 50-year-old grand challenge of biology. MIT Technol Rev 2020. 30th November 2020, https://www.technologyreview.com/2020/11/30/1012712/.
[66] Johnson AEW, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3(1):1–9. https://doi.org/10.1038/sdata.2016.35.
[67] Willemink MJ, Noël PB. The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. Eur Radiol 2019;29(5):2185–95. https://doi.org/10.1007/s00330-018-5810-7.
[68] Brajer N, Cozzi B, Gao M, Nichols M, Revoir M, Balu S, et al. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. JAMA Netw Open 2020;3(2):e1920733. https://doi.org/10.1001/jamanetworkopen.2019.20733.
[69] Manz CR, Chen J, Liu M, Chivers C, Regli SH, Braun J, et al. Validation of a machine learning algorithm to predict 180-day mortality for outpatients with cancer. JAMA Oncol 2020;6(11):1723–30. https://doi.org/10.1001/jamaoncol.2020.4331.
[70] Cabitza F, Campagner A, Balsano C. Bridging the "last mile" gap between AI implementation and operation:"data awareness" that matters. Ann Transl Med 2020;8(7). https://doi.org/10.21037/atm.2020.03.63.
[71] Müller H, Holzinger A, Plass M, Brcic L, Stumptner C, Zatloukal K. Explainability and causability for artificial intelligence-supported medical image analysis in the context of the European In Vitro Diagnostic Regulation. N Biotechnol 2022;70:67–72. https://doi.org/10.1016/j.nbt.2022.05.002.
[72] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30–6. https://doi.org/10.1038/s41591-018-0307-0.
[73] Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, et al. Superior skin cancer classification by the combination of human and artificial intelligence. Eur J Cancer 2019;120:114–21. https://doi.org/10.1016/j.ejca.2019.07.019.
[74] Bulten W, Balkenhol M, Belinga J-JA, Brilhante A, Çakır A, Egevad L, et al. Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. Mod Pathol 2021;34(3):660–71. doi:ar10.1038/s41379-020-0640-y.
[75] van der Gijp A, Schaaf MF, Schaaf I, Huige J, Ravesloot C, Schaik J, et al. Interpretation of radiological images: towards a framework of knowledge and skills. Adv Health Sci Educ Theory Pract 2014;19:565–80. https://doi.org/10.1007/s10459-013-9488-y.
[76] Göndöcs D, Dörfler V. AI-enabled organizational learning strategy. In: EGOS 2022: 38th colloquium of the European group for organization studies, Vienna, Austria; 2022. 7–9 July 2022, https://www.researchgate.net/publication/363049580.
[77] Chase WG, Simon HA. The mind's eye in chess. In: Chase WG, editor. Visual information processing. Academic Press; 1973. p. 215–81. https://doi.org/10.1016/B978-0-12-170150-5.50011-1.
[78] Chase WG, Simon HA. Perception in chess. Cogn Psychol 1973;4(1):55–81. https://doi.org/10.1016/0010-0285(73)90004-2.
[79] Dörfler V, Baracskai Z, Velencei J. Knowledge levels: 3-D model of the levels of expertise. In: AoM 2009: 69th annual meeting of the academy of management, Chicago, IL; 2009. 7–11 August 2009, https://www.researchgate.net/publication/308339223.
[80] Dreyfus HL, Dreyfus SE. From Socrates to expert systems : the limits of calculative rationality. Technol Soc 1984;6(3):217–33. https://doi.org/10.1016/0160-791X(84)90034-4.
[81] Dreyfus HL, Dreyfus SE. Mind over machine: the power of human intuition and expertise in the era of the computer. The Free Press; 1986/2000.
[82] Dreyfus HL, Dreyfus SE. The mistaken psychological assumptions underlying belief in expert systems. In: Costall A, Still A, editors. Cognitive psychology in question. St. Martin's Press; 1987. p. 17–31.

[83] Ericsson KA, Smith J, editors. Toward a general theory of expertise: prospects and limits. Cambridge University Press; 1991. https://books.google.co.uk/books?id=YaXOdNnCmA4C.
[84] Gobet F, Simon HA. Recall of random and distorted chess positions: implications for the theory of expertise. Mem Cognit 1996;24(4):493–503. https://doi.org/10.3758/BF03200937.
[85] Gobet F, Simon HA. Templates in chess memory: mechanism for re-calling several boards. Cogn Psychol 1996;31(1):1–40. https://doi.org/10.1006/cogp.1996.0011.
[86] Gobet F, Simon HA. Five seconds or sixty? Presentation time in expert memory. Cognit Sci 2000;24(4):651–82. https://doi.org/10.1016/S0364-0213(00)00031-8.
