Received: 8 September 2020 | Revised: 15 December 2020 | Accepted: 28 December 2020

DOI: 10.1111/jep.13535

ORIGINAL PAPER

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

WILEY

Co-designing diagnosis: Towards a responsible integration of Machine Learning decision-support systems in medical diagnostics

Olya Kudina PhD[1†] | Bas de Boer PhD[2†]

[1]Department of Values, Technology & Innovation, Section on Ethics and Philosophy of Technology, Delft University of Technology, Delft, the Netherlands

[2]Philosophy Department, University of Twente, Enschede, the Netherlands

Correspondence
Olya Kudina, Department of Values, Technology & Innovation, Section on Ethics and Philosophy of Technology, Delft University of Technology, Building 31, Jaffalaan 5, 2628 BX Delft, the Netherlands.
Email: o.kudina@tudelft.nl

Funding information
H2020 European Research Council, Grant/Award Number: 788321; 4TU Pride and Prejudice project under the High Tech for a Sustainable Future programme

Abstract

Rationale: This paper aims to show how the focus on eradicating bias from Machine Learning decision-support systems in medical diagnosis diverts attention from the hermeneutic nature of medical decision-making and the productive role of bias. We want to show how an introduction of Machine Learning systems alters the diagnostic process. Reviewing the negative conception of bias and incorporating the mediating role of Machine Learning systems in the medical diagnosis are essential for an encompassing, critical and informed medical decision-making.

Methods: This paper presents a philosophical analysis, employing the conceptual frameworks of hermeneutics and technological mediation, while drawing on the case of Machine Learning algorithms assisting doctors in diagnosis. This paper unravels the non-neutral role of algorithms in the doctor's decision-making and points to the dialogical nature of interaction not only with the patients but also with the technologies that co-shape the diagnosis.

Findings: Following the hermeneutical model of medical diagnosis, we review the notion of bias to show how it is an inalienable and productive part of diagnosis. We show how Machine Learning biases join the human ones to actively shape the diagnostic process, simultaneously expanding and narrowing medical attention, highlighting certain aspects, while disclosing others, thus mediating medical perceptions and actions. Based on that, we demonstrate how doctors can take Machine Learning systems on board for an enhanced medical diagnosis, while being aware of their non-neutral role.

Conclusions: We show that Machine Learning systems join doctors and patients in co-designing a triad of medical diagnosis. We highlight that it is imperative to examine the hermeneutic role of the Machine Learning systems. Additionally, we suggest including not only the patient, but also colleagues to ensure an encompassing diagnostic process, to respect its inherently hermeneutic nature and to work productively with the existing human and machine biases.

KEYWORDS
hermeneutics, Machine Learning, medical diagnosis, technological mediation

† The authors have contributed equally to the development of this article.

1 | INTRODUCTION

At the beginning of the 2010s, Drew et al asked a group of 24 radiologists to perform a familiar lung nodule detection task. The radiologists were asked to search for nodules on CT-scans in which, unbeknown to the doctors, a 29×50 mm image of a gorilla had been included. Strikingly, 83% of the radiologists did not notice the gorilla, even though that eye-tracking revealed that most of the radiologists who missed the gorilla looked directly at the place where it was located. In the psychological literature, this phenomenon is known as *inattentional blindness* during which attention to a particular task makes us blind to other salient phenomena. This suggests that expert radiologists search for particular anomalies located at particular places in the lungs, such that unexpected anomalies at unexpected locations might go unnoticed, potentially with severe medical consequences.[1] While bias is no stranger to medical encounters of doctors with patients, it is not clear what happens when it is coupled with the introduction of Machine Learning (ML) algorithms in assisting medical diagnosis.

One of the central promises of the use of ML in medical diagnostics is that it will make medical diagnoses more objective by eliminating forms of human bias.[2] Bias might be caused by deficiencies inherent to human perception such as discussed in the example above, but also of other biases arising in doctor-patient relations, which might be caused by prejudices on the side of the doctor.[3] Because of this, so it is postulated, ML diagnostic systems will be a significant improvement to human capabilities in clinical decision-making in terms of diagnosis, prognosis, and treatment recommendation.[4(p3)] In sum, the introduction of ML systems in medical diagnostics is often presented as an important augmentation to, or even as threatening to replace, human medical expertise.

