Big Data and the Threat to Moral Responsibility in Healthcare

Daniel W. Tigard

1 Introduction

Technological innovations in healthcare, perhaps now more than ever, are posing decisive opportunities for improvements in diagnostics, treatment, and overall quality of life. In particular, the use of big data and artificial intelligence (AI) stands to revolutionize healthcare systems as we once knew them. Indeed, only half of a century or so after the move from *doctors know best* to *patients know best,* we are seeing the potential transformation to *machines know best.* But machines bear important differences from both patients and practitioners; and thus, we are confronted with newfound questions. Among them, in this paper I want to explore: What effect do emerging technologies have on human agency and moral responsibility? How can patients, practitioners, and the general public best respond when responsibility becomes obscured?

The technologies I will have in mind are devices and programs that rely upon 'big data', defined as data "too large and complex to capture, process, and analyze using current computing infrastructure" and characterized in terms of volume, velocity, variety, veracity, and value (Gudivada et al. 2015). From there, we can take note of some of the benefits and challenges of such technologies, particularly as applied in medical contexts. Consider, for example, the *precision health* movement, which promises to harness our personal, genetic, and environmental details in an effort to tailor healthcare to each individual. Under the precision health paradigm, we may well see significant improvements in medical diagnostics (Mega et al. 2014; Jameson und Longo 2015) and more effective treatments, with fewer undesirable side-effects (Mirnezami et al.

D. W. Tigard (✉)
Institut für Geschichte und Ethik der Medizin, Technische Universität München, Ismaninger Straße 22, 81675 München, Germany
E-Mail: daniel.tigard@tum.de

G. Richter et al. (Hrsg.), *Datenreiche Medizin und das Problem der Einwilligung,*
https://doi.org/10.1007/978-3-662-62987-1_2

2012; Chen et al. 2016; Sugeir und Naylor 2018). If operationalized so as to include the identification of biomarkers of diseases and effective preventative measures, we might be better positioned to recommend uniquely healthier lifestyles, and thereby contribute to longer, flourishing lives (Galli 2016; Gambhir et al. 2018). While we might be concerned by medicine becoming dehumanized (Dalton-Brown 2020), some authors promote benefits such as AI serving simply as a "digital scribe", thereby *increasing* patients' interactions with human practitioners (Coiera et al. 2018; Topol 2019).

There are certainly reasons for optimism, but we must also note that technological innovations bring newfound challenges. For instance, the success of AI diagnostic systems, along with precision health and other advances, depends upon harnessing massive quantities of highly sensitive data. Patients' medical records and seemingly benign lifestyle information must be stored and shared, raising justifiable concerns over privacy and data security (Mizani und Baykal 2015; Berger und Schneck 2019). With increasing reliance upon machines, physicians risk compromising patients' trust and public perceptions of the healthcare system as a whole (Sparrow and Hatherly 2020). Even if properly harnessed, data-driven systems face difficulties in helping us in ways that appear transparent, fair, or compassionate, considering the concerns for "black box" decision-making, biases in the data, and AI's potential to erode the patient-physician relationship (DeCamp und Tilburt 2019; Ploug und Holm 2020). Lastly, for now, we must work to assure that emerging medical technologies are widely available and not used to exacerbate inequalities or discrimination (Stiles und Appelbaum 2019).

In considering both the prospects and potential problems, we might wonder, can we harness the benefits of digital medical technologies while managing the challenges? Here some will respond negatively, saying the risks are too great; we would be better off with traditional models of treatment and clinical research.[1] I suspect, however, that most readers will be at least somewhat optimistic. The question for them is, then: *How* exactly can we assure that the benefits outweigh the costs? Here is where we see a wide range of efforts, with researchers focused on increasing privacy, security, or transparency in data-based models; on decreasing harmful biases, discrimination, and inequalities; on revising models of informed consent, so as to make the use of patient data morally acceptable, and so on. And while these efforts may well help to usher in the new era of digitalized medicine, my assumption throughout this essay will be that harms will still occur. Errors will be made, both by human practitioners and by the technologies they employ.[2] The

[1] Although few have made an outright case against the development and use of innovations like medical AI. This is rather surprising, considering that we see campaigns against the use of killer robots (Sparrow 2007), sex robots (Richardson 2016), and self-driving cars (Madrigal 2018). Still, some authors appear committed to raising worries over medical AI (e.g. McDougall 2019; Morley et al. 2019; Sparrow and Hatherly 2020).