[87] Larkin JH, McDermott J, Simon DP, Simon HA. Models of competence in solving physics problems. Cognit Sci 1980;4(4):317–45. https://doi.org/10.1016/S0364-0213(80)80008-5.
[88] Luo Y, Tseng H-H, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR Open 2019;1(1):20190021. https://doi.org/10.1259/bjro.20190021.
[89] Poursabzi-Sangdeh F, Goldstein DG, Hofman JM, Wortman Vaughan JW, Wallach H. Manipulating and measuring model interpretability. In: 2021 CHI conference on human factors in computing systems, virtual conference; 2021. https://doi.org/10.48550/arXiv.1802.07810. 8-13 May 2021.
[90] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016. https://doi.org/10.48550/arXiv.1602.04938. 2016.
[91] Yang Q, Steinfeld A, Zimmerman J. *Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes 2019 CHI conference on human factors in computing systems*. 2019. https://doi.org/10.1145/3290605.3300468. 4-9 May 2019.
[92] Dörfler V, Cuthbert G. Dubito Ergo Sum: Exploring AI Ethics, HICSS 57: Hawaii International Conference on System Sciences, Honolulu, HI. 2024. p. 5587–96. https://hdl.handle.net/10125/107056.
[93] Müller H, Mayrhofer MT, van Veen E-B, Holzinger A. The ten commandments of ethical medical AI. Computer 2021;54(7):119–23. https://doi.org/10.1109/MC.2021.3074263.
[94] Trocin C, Mikalef P, Papamitsiou Z, Conboy K. Responsible AI for digital health: a synthesis and a research agenda. Inf Syst Front 2023;25(6):2139–57. https://doi.org/10.1007/s10796-021-10146-4.
[95] Barredo-Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012.
[96] Kumar P, Dwivedi YK, Anand A. Responsible artificial intelligence (AI) for value formation and market performance in healthcare: the mediating role of patient's cognitive engagement. Inf Syst Front 2023;25(6):2197–220. https://doi.org/10.1007/s10796-021-10136-6.
[97] Sambasivan N, Holbrook J. Toward responsible AI for the next billion users. Interactions 2018;26(1):68–71. https://doi.org/10.1145/3298735.
[98] Wang Y, Xiong M, Olya H. Toward an understanding of responsible artificial intelligence practices. 2020. 2020-01-07.
[99] Pearl J, Mackenzie D. The book of why: the new science of cause and effect. Basic books; 2018.
[100] Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: statistics, structure, and abstraction. Science 2011;331(6022):1279–85. https://doi.org/10.1126/science.1192788.
[101] Poole DI, Goebel RG, Mackworth AK. Computational intelligenceVol. 1. New York: Oxford University Press; 1998.
[102] McCarthy J. Recursive functions of symbolic expressions and their computation by machine, part I. Commun ACM 1960;3(4):184–95. https://doi.org/10.1145/367177.367199.
[103] Newell A, Shaw JC, Simon HA. Elements of a theory of human problem solving. Psychol Rev 1958;65(3):151–66. https://doi.org/10.1037/h0048495.
[104] Johnson DW, Johnson RT, Holubec EJ. The new circles of learning: cooperation in the classroom and school. ASCD; 1994.
[105] Lacave C, Díez FJ. A review of explanation methods for Bayesian networks. Knowl Eng Rev 2002;17(2):107–27. https://doi.org/10.1017/S026988890200019X.
[106] Swartout W, Paris C, Moore J. Explanations in knowledge systems: design for explainable expert systems. IEEE Expert 1991;6(3):58–64. https://doi.org/10.1109/64.87686.
[107] Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. Behav Brain Sci 2017;40:E253. https://doi.org/10.1017/S0140525X16001837.
[108] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. *TensorFlow: a system for large-scale machine learning* OSDI'16: 12th USENIX conference on operating systems design and implementation, Savannah, GA. 2016. https://doi.org/10.48550/arXiv.1605.08695.
[109] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015;349(6245):255–60. https://doi.org/10.1126/science.aaa8415.
[110] Michalski RS, Carbonell JG, Mitchell TM. Machine learning: an artificial intelligence approach. Springer-Verlag; 1983.
[111] Holzinger A. From machine learning to explainable AI. In: 2018 world symposium on digital intelligence for systems and machines (DISA), Košice, Slovakia; 2018. https://doi.org/10.1109/DISA.2018.8490530. 23-25 August 2018.
[112] Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv 2017. https://doi.org/10.48550/arXiv.1712.09923. Article 1712.09923.