In this paper, we critically scrutinize the promise of ML in diagnostic practice by drawing attention to its relationship with medical expertise. First, we briefly discuss the ethical issues often discussed in relation to the introduction of ML into healthcare broadly conceived. Second, we flesh out a hermeneutic understanding of medical expertise and the diagnostic process, in which biases have a productive, rather than a distorting role. Third, we make clear how ML can be understood as mediating the hermeneutic process through which a medical diagnosis is established. On the basis of this, we suggest that the introduction of ML systems in medical diagnostics should not be framed as requiring to make a choice for either the expertise of clinicians or the alleged objectivity of ML systems. Finally, we offer some starting points for how ML can be seen as a dialogical partner for medical experts.

2 | FROM BIAS TO THE QUESTION OF EXPERTISE

Since ML systems are often presented as a solution to the problem of bias, it should not come as a surprise that both developers of ML systems and doctors that critically reflect on ML search for biases that might be present in algorithms used to make medical diagnoses. When ML systems also suffer from biases, they effectively undermine the promise of developing a more objective way of clinical decision-making. For example, the data-sets on which ML systems rely might be biased towards particular healthcare systems, as was the case when IBM launched Watson for Oncology. This assistive system was based on data collected in the American healthcare system, having a bias towards specific ways of drug-prescription that are deemed normal in the USA, but did not align with cultures of drug-prescription in other countries, such as Taiwan.[5] Furthermore, existing datasets typically exist for medical problems suffered by white men, leading to a poor performance rate when applied to other groups, such as younger black women.[4(p5)] For ML systems to live up to their promise of objectivity, it is thus crucial to identify and eliminate such biases in datasets.

This also explains the centrality of another concern: the opacity of the algorithms on the basis of which clinical decisions are made. Algorithms can be opaque to users because they lack the appropriate training allowing to understand how the algorithm comes to a certain diagnosis, or when it is inherent to the design of the algorithm that its workings are not intelligible to humans. In both cases, opacity hampers the possibility of detecting potential biases. And if the opaqueness of algorithms can indeed not be circumvented, then also their potential to make medical diagnoses more objective by eliminating bias cannot be properly assessed. As a result, researchers are worried about the potentially ethically problematic outcomes that can be expected when ML systems are constructed as black boxes, making it more likely that problems such as the ones mentioned in the previous paragraph might remain unnoticed.[6]

The focus on bias of clinicians and developers is to a large extent mirrored in policy documents discussing the impact of ML, in healthcare and beyond.[7-10] Some of the frequently discussed risks concern the individual harm that can be induced due to the algorithms that make decisions about treatment on the basis of biased datasets or the unfair advantage that people that are represented in (biased) datasets have over the ones that are typically underrepresented.[11(pp23-24)] In order to avoid such biases and to prevent harm and unfairness, it is often stressed that algorithms should be designed in accordance with principles of transparency and/or explainability.[12]

The opacity of ML systems is especially concerning since clinicians reportedly tend to find it challenging to counter algorithm-based judgements and provide independent diagnoses or suggestions for treatment, affecting how they value their own judgements.[13] As a consequence, if the decisions of ML systems are biased, then it seems likely that these biases are reproduced or reified due to them remaining effectively unchallenged.[7(p181)] Insofar as ethical discussions take objectivity (or the absence thereof) in clinical decision-making to be the central issue at stake, the negative impact of bias must be a focal point, as it is this issue that makes it that ML systems cannot live up to their promises.

However, more recently, ethical discussions on the use of ML in clinical decision-making have started to address ethical concerns

related to the introduction of ML beyond the narrow focus on bias.[4] In fact, so it is argued, the belief that algorithms are—in contrast with human beings—harbours of objectivity is a "carefully crafted myth".[4(p4),14] While algorithms might outperform humans when it comes to pattern recognition, their ability to attach meaning to patterns or make inferences on the basis of them remains unclear.[2] In one way or another, this suggests that instead of speaking of a competition between humans and ML systems, discussions about how to integrate ML in healthcare practices should be augmented through exploring what kind of collaborations between doctors and ML systems are *desirable*.[15] For instance, in the field of mental healthcare, physicians and patients engage in developing ML systems in the patients' smartphones for onset symptoms detection.[16] In pathology, collaborative efforts take place to design diagnostic AI assistant that capitalizes on the mental models of the clinicians, while utilizing optimization techniques of ML systems.[17] Radiologists propose strategies on how to practically integrate ML systems for collaboration in the work practice: while they can remove the workload by taking on normal examinations (eg, head CTs or MRIs for headache), the current business strategies do not allow integrating the input of ML systems in the administrative flows or reimbursement schemes.[18] While evidence on including ML systems as collaborators continues to surface, the early practice-driven efforts already hint at the adjustments to the healthcare process and the reconfiguration of the medical profession[19] that the recognition of ML systems as collaborators requires.