[2] For example, in a report on deep learning in detecting metastatic breast cancer, Wang et al. (2016) found their system's success rate to be 92.5 %, while human pathologists identified 96.6 % of the images correctly. Deep learning and human diagnosis combined showed a 99.5 % success rate. The reduction in human error – 85 %, as the authors report – should undoubtedly be celebrated. My focus is simply on who is to blame when things go wrong, even if such cases are minimized.

ethical questions that arise, then, are: How should we deal with the inevitable failures? Who should be held morally responsible when it appears that a machine has cost us a life? And who, or what, *can* be plausibly held responsible?

As I will argue, our ability to locate responsibility may become threatened with the use of data-driven medical technologies, in which case we are left with a difficult choice of trade-offs. It might seem, on the one hand, that we must exercise extreme caution or perhaps restraint in our use of state-of-the-art systems. Yet, the prospect of losing out on some of the exciting benefits is unlikely to be widely appealing. Thus, on the other hand, we can proceed with innovative digital healthcare models, but in doing so we take on a degree of risk and might need to loosen our commitment to locating moral responsibility. In either case, the use of AI and big data calls for new ways of thinking about responsibility in healthcare. To show this, in Sect. 2, I briefly summarize notions of moral agency and responsibility as they have been developed in the philosophical literature. In Sect. 3, I clarify why responsibility is so important in high-stakes domains, such as healthcare, and explain how the use of AI and big data might challenge our basic conceptions. Then, in Sect. 4, I outline numerous proposals for how we can respond when moral responsibility becomes obscured. In Sect. 5, I close by suggesting that all of us, as members of the moral community, can help to adapt our existing mechanisms in ways that accommodate emerging healthcare systems and technological environments.

2 Agency and Responsibility Preliminaries

In order to make the case that responsibility is potentially threatened by the use of emerging medical technologies, I must first make clear what is meant when speaking of responsibility. To begin, it should be stated that in the philosophical literature, "responsibility" is used widely – if not unanimously – as a neutral term. Granted, the word is often used in science and technology studies to entail something *positive* or *morally good* (e.g. 'responsible research and innovation'). However, as its conceptual roots, being responsible is neither good nor bad. To see this, consider that humans can be *responsible* for doing terrible things: lying, murdering, and so on. Consider also that a person who throws a life-vest to a drowning child can be thought of as *responsible* for saving the child's life. Thus, while it might clash with our everyday usage, saying simply "X is responsible" does not yet give us a positive or negative evaluation of X. We need to know what X is responsible *for* and why (cf. Loh 2019; Tigard 2020a).

I should also point out that I am concerned primarily with notions of *moral* responsibility, as opposed to legal notions, like liability. However, concepts of morality can still serve to inform and substantiate legal concepts.[3] Accordingly, if we see that

[3]This is not to say that moral and legal responsibility have similar meanings or functions. See Shoemaker (2013) for arguments dispelling these common connections.

moral notions (like blame) are unclear with the emergence of novel technologies, we may have reason to think legal notions (like liability) will be difficult to apply in new sorts of cases. In this way, the pursuit of clarifying moral responsibility in cases of emerging technology can be appreciated across various regulatory domains.

So, what exactly is moral responsibility? When can we appropriately say someone is morally responsible? Typically, in order to think of a person as morally responsible, we must think of that person a moral agent. She must be capable of initiating action, knowingly and freely – these features are often referred to, respectively, as the 'epistemic' and 'control' conditions. If one doesn't know what she is doing, or doesn't understand something pertinent about the situation, she is usually *not* thought to be fully responsible (think of how we treat children). If one is somehow not in control of her actions – whether due to external factors (like coercion) or internal characteristics (like spasms) – she is similarly not fully responsible.[4] Morally responsible agents, then, are persons who act knowingly and freely in ways that can be evaluated, either by others or by themselves, as morally good or bad. Common instances of these evaluations are praise and blame: for example, someone who knowingly and freely saves a life is appropriately praised, while someone who knowingly and freely murders is appropriately blamed.

There are, of course, many complexities lurking in the ideas presented thus far, most of which cannot be adequately addressed here. But I want to draw attention to one important development, since it will resurface in later sections. Recent responsibility theorists have argued that moral responsibility is not a singular phenomenon (Watson 2004; Shoemaker 2011). The basic idea is hopefully rather intuitive, namely, that there are several ways in which a person can be thought of as morally responsible. Gary Watson is commonly seen as one of the pioneers of *responsibility pluralism.* In his 1996 essay "Two Faces of Responsibility", Watson showed that we can think of a person's action as representing their underlying values or character, what it is they stand for. In this sense, we are *attributing* the action to that person, judging that it represents who they are. But unlike responsibility-as-attributability, Watson suggested, we could also hold others *to account,* namely by actively communicating, expressing expectations (whether met or failed), and rewarding or punishing. Responsibility-as-accountability, in this way, is distinct from – and not always coincident with – attributability. Drawing the divisions further, David Shoemaker showed that responsibility comes apart into three sorts. Aside from attributing some action and outwardly blaming or praising the person for it, we can demand reasons.[5] Considering our efforts to understand *why* someone behaved as

[4] Some will notice that I've said one isn't "*fully* responsible." With this usage I mean to recognize that, due to the great diversity of the human condition, concepts like agency and responsibility are not all-or-nothing features. See Shoemaker (2015: 120–122).