[113] Holzinger A, Plass M, Kickmeier-Rust M, Holzinger K, Crişan GC, Pintea C-M, et al. Interactive machine learning: experimental evidence for the human in the algorithmic loop. Appl Intell 2019;49(7):2401–14. https://doi.org/10.1007/s10489-018-1361-5.
[114] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.
[115] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8. https://doi.org/10.1038/nature21056.
[116] Ting DSW, Cheung CY-L, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318(22):2211–23. https://doi.org/10.1001/jama.2017.18152.
[117] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35(8):1798–828. https://doi.org/10.1109/TPAMI.2013.50.
[118] Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform 2016;3(2):119–31. https://doi.org/10.1007/s40708-016-0042-6.
[119] Saunders MN, Lewis P, Thornhill A. Research methods for business students. 9th ed. Pearson Education; 2023.
[120] Finlay L. Debating phenomenological research methods. Phenomenol Pract 2009;3(1):6–25. https://doi.org/10.29173/pandpr19818.
[121] Heidegger M. Being and time. Harper; 1927/1996.
[122] Heidegger M. The basic problems of phenomenology. Indiana University Press; 1975/1988.
[123] Husserl E. Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: first book: general introduction to a pure phenomenology. Kluwer Academic Publishers; 1913.
[124] Husserl E. Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: second book: studies in the phenomenology of constitution. Kluwer Academic Publishers; 1913.
[125] Husserl E. Crisis of European sciences and transcendental phenomenology. Northwestern University Press; 1936/1970.
[126] Schütz A. The phenomenology of the social world. Northwestern University Press; 1967.
[127] Schütz A. On phenomenology and social relations: selected writings. University of Chicago Press; 1970.
[128] Bas A, Sinclair M, Dörfler V. Sensing: the elephant in the room of management learning. Manag Learn 2023;54(4):489–510. https://doi.org/10.1177/13505076221077226.
[129] Furnari S. Interstitial spaces: microinteraction settings and the genesis of new practices between institutional fields. Acad Manage Rev 2014;39(4):439–62. https://doi.org/10.5465/amr.2012.0045.
[130] Fisher G, Mayer K, Morris S. From the editors - phenomenon-based theorizing. Acad Manage Rev 2021;46(4):631–9. https://doi.org/10.5465/amr.2021.0320.
[131] Langley A. What is "this" a case of? Generative theorizing for disruptive times. J Manag Inq 2021;30(3):251–8. https://doi.org/10.1177/10564926211016545.
[132] Ployhart RE, Bartunek JM. Editors' comments: there is nothing so theoretical as good practice – a call for phenomenal theory. Acad Manage Rev 2019;44(3):493–7. https://doi.org/10.5465/amr.2019.0087.
[133] Pratt MG. Fitting oval pegs into round holes: tensions in evaluating and publishing qualitative research in top-tier North American journals. Organ Res Methods 2008;11(3):481–509. https://doi.org/10.1177/1094428107303349.
[134] Pratt MG. From the editors: for the lack of a boilerplate: tips on writing up (and reviewing) qualitative research. Acad Manage J 2009;52(5):856–62. https://doi.org/10.5465/amj.2009.44632557.
[135] Saunders MN, Townsend K. Reporting and justifying the number of interview participants in organization and workplace research. Br J Manag 2016;27(4):836–52. https://doi.org/10.1111/1467-8551.12182.
[136] Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.
[137] Braun V, Clarke V. Thematic analysis. In: APA handbook of research methods in psychology, vol 2: research designs: quantitative, qualitative, neuropsychological, and biological. American Psychological Association; 2012. p. 57–71. https://doi.org/10.1037/13620-004.
[138] Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well Being 2014;9:26152. https://doi.org/10.3402/qhw.v9.26152.
[139] Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research: notes on the Gioia methodology. Organ Res Methods 2013;16(1):15–31. https://doi.org/10.1177/1094428112452151.
[140] Olekanma O, Dörfler V, Shafti F. Stepping into the participants' shoes: the trans-positional cognition approach (TPCA). Int J Qual Methods 2022;21:1–15. https://doi.org/10.1177/16094069211072413.
[141] Dörfler V, Eden C. Research on intuition using intuition. In: Sinclair M, editor. Handbook of research methods on intuition. Edward Elgar Publishing; 2014. p. 264–76. https://doi.org/10.4337/9781782545996.00031.
[142] Dörfler V, Stierand M. Bracketing: a phenomenological theory applied through transpersonal reflexivity. J Organ Chang Manag 2021;34(4):778–93. https://doi.org/10.1108/JOCM-12-2019-0393.

[143] Stierand M. Developing creativity in practice: explorations with world-renowned chefs. Manag Learn 2015;46(5):598–617. https://doi.org/10.1177/1350507614560302.