Insofar as a medical diagnosis is concerned with the *interpretation* of the patient and her health status, ethical discussions that narrowly conceive of an ethics of AI as an ethics of bias might neglect the way ML systems shape medical expertise. After all, *if* clinical decision-making is more than simple pattern recognition and requires another form of expertise, it is crucial to explore what this expertise is, and in what sense ML systems might contribute to it. This we will do in the next two sections of this paper.

3 | EXPERTISE AND MEDICAL DIAGNOSIS: GADAMER'S NOTION OF FORE-UNDERSTANDING

Recently, it has been argued that the demands of transparency and explainability—while important—hold ML algorithms "to an unrealistically high standard [...], possibly owing to an unrealistically high estimate of the degree of transparency attainable from human decision-makers".[20(p662)] Regardless of whether it is justified that the standards we set for ML are exceptionally high, Zerili et al importantly point to the need to clarify what we take medical (or diagnostic) expertise to be, and if and how it can be outperformed by ML. In other words, a discussion about the potential biases in ML systems must be informed by a discussion about what we consider good *human* forms of decision-making[21] and the nature of expertise exercised by clinicians.

Recently, Grote and Berens argued that the use of ML in diagnostic practice changes the epistemic conditions under which medical expertise is exercised.[22] They note that medical diagnosis is often not a solitary activity of a clinician, but one that also involves discussions with other clinicians that function as peers to diagnostic judgements. The peers offer epistemic import that might support or criticize a certain judgement, making diagnostic expertise effectively distributed among different individuals. These different individuals can engage in a dialogue each providing reasons for or against a certain diagnosis, and this dialogical process eventually will improve the diagnostic process.[22(p207)] Within such diagnostic processes, clinicians use all kinds of technologies (eg, imaging technologies) that influence, and might support their judgements, making those also a crucial part of diagnostics already.[23] However, what is crucially different about the involvement of ML as a diagnostic peer is that—insofar the inferences it makes cannot be articulated—clinicians are unable to judge whether or not their import is epistemically credible.[22(p207)]

The idea that medical diagnoses presuppose some form of situated or distributed expertise nicely illustrates the uncertainties that ML might introduce into diagnostic practices. However, and this is what they seemingly have in common with developers of ML systems, Grote and Berens[22] conceive of expertise as a form of knowledge that is propositionally available to its bearer, such that the steps that one makes to come to a certain diagnosis (a) can be reconstructed as a logical argument, and (b) that this reconstruction adequately represents the expertise exercised to come to a diagnosis. Yet, and this is what we will further clarify in this section, there might be another way of thinking about expertise; one that conceives of it as a hermeneutic process.

An image of a physician as an objective judge who weighs in different concerns in a logical inductive manner and iteratively verifies the conclusions became dominant in medicine in the 19th century.[24] This approach to diagnostics and medical expertise was facilitated by the introduction of medical technologies such as a stethoscope and X-Ray imaging. Medical tools facilitate the diagnostic process by providing a supposedly direct view into the body of the patient through medical imaging and the quantitative representation of bodily concerns. Leder challenged this model of diagnosis and expertise as untenable in view of the value-laden and historically situated nature of both the physician and the patient, as well as the tacit experiential knowledge that also shapes medical expertise and resists quantification.[25] Instead, building on Gadamer,[26] Leder puts forth a model of diagnosis as an inherently hermeneutic enterprise.