[5] It must be noted that while Shoemaker expands upon Watson's two-fold division, Shoemaker's account of attributability and accountability differs from Watson's account.

they did, we see a sort of responsibility-as-answerability, wherein we evaluate not one's character, but their judgment or decision-making processes (Shoemaker 2011, 2015).

Again, this depiction of moral responsibility glosses over a host of complexities and objections. One might think, for example, that responsibility doesn't *really* mean two or three separate things – it is still a singular, albeit complex, phenomenon (McKenna 2018). For present purposes, I can remain neutral on this issue and accept that responsibility might be a plural enterprise, or it might be singular, in which case I find it safe to assume nonetheless that there are multiple ways we think of others as responsible. That is, attributing an action to one's character, holding her to account for it, and demanding answers are all mechanisms by which we hold moral agents (those who act knowingly and freely) as morally responsible. With this in mind, I turn to an examination of responsibility's importance in high-stakes domains, like healthcare, and of how emerging technologies present newfound social and ethical challenges.

3 Responsibility Gaps and Data-Driven Technology

In recent debates in technology ethics, concerns have been raised that machines are becoming sophisticated enough to behave in ways that go beyond our awareness or control. The notion of a 'responsibility gap' was first put forward by Andreas Matthias (2004: 175), who claimed that autonomous, learning machines "create a new situation, where the manufacturer/operator of the machine is *in principle* not capable of predicting the future machine behavior." As a result, Matthias argues, the manufacturer or operator cannot be held responsible. In a fuller sense, responsibility gaps are situations involving two basic features: (a) it seems fitting to hold someone responsible, but (b) there is no one who it is fitting to hold responsible.[6]

What is important to notice is the appeal to some of the most traditional and intuitive criteria for moral responsibility. A machine's manufacturer and users will be unable to predict its behavior; that is, they do not always *know* exactly what will happen when such devices are put to use. By contrast, those who are more optimistic could argue that the use of big data can help us to *better* understand – say, a patient's condition or treatment options – and thereby stands to *enhance* agential capacities (Fakoor et al. 2013). Yet, it is commonly acknowledged that algorithmic processes lack transparency, and that systems based, for example, on artificial neural networks arrive at their outputs by way of *hidden* layers of coding. Although recent works aim to address the problems of transparency (Adadi und Berrada 2018; Wachter et al. 2018; Arrieta et al. 2020), it seems that manufacturers and users will often remain incapable of knowing exactly why some

[6]This helpful formulation is adapted from Köhler, Roughley and Sauer (2017). On their account, however, responsibility gaps are primarily a matter of accountability, which I find too narrow to account for the pluralism established above. I discuss this further in Tigard (2020b).

output is given. If indeed knowledge is a necessary component of moral responsibility, as traditional theories suggest, it appears that the loss of knowledge brings about the loss of responsibility. After all, it seems unfair to blame someone who simply did not know.

But surely, some might think, the users are still in *control* of what happens with their devices, namely the decision to deploy the technology in the first place, and this alone should be enough to locate those who are morally responsible for any harms. However, among the key concerns articulated by Matthias and others is the fact that the manufacturers and operators have given up control; their machines now act autonomously in the sense of determining their own behavior (Matthias 2004; Hellström 2013). Indeed, one of the primary motivations to deploy automated technologies is that humans no longer need or want to retain control. The labor, including decision-making processes, can be outsourced (Vallor 2015; Danaher 2016a; Smids et al. 2019). The problem, according to numerous critics, is that there are domains in which we should not be outsourcing importantly human work.[7]

One of the most prominent critics of technology – specifically, its use in highly sensitive domains like warfare and healthcare – is Robert Sparrow. In a widely cited article, Sparrow (2007) argues that in our use of sophisticated technologies, there are three possible loci of responsibility, none of which turn out to be responsible for potential harms. First, like Matthias, Sparrow argues that the programmers of autonomous machines will be unable to predict or control its behavior. If we assume that all necessary precautions were taken – for example, autonomous weapons being programmed so as to fire only upon enemy combatants – it will be unfair to hold them responsible in the event of any unintended consequences, such as robots misfiring upon civilians.[8] Similarly, and perhaps more controversially, the users of autonomous devices – such as a commanding officer – cannot be appropriately held responsible. For, on Sparrow's account, the user likewise was not in control and could not foresee the unfortunate consequences. Finally, the technology itself, for Sparrow (among others), cannot be plausibly held responsible since machines are clearly incapable of suffering punishment – a position I take issue with, below, by suggesting we move away from purely retrospective approaches to responsibility.