[144] Stierand M, Dörfler V. Researching intuition in personal creativity. In: Sinclair M, editor. Handbook of research methods on intuition. Edward Elgar Publishing; 2014. p. 249–63. https://doi.org/10.4337/9781782545996.00030.
[145] Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: a tool for assessing physician practice variation. Am J Med Qual 2005;20(3):151–7. https://doi.org/10.1177/1062860605274520.
[146] Hoffman RR, Johnson M. The quest for alternatives to "levels of automation" and "task allocation". In: Mouloua M, Hancock PA, editors. Human performance in automated and autonomous systems. CRC Press; 2019. p. 43–68.
[147] Shneiderman B. Design lessons from AI's two grand goals: human emulation and useful applications. IEEE Trans Technol Soc 2020;1(2):73–82. https://doi.org/10.1109/TTS.2020.2992669.
[148] Shneiderman B. Human-centered artificial intelligence: reliable, safe & trustworthy. Int J Hum-Comput Interact 2020;36(6):495–504. https://doi.org/10.1080/10447318.2020.1741118.
[149] Anderson SL. Philosophical concerns with machine ethics. In: Anderson M, Anderson SL, editors. Machine ethics. Cambridge University Press; 2011. p. 162–7. https://doi.org/10.1017/CBO9780511978036.014.
[150] Coeckelbergh M. 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 Mach 2020. https://doi.org/10.1007/s11023-020-09554-3.
[151] Dennett DC. When HAL kills, who's to blame? computer ethics. In: Stork DG, editor. HAL's legacy: 2001's computer as dream and reality. MIT Press; 1998. p. 351–65. https://doi.org/10.7551/mitpress/3404.003.0018.
[152] Floridi L, Sanders JW. On the morality of artificial agents. Minds Mach 2004;14(3):349–79. https://doi.org/10.1023/B:MIND.0000035461.63578.9d.
[153] Moor JH. What is computer ethics? Metaphilosophy 1985;16(4):266–75. https://doi.org/10.1111/j.1467-9973.1985.tb00173.x.
[154] Balasubramanian N, Ye Y, Xu M. Substituting human decision-making with machine learning: implications for organizational learning. Acad Manage Rev 2020;0(ja). https://doi.org/10.5465/amr.2019.0470. null.
[155] Čerka P, Grigienė J, Sirbikytė G. Liability for damages caused by artificial intelligence. Comput Law Secur Rev 2015;31(3):376–89. https://doi.org/10.1016/j.clsr.2015.03.008.
[156] Daly A, Hagendorff T, Li H, Mann M, Vidushi M, Wagner B, et al. AI, governance and ethics: global perspectives. In: Pollicino O, de Gregorio G, editors. Constitutional challenges in the algorithmic society. Cambridge University Press; 2021. p. 182–201. https://doi.org/10.1017/9781108914857.010.
[157] Moser C, den Hond F, Lindebaum D. Morality in the age of artificially intelligent algorithms. Acad Manag Learn Edu 2022;21(1):139–55. https://doi.org/10.5465/amle.2020.0287.
[158] Samek W, Müller K-R. Towards explainable artificial intelligence. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R, editors. Explainable AI: interpreting, explaining and visualizing deep learning. Springer International Publishing; 2019. p. 5–22. https://doi.org/10.1007/978-3-030-28954-6_1.
[159] Samek W, Wiegand T, Müller K-R. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. 2017. https://doi.org/10.48550/arXiv.1708.08296.
[160] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev: Data Min Knowl Discov 2019;9(4):e1312. https://doi.org/10.1002/widm.1312.
[161] Bologna G, Hayashi Y. Characterization of symbolic rules embedded in deep DIMLP networks: a challenge to transparency of deep learning. J Artif Intell Soft Comput Res 2017;7(4):265–86. https://doi.org/10.1515/jaiscr-2017-0019.
[162] Argyris C. The executive mind and double-loop learning. Organ Dyn 1982;11(2):5–22. https://doi.org/10.1016/0090-2616(82)90002-X.
[163] Argyris C. Double-loop learning in organizations: a theory of action perspective. In: Smith KG, Hitt MA, editors. Great minds in management: the process of theory development. Oxford University Press; 2005. p. 261–79.
[164] Schön DA. Deutero-learning in organizations: learning for. Organ Dyn 1975;4(1):2–16. https://doi.org/10.1016/0090-2616(75)90001-7.
[165] Davenport TH, Seseri R. What is a minimum viable AI product? MIT Sloan Manag Rev 2020. 15th December 2020, https://sloanreview.mit.edu/article/what-is-a-minimum-viable-ai-product/.