The hermeneutic model of the clinical encounter suggests that the doctor iteratively interprets the patient's symptoms and their visualization by instruments against her own background knowledge and experience to arrive at a diagnosis. The text to interpret here appears in the integration of the bodily signals of the patient (the experiential text), their stories combined with doctor's hypotheses (the narrative text), the recorded results of the exams (the physical text) and the instrumental input of graphs and numbers (the instrumental text).[25] The doctor reads these texts as an embodied individual bringing in her

own conceptual and experiential frameworks that incorporate tacit knowledge and the relevant technological input. Medical diagnosis and expertise are thus hermeneutic not by method, but ontologically and epistemologically.[24(p131)] Following Leder, "[in] its attempt to expunge interpretive subjectivity, modern medicine thus threatens to expunge the subject [doctor and patient as the interpreters]. This can lead to an undermining of medicine's [...] hermeneutic telos".[25(p22)]

The hermeneutic model of diagnosis and expertise helps to reframe the nature and role of bias. Gadamer, whose work inspired Leder's hermeneutic approach to medicine, understood bias as a productive pre-judgement and fore-understanding that starts the process of interpretation.[26] Such pre-judgements form an effective history from which any act of interpretation departs because these allow an entry into the mindset of another time, place or object. Gadamer discards the modern negative meaning of bias as prejudice and instead relies on its ancient meaning as prior awareness or pre-judgement.[26(p273)] Also in medicine, bias denotes the cumulative potential of the preconceptions, provisional judgements and prejudices that direct a physician to the patient and their illness, being an inalienable part of her hermeneutic situatedness.

However, acknowledging the productive role of bias for medical diagnosis and expertise does not mean that they are a matter of opinion or preference. As mentioned earlier, medical diagnosis always also presupposes following best practices of consensual validation with colleagues and with an eye to instrumental decision support. Gadamer similarly suggests that interpretation relies on making oneself aware of own biases and how they direct us in viewing new phenomena, even though it is never possible to fully expel them. Viewed as such, bias appears as enabling clinicians to exercise expertise when coming to a medical diagnosis rather than constituting a hindrance to clinical interpretation: "By acknowledging the interpretive nature of clinical understanding, we leave behind the dream of a pure objectivity. Where there is interpretation there is subjectivity, ambiguity, room for disagreement".[25(p10)]

A potential caveat to Gadamer's hermeneutics when applied to the medical diagnosis is that it primarily concerns human bias. Becoming aware of the productive role of bias in decision-making becomes even more difficult when medical diagnosis concerns not just human but also machine bias, for example, in ML algorithms. However, simultaneously Gadamer's hermeneutics points to the impossibility of eradicating bias, because it is an inalienable by-product of both human engineers and designers that developed the AI-assisting decision-support systems, the clinicians eventually using these systems, as well as the ML systems themselves. Indeed, from the perspective of Gadamer's hermeneutics, the very idea of asking algorithms to be completely free of bias places far too high demands on them when compared with human actors. Just as that Zerili et al have argued that demands of algorithms to be fully transparent presuppose an unrealistically high degree on the transparency available on human-decision making,[20] the same can be said about the ability of humans to have full access to their own biases and those of others. Put differently, also human decision-making seems to be, from a hermeneutic point of view, to a large degree "opaque".

A hermeneutic perspective thus points to the need to anticipate and identify the productive role of ML in medical decision-making and act responsibly in light of the non-neutral hermeneutic role of algorithms, instead of focusing on expelling machine bias to ensure the objectivity of the medical diagnosis. This can be done by considering interactions between doctors and algorithms not in the abstract, but as embedded in specific practices. In such practices, once a bias in algorithmic suggestions is noted, doctors can start to identify its relevance within the intricacies of the case and compare it against their experiences and those of their colleagues. As will become clear below, this implies that ML systems should not be treated as offering immediately actionable suggestions before entering specific practices. In the next section, we show how the philosophical approach of postphenomenology can be helpful in this regard to reconceptualize the role of ML algorithms as active mediators in medical encounters.

4 | POSTPHENOMENOLOGY AND MEDICAL DIAGNOSIS

In the previous section, we have argued that a medical diagnosis can be fruitfully understood as a hermeneutic process in which doctors and patients work together towards a medical diagnosis. Having expertise in this process thus both involves a certain fore-understanding of medical diseases and classifications, as well as the capacity to match this knowledge with, and update it in light of the patient's report and instrumental input. In this section, we make clear how ML must be considered as mediating the hermeneutic process through which medical diagnoses are established. To do so, we draw on postphenomenology, an approach within the philosophy of technology concerned with how technologies shape the world to which human beings relate.[27,28]