In this way, if we give up control to automated technologies, we risk facing situations where a harm has occurred and (a) someone should be held responsible, but (b) there is no one who can be fittingly held responsible. And while we might not be able to articulate exactly why it bothers us, the prospect of facing ambiguous harms may be quite unsettling (Danaher 2016b). The question, then, is: Do digital medical technologies

[7] Common examples of such domains are warfare (Sharkey 2010; Asaro 2012), healthcare (Char et al. 2018), and care for the elderly (Sparrow 2016).

[8] Here, the legal-minded reader might have notions of negligence in mind. Indeed, for Sparrow (2007: 69) responsibility for programmers "will only be fair if the situation described occurred as a result of negligence."

create such responsibility gaps? To what extent are data-driven research and clinical treatment creating potentially ambiguous harms?

In a recent editorial on precision medicine, medical futurist Bertalan Mesko (2017: 239) explains that the increasing availability of genome sequencing, wearable health sensors, and hand-held devices are producing vast quantities of data, such that "it has become impossible for a physician to analyze all those data or simply to be up-to-date." As a result, we see the growing need for algorithms that can learn independently, integrating patient information and masses of medical literature, and detect diseases from images or samples (Topol 2019). But because these algorithms are too complex to be understood by human practitioners, or even by the designers and programmers, it seems that we are fast approaching a model of healthcare wherein machines "know" more about a patient's condition, and possibly more about how to treat it, than both the patient and practitioner. A similar point was recently made by Thomas Grote and Philipp Berens (2020: 1) who suggest that the use of "current machine learning algorithms challenges the epistemic authority of clinicians." Their argument begins by accepting that there is great promise in machines' capacities to analyze electronic health data. Once fully operational, the processing of big data may well allow us to more effectively diagnose, treat, and prevent diseases on an individual level. However, there will be cases of disagreement – say, between the diagnosis provided by an algorithm and that of human practitioners.[9] How, in such cases, should patients and healthcare providers proceed?

On the account of Grote and Berens, there is "little that the clinician might do on epistemic grounds to resolve the disagreement" (2019: 3). In cases of human-to-human peer disagreement, we have opportunities to engage in deliberation, to see others' evidence and reasoning, and perhaps arrive together at a more accurate understanding. But when one disagrees with the findings of an algorithm, these sorts of procedures are simply not possible. This is due largely to the fact that algorithmic processes are, by their nature, opaque to human comprehension. Here we see why many researchers are concerned with revealing the "black box" nature of deep-learning systems, and why we might want to work toward developing AI systems that are *explainable* (Wachter et al. 2018; Arrieta et al. 2019). Perhaps in time, we will be better able to deliberate and coordinate with AI in healthcare contexts. However, as some argue, the black box systems being introduced are not conducive to informed decision-making or shared deliberation, and thereby risk subverting patient values.[10]

Data-driven technologies may well bring about notable net benefits in medical and health-economic outcomes (Chen et al. 2016). Yet, when harms occur nonetheless – even if in fewer cases – we will want to learn why, whether in order to assure that similar

[9] The potential for disagreement can be seen by, again, referring to the differing diagnostic success rates between humans and machine learning systems – see note 2. Although, there, the differing rates showed higher success with human diagnosis, surely there will be individual cases where machines are correct while humans fail.

[10] Along with Grote and Berens (2020), see McDougall (2019) and Bjerring and Busch (2020).

harms do not reoccur, or simply to satisfy our psychological need to hold someone responsible. And while the opacity of machine learning systems appears to preclude undertaking such actions, notice that the notion of responsibility at work here can be characterized as 'answerability' as outlined above. When we evaluate decision-making processes and seek explanations, reasons or justifications, we are attempting to hold others answerable for their conduct. Unsurprisingly, this becomes increasingly difficult the more we rely upon *unexplainable* systems. For *some* sorts of responsibility, then, it may be that there is a potential gap created by data-driven medical technologies. But for other sorts, we might still find and create social and institutional mechanisms for addressing the threat. I turn next to outlining such proposals.

4 Rethinking Responsibility for Data-Driven Healthcare

Recall that the key threats to moral responsibility or "gaps" are defined as situations where (a) it seems fitting to hold someone responsible, but (b) there is no one who it is fitting to hold responsible. As suggested above, there is a sense in which data-driven medical technologies pose such challenges, to both our current moral and legal frameworks.[11] And although legal gaps are a legitimate cause for concern, it seems that local and international legal systems can continue to make progress in regulating undesirable effects upon data-subjects (cf. Wachter und Mittelstadt 2019; McMahon et al. 2020). For this reason, my primary concern has been the effects of emerging technologies upon *moral* responsibility.

Indeed, it seems quite unclear how to resolve the ethical challenges and, accordingly, that we face a difficult choice of trade-offs. On the one hand, as critics will claim, we might need to exercise extreme caution or restraint in our use of data-driven systems. The risks are simply too great, both in terms of physical harm to patients but also considering the additional psychological harms to patients, families, and communities left in the dark as to who is responsible. Consider, for example, marginalized populations facing additional discrimination as a result of racially biased healthcare algorithms (Obermeyer et al. 2019). The fact that responsibility is not entirely clear certainly counts against the deployment of such systems. On the other hand, in order to harness the net benefits of emerging medical technologies, we could continue to work toward mitigating the harms and make explicit any trade-offs, namely the need to loosen or at least rethink our search for responsibility when patients come to harm. In this final section, I offer strategies that lean toward the latter approach, addressing the question of how we – as patients, practitioners, and concerned citizens – can respond in the face of potentially obscured responsibility.

[11] As raised by Matthias (2004: 176), the challenge concerns both "the moral framework of society and the foundation of the liability concept in law." For Köhler, Roughley und Sauer (2017), the key concern is existing regulatory gaps, which explains their narrow focus on accountability.

Against the backdrop of the responsibility gap, the first mechanism is a response to the feature (a) that it seems fitting to hold someone responsible. In cases where this condition is not present, it would be far less concerning that (b) there is no one who can be appropriately held responsible. As John Danaher (2016b) suggests, it is understandable that individuals negatively impacted by technological systems are inclined to impose blame. Yet, it is not clear that they should be so inclined.[12] Surely, there are cases where harm is utterly ambiguous – perhaps the source is unidentifiable or simply does not exist. Picture, for example, those who place blame upon fate or the gods when they experience a particularly unlucky situation. Of course, in no way do I want to discount our natural psychological responses to distressing situations; for, it seems that such responses can be extremely valuable, in terms of revealing and affirming the things we care about (Tigard 2019a).

Nonetheless, in controlled environments, such as healthcare settings, it seems that we can work to manage the expectations that would give rise to distressful responses. For example, we may find it entirely natural for patients to seek responsible individuals in the aftermath of a harmful error. Yet, we can work to ensure that patients and the wider public become increasingly aware of the basic functionality of emerging technologies. In particular, efforts can be made to carefully convey the importance of data in healthcare diagnostics and treatment systems, and that patients can support the success of these emerging technological models (e.g. McGonigle 2016; Wiggins und Wilbanks 2019). Given the motivation to retain patient autonomy, it is clear also that we should undertake such measures as developing new modes of informed consent (e.g. Hansson et al. 2006; Steinsbekk 2013; Ploug und Holm 2016). What will be more difficult to convey, however, will be the nature of "black box" AI systems and applications of machine-learning to healthcare. But by working to raise awareness of the benefits and challenges of data-driven healthcare, we might help to promote informed participation, improvements in research and treatment, as well as revised expectations on how responsibility can *or cannot* be located in single individuals. If it seems less fitting to hold someone responsible, the potential for facing a responsibility gap can be mitigated.

Next, along with revising expectations – namely those that give rise to (a) – we can work to assure that *there is* someone to be held responsible in the event of unfortunate outcomes, however unforeseen they may be. This approach may seem rather intuitive, but I want to make clear what is at stake for our understanding of moral responsibility. On many traditional and contemporary theories, the most appropriate individuals to hold responsible are those who are indeed seen as responsible. As suggested above, this means identifying the person or group with sufficient knowledge and control of the conduct in question. Even on the pluralistic views of responsibility, those who are most appropriately held accountable, answerable, or attributed with the harm are those who

[12] For conveying this idea in conversation, I'm indebted to Peter Königs.

somehow deserve it. But as we saw, given the technological threat to responsibility, it may be that no one knew enough or had sufficient control of the conduct to the extent that they deserve to be held responsible in any way. This, however, does not mean that there is no one who can *take* responsibility. We simply need to shift our understanding of responsibility in order to accommodate data-driven technologies. Allow me to briefly expand upon this second strategy.