From a postphenonomenological perspective, when people use technologies, these always mediate human perceptions and actions in view of their design and inherent scripts.[27] However, technologies never fully determine how they are used because the totality of human experiences and prior conceptions, coupled with the specific sociocultural settings, productively inform specific technological mediations. Verbeek calls this phenomenon "the co-shaping of subjects and objects"[28] to designate that not only technological use and its effects are influenced by specific users, but also the agency and subjectivity of those users get shaped in relation to technologies at hand. Viewed through the prism of the technological mediation approach, ML decision-support systems are thus not passive providers of data or neutral diagnostic instruments but actively take part in the diagnostic process, both by providing hermeneutic input and by being a co-interpreter alongside the doctors. ML decision-support systems thus help to shape specific diagnostic pre-judgements and biases, making the medical expertise not solely a human affair but one that is mediated by technologies.

ML-based decision-support systems significantly expand and complicate the hermeneutic model of clinical encounter as put forth by Leder.[25] In Leder's model, the doctor has to reconcile different streams of information about the patient in an iterative way: the ones from initial anamnesis, patient's account and examination, and the others that appear on the screen of the decision-support system, guided by numerical representations of lab results and correlations with evidence-based treatments in similar patient histories. However, as Tschandl et al[29] found in their empirical studies on the interaction of clinicians with ML-based support for skin cancer diagnosis, the line between supporting medical decisions and determining them may be thin if not carefully reflected upon. The statistically ranked and at times colour-coded manner in which ML systems visualize the results and suggest treatments can change the doctor's mind regarding their initial diagnosis.[5,29] Tschandl et al further found that the ML suggestions helped less experienced specialists and general practitioners improve the accuracy of their diagnosis by 26% by changing their initial diagnosis in favour of the one suggested by the ML system when their initial diagnosis was not at least the second or third option suggested by the ML system.[29] More experienced specialists, on the contrary, insisted on their original diagnosis after checking the suggested alternatives and which eventually turned out to be correct.[29(p4)] The experience and confidence of doctors when interpreting and combining various stages of the diagnostic process were determining factors in an accurate diagnosis, whereby the ML suggestions were perceived as alternatives to consider and verify the diagnosis, as a matter of second opinion. Viewed through the technological mediation lens, the doctors acknowledged and scrutinized the productive role of ML in a diagnostic process, making a decision a matter of weighing in both inputs as an intersection between the interpretative horizons of the doctor and the machine.

However, as Tschandl et al also note,[29(p4)] once the doctors gain trust in the ML systems to help them reach a correct diagnosis, the trust may lead them astray when the ML systems become faulty, for example, tainted with biased datasets, applied to an unintended target group or when under malicious attacks. This further challenges the epistemic credibility of ML systems in medicine, as suggested by Grote and Berens,[22] and in parallel strengthens their proposal about introducing diagnostic soundboards in the form of peer panels when ML systems are involved. The case of South Korean doctors as early users of ML-based decision-support systems in cancer treatment suggests that such collaborative diagnostic practices are possible and helpful in reaching a correct diagnosis.

In South Korea, ML systems became involved in accompanying the diagnosis starting from 2016 in several hospitals.[30] To maintain diagnostic transparency and treat the ML system as a recommender and not as a definitive judge, a team of at least five doctors, senior and junior, would correlate the options suggested by the ML system with their own ones to jointly reach a decision.[5] As a positive side-effect, the open manner in which the ML system showed the diagnostic data and the treatment options on a big screen on the wall levels out the decision-making process. It allowed junior doctors to reflect on the data in an open manner, debate the recommendation of the ML system and the hypothesis of their senior colleagues and thus level the hierarchy of the diagnostic process. Such a reflective and collaborative manner of introducing ML-based systems in medicine explicitly addresses both human and machine biases within the iterative diagnostic process: even though it does not offer a way to eliminate machine bias, it can help productively integrating bias into medical practices by creating the opportunity to compare what the algorithm is offering against the expertise of a doctor and her colleagues. The South Korean case was supported by the recent findings of Tschandl et al, demonstrating that "aggregated AI-based multiclass probabilities and crowd wisdom significantly increased the number of correct diagnoses in comparison to individual [doctors] or AI in isolation".[29(p4)]