Unlike the characteristic cases of an agent causing harm and subsequently being held responsible, either by herself or others, we can imagine cases where someone *takes* responsibility – that is, for an unfortunate situation or outcome that she did not directly cause. With these latter sorts of cases in mind, Elinor Mason (2019) argues that often we *should* take responsibility for harms to which we are somehow connected. Mason appeals to Bernard Williams's famous case of the lorry driver who "through no fault of his, runs over a child" (1981). Here Williams makes clear that there is a unique response we are expected to have when we play a causal role in unfortunate events, and that others would think poorly of those who do not show some sort of regret or remorse.[13] For Mason, such situations are an opportunity to take 'ownership' of an action by displaying to those who are affected that we are invested in our relationships, and committed to securing others' trust, even where we were not *morally* at fault. While Mason is concerned to show the importance of taking responsibility within interpersonal relations, as I see it, the very same mechanism extends to professional relations (Tigard 2019b). Particularly with the use of emerging technology in healthcare, where the trust between patients and practitioners is coveted but vulnerable, we see good reason to expect someone to take ownership and help patients through any harms they may incur. It might be an attending physician, hospital counselor, or possibly a third-party[14] – by taking responsibility, we can minimize the negative impact of facing ambiguous harms. Where someone effectively takes responsibility, in this way, the threat of a gap is again decreased.

For the final strategy, I maintain my focus on feature (b), where there is supposedly no one who can be held responsible. Here I admit that my proposal requires a more radical rethinking of how we assign moral responsibility; for the suggestion is that we might work to develop ways of holding technology itself responsible. To be sure, this idea is easily dismissible, both by technology critics like Matthias and Sparrow and by those who are more optimistic about our lives with technology (e.g. Danaher 2019;

[13] Williams dubbed this response "agent-regret", though the precise label is not of particular importance here.

[14] McMahon, Buyx, und Prainsack (2020) introduce the notion of *harm mitigation bodies*, which would serve precisely these purposes. HMBs are a notable step forward in protecting data-subjects from ambiguous harms; still, one might worry that the 'bodies' are not sufficiently connected to the subject or to the harm.

Nyholm 2020). Thinking back to the pluralistic account offered above, what would it even mean to hold technology responsible?[15]

First, we might sensibly evaluate a machine's behavior, but it would appear difficult to demand answers and perhaps impossible, given the opaque nature of deep-learning systems. Indeed, as I suggested, answerability will be particularly threatened by the increasing use of data-driven healthcare. Still, in our efforts to increase transparency and to develop explainable AI, we increase the prospect of grasping reasons for its outputs. Second, attributing an action to a machine's underlying values or character may initially appear absurd. After all, machines cannot be said to have characters in the way we do; they cannot *care* or be committed to someone or something. Nonetheless, in a growing volume of work we see the idea that, starting from their design, technological devices can *and do* reflect values (e.g. Winner 1980; Friedman 1997). Granted, the values *reflected* are likely only those of the human designers or users. Thus, by attributing some characteristic behavior to technology itself, we plausibly hold the designers or users responsible in this way. But even here, practitioners and healthcare institutions can be encouraged to carefully evaluate the "character" of technologies – and of companies that design them – and to consider how well those values cohere with their own and with their patients' values.[16] Third, accountability for technology itself will often be quickly dismissed, for the reason that holding someone accountable entails punishment or demanding reparations, and these measures cannot be applied to machines. Again, in a traditional sense, only humans are truly held to account in these retrospective ways. But some devices and programs can be designed to learn from the positive and negative reinforcement implicit in praise and blame (cf. Hellström 2013). If we move away from accountability only as retribution, and toward mechanisms of rewarding or punishing for the sake of encouraging or discouraging future behavior, we see plausible ways in which machine-learning systems can be held accountable. Thus, to the extent that our conceptions of responsibility are adaptable to the use of emerging technologies, it seems that there is no responsibility gap in healthcare (cf. Tigard 2020b).

In sum, the underlying motivation for the strategies outlined here has been to encourage new ways of thinking about responsibility, and thereby to consider various approaches for mitigating the potential threat to responsibility in data-driven healthcare. Emerging technologies will undoubtedly pose challenges for our common understandings of moral responsibility. It is because of this we see reasons to shift

[15] In Tigard (2020c), I offer a general framework for approaching this question. In short, we must shift away from a 'property view' of responsibility, where we seek (and fail to find) qualities like consciousness and empathy in machines, and toward a 'process view' wherein responsibility is a matter of our relationships and interactions. Similarly, see Coeckelbergh's relational account (2009, 2010).

[16] For example, it may be that an AI oncology treatment system fails to consider a patient's preference for purely palliative care, in which case the clinic should question its use of the system with this patient. See McDougall (2019).

expectations and assure informed patient participation, to encourage relief through practitioners taking responsibility, and perhaps to locate notions of responsibility in technology itself.

5 Conclusion

I began with the caveat that moral responsibility is distinct from legal notions, such as liability. Where concepts of the latter sort are lacking, or where their meaning becomes threatened by newfound technologies, we see opportunities for legal experts to come together and assess how the regulatory gaps can and should be filled. No doubt, the emergence of AI and big data in healthcare is presenting a wealth of such opportunities. But what of the moral domain? How exactly can we work to assure that the potential threats to moral responsibility are likewise mitigated? Here, it seems, we would do well to bear in mind that the opportunities are not reserved for experts alone. Considering that we are all members of the moral community, it is up to all of us to see that the ways in which we interact and relate to one another can accommodate our emerging technologies and environments. The ways in which we hold one another responsible – and how we hold ourselves responsible – can undergo adaptations to better fit our increasing use of devices to which we cede some degree of control and which may "know" more about our health than human practitioners. As I have suggested, we can work to assure informed participation in data-driven systems, and shift expectations of patients and of the public, so that when harms occur nonetheless we may be less inclined to blame where there is no one responsible. We can and should see that someone who is sufficiently connected to the harm is prepared to take responsibility, that is, even where they were not strictly at fault. We may also need to start taking more seriously the idea that, in some ways, technology itself might be a plausible loci of responsibility. If we are willing to rethink moral responsibility, it seems we will be much better positioned to harness the benefits and manage the challenges of data-driven healthcare, and to ease the distress of the occasional ambiguous harm.

References

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on Explainable Artificial Intelligence (XAI). IEEE Access 6:52138–52160

Arrieta AB, Díaz N, Del Ser J, Bennetot A, Tabik S, Barbado A, Chatila R (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Asaro P (2012) On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. Int Rev Red Cross 94:687–709

Berger KM, Schneck PA (2019) National and transnational security implications of asymmetric access to and use of biological data. Front Bioeng Biotechnol 7(21).

Bjerring JC, Busch J (2020) Artificial intelligence and patient-centered decision-making. Philos Technol 1–23
Char DS, Shah NH, Magnus D (2018) Implementing machine learning in healthcare – addressing ethical challenges. N Engl J Med 378:981–983
Chen Y, Guzauskas GF, Gu C et al (2016) Precision health economics and outcomes research to support precision medicine: big data meets patient heterogeneity on the road to value. J Pers Med 6(4):20
Coeckelbergh M (2009) Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents. AI & Soc 24:181–189
Coeckelbergh M (2010) Robot rights? Towards a social-relational justification of moral consideration. Ethics Inf Technol 12(3):209–221
Coiera E, Kocaballi B, Halamka J, Laranjo L (2018) The digital scribe. NPJ Digit Med 1(1):1–5
Dalton-Brown S (2020) The ethics of medical ai and the physician-patient relationship. Camb Q Healthc Ethics 29(1):115–121
Danaher J (2016a) The threat of algocracy: Reality, resistance and accommodation. Philos Technol 29(3):245–268
Danaher J (2016b) Robots, law and the retribution gap. Ethics Inf Technol 18(4):299–309
Danaher J (2019) Automation and Utopia: human flourishing in a world without work. Harvard University Press, Cambridge
DeCamp M, Tilburt JC (2019) Why we cannot trust artificial intelligence in medicine. Lancet Digital Health 1(8):e390
Fakoor R, Ladhak F, Nazi A, Huber M (2013) Using deep learning to enhance cancer diagnosis and classification. In: Proceedings of the international conference on machine learning (Vol. 28). New York, USA: ACM.
Friedman B (1997) Human values and the design of computer technology. Cambridge University Press, Cambridge
Galli SJ (2016) Toward precision medicine and health: opportunities and challenges in allergic diseases. J Allergy Clin Immunol 137(5):1289–1300
Gambhir SS, Ge TJ, Vermesh O, Spitler R (2018) Toward achieving precision health. Sci Trans Med 10(430):eaao3612
Grote T, Berens P (2020) On the ethics of algorithmic decision-making in healthcare. J Med Ethics 46(3):205–211
Gudivada VN, Baeza R, Raghavan VV (2015) Big data: promises and problems. Computer 3:20–23
Hansson MG, Dillner J, Bartram CR, Carlson JA, Helgesson G (2006) Should donors be allowed to give broad consent to future biobank research? Lancet Oncol 7(3):266–269
Hellström T (2013) On the moral responsibility of military robots. Ethics Inf Technol 15(2):99–107
Jameson JL, Longo DL (2015) Precision medicine – personalized, problematic, and promising. New Engl J Med 372(23):2229–2234
Köhler S, Roughley N, Sauer H (2017) Technologically blurred accountability? In: Ulbert C et al (ed) Moral agency and the politics of responsibility. Routledge, London
Loh J (2019) Responsibility and robot ethics: a critical overview. Philosophies 4(4):58
Madrigal A (2018) 7 Arguments against the Autonomous-Vehicle Utopia. The Atlantic, 20
Mason E (2019) Between strict liability and blameworthy quality of will: taking responsibility. In: Shoemaker D (ed) Oxford studies in agency and responsibility, Vol 6. Oxford University Press, Oxford, pp 241–264