Viewed through the prism of technological mediation, ML-based decision-support systems do not surround the doctor with a mute wall of numbers and graphs but help to bring the real world in through continuous feedback loops, learning and engagement with the technology and other doctors. As becomes visible in the examples discussed above, it does not seem productive to think of ML systems as potential complete replacements of existing clinical practices, but instead as potential collaborators that function within the *collective* practice of coming to adequate diagnoses and treatment. ML systems can thus be said to mediate what medical expertise is: an integral part of it is being able to *not* consider the treatments and diagnoses offered by ML systems as immediately actionable, but as something to be integrated into collective diagnostic practices. Instead of treating ML systems as black-boxes, medical expertise now also consist of developing the ability to treat them as conversational partners to enter into a dialogue with. This, then, requires to contrast a ML system with one's own biases and treating it as an equally biased dialogical partner. When doing so, medical diagnosis that is accompanied by ML systems becomes an even more nuanced hermeneutic enterprise without blind trust either in the human expertise or in the machine's suggestions. Potentially, this new way of diagnosing becomes less individual and more team-based and where the effectiveness of diagnosis depends on not treating machines as competitors but as collaborators.

5 | DISCUSSION: HOW MACHINE LEARNING RE-DISTRIBUTES EXPERTISE AND CO-DESIGNS DIAGNOSIS

With the aid of the technological mediation approach, we showed how decision-supporting ML systems change the hermeneutic process through which medical diagnoses are made, as well as the role of expertise when coming to a diagnosis. Important to highlight is that this perspective implies that it is not needed to make a choice for either the expertise of clinicians or the alleged objectivity of ML systems; a hermeneutic perspective in technological mediation reveals that clinical expertise and ML systems are co-extensive. This implies that we should recognize that ML systems and clinicians inevitably are dialogical partners during the diagnostic process.

Tschandl et al have recently demonstrated how ML systems can help doctors to identify better a specific type of skin lesions, pigmented actinic keratoses.[29] Backward engineering the algorithmic workings, Tschandl et al found that whereas the ML system focused on the blemish as well as on the area around it, doctors tend to focus only on the blemish itself. Expanding the area of attention allowed ML systems to spot chronic UV damage surrounding the blemish, which causes actinic keratosis. The researchers integrated this finding into training medical resident students, whose accuracy in detecting actinic keratoses consequently increased from 32,5% to 47,3%.[29(p4)] The researchers suggest that learning from the ML systems helps expand the areas of doctors' attention and highlights the value of human collaboration with ML systems.

This example suggests that a focus on human-machine collaboration rather than competition can help to improve the accuracy of medical diagnosis and expand the areas of medical attention. This new form of collaboration should acknowledge the mediating non-neutral import of ML systems. On the one hand, it shows that doctors are not—and never have been—alone in making medical decisions. On the other hand, accounting for the productive role of ML systems in doctor's decision-making dispels the idea of objectivity and de-biasing in the medical practice, rather drawing attention to its inherently hermeneutic nature. From this perspective, any collaboration between clinicians and ML systems presupposes that medical expertise also consists of being able to treat the latter as a conversational partner (just as other team members) that does not offer immediately actionable input, but instead as putting forward its own biases that can be compared against the biases of other team members.

The technological mediation lens helps to expand Leder's hermeneutic model of diagnosis with the active impact of technologies. Highlighting the mediating role of ML systems in the medical diagnosis would help to make what Leder calls "the hermeneutic telos"[25(p22)] of medical decision-making more nuanced. It helps to bring the coherent overview of the patient by preventing her experience from getting lost in the troves of data by increasing opportunities for hermeneutic dialogue with the patient, the colleagues and the machine. ML systems can contribute to the interpretative coherence, collaboration and effectiveness of the diagnosis by confronting the doctor with evidence-based alternative possibilities for diagnosis (which also mitigates physician's biases), and encouraging consultations with other physicians to account for the inaccuracies in the ML systems and the broader social factors that they miss (which additionally mitigates machine's biases).

Doctor's participation in the development and/or tailoring of the ML-based decision-support systems to their specific practice can increase the diagnostic effectiveness. The visual way in which the ML systems communicate the findings may present an undue influence in the doctor's decisions, while not all ML support features are relevant for the practice at hand.[29] As Tschandl et al suggest, the form of machine support should be proportional to the task and the physicians can effectively contribute to the joint development and tailoring of the ML systems in medical practice.[29(pp2,4)] The increased interaction between the doctors and the ML systems essentially transforms medical diagnosis to a form of co-design, whereby all actors co-shape each other.

While in this paper we focused on a diagnostic moment, our research points to a further direction to explore in the future research: how the technologically mediated diagnostic moment in parallel shapes the medical infrastructure, for example, the doctor-nurses relations, the hidden costs of embedding AI technology in the hospital, the hospital organization, etc. Bringing attention to the productive nature of bias in medical diagnosis demonstrates that it is short-sighted to consider the technological factor alone but to see it in the systematic and sociocultural embedding.

Acknowledging the mediating role of ML systems in clinical decision-making essentially points to a triad of diagnostic co-design: an iterative hermeneutic process between doctors, patients and the ML system. The quality of the interaction between the doctor and the ML systems depends on examining the hermeneutic role of the technology, how it simultaneously expands and narrows medical attention, highlights certain aspects, while disclosing others, thus mediating medical perceptions and actions. Including not only the patient, but also other colleagues in the process helps to ensure an encompassing diagnostic process, to respect its inherently hermeneutic nature and to work productively with existing human and machine biases. In this paper, we have primarily focused on two parts of the triad of co-design: doctors and the ML systems. While elsewhere we have discussed in more detail how ML might shape the relation between doctors and patients,[31] a detailed analysis of this is beyond the scope of the current paper. However, let us conclude with a few words on how the understanding of medical expertise in the collaboration between clinicians and ML systems can be used to think about the role of patients in the diagnostic triad. It is argued that ML will reduce the time the doctors spend on making diagnoses and searching for treatments that the doctors can consequently redirect to the interaction with patients.[2,32] Advocates of introducing ML in healthcare in general, and in medical diagnostics in specific, allude to the objectivity of ML as a means to make medical practice "more human". Our analysis, however, suggests that instead of understanding ML as a way to solve such concerns, we should rather ask how it shapes medical expertise and how it shapes the interactions between doctors and patients. After all, the question of whether or not medical practice eventually *will* become more human crucially depends on how ML shapes how patients, the most important stakeholders in medical practice, are made present.

One of the potential pitfalls of ML is that it bears the threat of turning the triad of diagnostic co-design into a dyad: since ML systems rely and make decision on the basis of quantifiable datasets, they implicitly present patients as data, and potentially move the patient's own narrative and experiences to the background.[13] However, as we saw in Leder's account,[25] this information is crucial for how doctors test their hypotheses, and eventually come up with a diagnosis. Therefore, ML places an extra demand on patients to be explicitly vocal about their (medical) biography and personal context that otherwise remain invisible to ML systems. It cannot be expected from every patient that she is capable of doing so, which points to an important concern for doctors

working with ML systems that should be a critical part of medical expertise: the responsibility of ensuring that patients are able to narrate their experiences and context is magnified, as well as the capability to continue integrating these narrations into the diagnostic triad. In other words, it requires active work to keep the diagnostic triad intact and prevent that patient experience disappear from view. From this perspective, keeping medicine "human" consists of maintaining the existence of the diagnostic triad between doctors, ML systems, and patients, rather than eliminating it through an unrealistic pursuit for purified objectivity.

ACKNOWLEDGEMENTS

Olya Kudina's work on this paper has been supported financially by the project Value Change that had received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 788321. Bas de Boer's work on this paper has been supported financially by the project *Pride and Prejudice* that had received funding from 4TU under the High Tech for a Sustainable Future programme.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICAL APPROVAL

The research conducted in the paper did not involve any human and/or animal participants.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Olya Kudina https://orcid.org/0000-0001-5374-1687
Bas de Boer https://orcid.org/0000-0002-2009-2198

REFERENCES

1. Drew T, Vo MLH, Wolfe JM. The invisible gorilla strikes again: sustained inattentional blindness in expert observers. *Psychol Sci*. 2013;24:1848-1853. https://doi.org/10.1177/0956797613479386.
2. Topol E. *Deep Medicine. How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books; 2019.
3. O'Sullivan ED, Schofield SJ. Cognitive bias in clinical medicine. *J R Coll Physicians Edinb*. 2018;48:225-232. https://doi.org/10.4997/JRCPE.2018.306.
4. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. 2020;260:113172. https://doi.org/10.1016/j.socscimed.2020.113172.
5. Ross C, Swetlitz I. IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close" *STAT*. September 5, 2017. https://www.statnews.com/2017/09/05/watson-ibm-cancer/. Accessed August 4, 2020.
6. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med*. 2018;378:981-983. https://doi.org/10.1056/NEJMp1714229.
7. Schönberg D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *Int J Law Inf Technol*. 2019;27:171-203. https://doi.org/10.1093/ijlit/eaz004.
8. Rowley Y, Turpin R, Walton S. The emergence of artificial intelligence and machine learning algorithms in healthcare: recommendations to support governance and regulation [Position paper]. BSI, Association for Advancement of Medical Instrumentation; 2019. https://www.bsigroup.com/globalassets/localfiles/en-gb/about-bsi/nsb/innovation/mhra-ai-paper-2019.pdf. Accessed September 2, 2020.
9. Floridi L, Cowls J, Beltrametti M, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Mind Mach*. 2018;28:689-707.
10. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1:389-399.
11. Whittaker M, Crawford K, Dobbe R, et al. AI Now Report 2018. *AI Now Institute*; 2018. https://ainowinstitute.org/AI_Now_2018_Report.pdf. Accessed September 2, 2020.
12. Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a "Right to explanation". *AI Mag*. 2017;38:50-57. https://doi.org/10.1609/aimag.v38i3.2741.
13. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318:517-518. https://doi.org/10.1001/jama.2017.7797.
14. Gillespie T, Boczkowski PJ, Foot KA. *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: The MIT Press; 2014.
15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56.
16. Torous J, Wisniewski H, Bird B, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *J Technol Behav Sci*. 2019;4:73-85.
17. 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. Paper presented at: Proceedings of the ACM on Human-Computer Interaction; November 2019:104. https://doi.org/10.1145/3359206.
18. Paul HY, Hui FK, Ting DS. Artificial intelligence and radiology: collaboration is key. *J Am Coll Radiol*. 2018;15:781-783.
19. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? *npj Digit Med*. 2020;3:86.
20. Zerili J, Knott A, Maclaurin J, Gavaghan C. Transparency in algorithmic and human decision-making: is there a double standard? *Philos Technol*. 2019;32:661-683.
21. Coeckelbergh M. *AI Ethics*. Cambridge, MA: The MIT Press; 2020.
22. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics*. 2020;46:205-211. https://doi.org/10.1136/medethics-2019-105586.
23. van Baalen S, Carusi A, Sabroe I, Kiely DG. A social-technological epistemology of clinical decision-making as mediated by imaging. *J Eval Clin Pract*. 2017;23:949-958. https://doi.org/10.1111/jep.12637.
24. Svenaeus F. *The Hermeneutics of Medicine and the Phenomenology of Health: Steps Towards a Philosophy of Medical Practice*. Vol 5. Dordrecht: Springer Science & Business Media; 2013.
25. Leder D. Clinical interpretation: the hermeneutics of medicine. *Theor Med*. 1990;11:9-24.
26. Gadamer H-G. *Truth and Method*. New York: Crossroad; 2004 /1975.
27. Ihde D. *Philosophy of Technology: An Introduction*. New York: Paragon House; 1993.
28. Verbeek P-P. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Pennsylvania State University Press; 2005.
29. Tschandl P, Rinner C, Apalla Z, et al. Human–computer collaboration for skin cancer recognition. *Nat Med*. 2020;26:1229-1234. https://doi.org/10.1038/s41591-020-0942-0.

30. Yoon S-W. Korea's third AI-based oncology center to open next month. *The Korea Times*. March 16, 2017. http://www.koreatimes.co.kr/www/tech/2017/03/129_225819.html. Accessed August 4, 2020.
31. de Boer B, Kudina O. What is morally at stake when using algorithms to make medical diagnoses? Expanding the discussion beyond risks and harms. *Theor Med Bioeth*. In press.
32. Chung J, Zink A. Hey Watson - Can I sue you for malpractice? Examining the liability of Artificial Intelligence in medicine. *Asia Pac J Health Law Ethics*. 2018;11:51-80.

How to cite this article: Kudina O, de Boer B. Co-designing diagnosis: Towards a responsible integration of Machine Learning decision-support systems in medical diagnostics. *J Eval Clin Pract*. 2021;27:529–536. https://doi.org/10.1111/jep.13535