Matthias A (2004) The responsibility gap: ascribing responsibility for actions of learning automata. Ethics Inf Technol 6(3):175–183
McDougall RJ (2019) Computer knows best? The need for value-flexibility in medical AI. J Med Ethics 45(3):156–160
McKenna M (2018) Shoemaker's responsibility pluralism: reflections on Responsibility from the Margins. Philos Stud 175(4):981–988
McMahon A, Buyx A, Prainsack B (2020) Big data governance needs more collective responsibility: the role of harm mitigation in the governance of data use in medicine and beyond. Med Law Rev 28(1):155–182
Mega JL, Sabatine MS, Antman EM (2014) Population and personalized medicine in the modern era. J Am Med Assoc 312(19):1969–1970
McGonigle IV (2016) The collective nature of personalized medicine. Genet Res 98
Mesko B (2017) The role of artificial intelligence in precision medicine. Exp Rev Precis Med Drug Dev 2(5):239–241
Mirnezami R, Nicholson J, Darzi A (2012) Preparing for precision medicine. New Engl J Med 366(6):489–491
Mizani MA, Baykal N (2015) Policymaking to preserve privacy in disclosure of public health data: a suggested framework. J Med Ethics 41(3):263–267
Morley J, Machado C, Burr C, Cowls J, Taddeo M, Floridi L (2019) The debate on the ethics of AI in health care: a reconstruction and critical review. SSRN *3486518*.
Nyholm S (2020) Humans and robots: ethics, agency, and anthropomorphism. Rowman & Littlefield, London
Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366(6464):447–453
Ploug T, Holm S (2016) Meta consent – a flexible solution to the problem of secondary use of health data. Bioethics 30(9):721–732
Ploug T, Holm S (2020) The right to refuse diagnostics and treatment planning by artificial intelligence. Med Health Care Philos 23(1):107–114
Richardson K (2016) The asymmetrical 'relationship': parallels between prostitution and the development of sex robots. ACM Digital Library
Sharkey N (2010) Saying "no!" to lethal autonomous targeting. J Mil Ethics 9(4):369–383
Shoemaker D (2011) Attributability, answerability, and accountability: toward a wider theory of moral responsibility. Ethics 121(3):602–632
Shoemaker D (2013) Blame and punishment. In: Coates J, Tognazzini N (eds.) Blame: Its Nature and Norms. Oxford University Press, pp 100–118
Shoemaker D (2015) Responsibility from the margins. Oxford University Press, Oxford
Smids J, Nyholm S, Berkers H (2019) Robots in the workplace: a threat to—or opportunity for—meaningful work? Philos Technol 1–20
Sparrow R (2007) Killer robots. J Appl Philos 24(1):62–77
Sparrow R (2016) Robots in aged care: a dystopian future? AI & Soc 31(4):445–454
Sparrow R, Hatherley J (2020) High hopes for "Deep Medicine"? AI, economics, and the future of care. Hastings Cent Rep 50(1):14–17
Steinsbekk KS, Kare MB, Solberg B (2013) Broad Consent Versus Dynamic Consent in Biobank Research: Is Passive Participation an Ethical Problem? Eur J Hum Genet 21(9):897–902
Stiles D, Appelbaum PS (2019) Cases in precision medicine: concerns about privacy and discrimination after genomic sequencing. Ann Intern Med 170(10):717–721
Sugeir S, Naylor S (2018) Critical care and personalized or precision medicine: who needs whom? J Crit Care 43:401–405
Tigard D (2019a) The positive value of moral distress. Bioethics 33(5):601–608

Tigard D (2019b) Taking the blame: appropriate responses to medical error. J Med Ethics 45(2):101–105
Tigard D (2020a) Responsible AI and moral responsibility: a common appreciation. AI and Ethics, forthcoming.
Tigard D (2020b) There is no techno-responsibility gap. Philos Technol, forthcoming
Tigard D (2020c) Artificial moral responsibility: how we can and cannot hold machines responsible. Camb. Q. Healthc Ethics, forthcoming
Topol E (2019) Deep medicine: how artificial intelligence can make healthcare human again. Hachette, UK
Vallor S (2015) Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character. Philos Technol 28(1):107–124
Wachter S, Mittelstadt B, Russell C (2018) Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv. JL & Tech 31:841
Wachter S, Mittelstadt B (2019) A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Bus Law Rev 2019(2):494–620
Wang D, Khosla A, Gargeya R, Irshad H, Beck AH (2016) Deep learning for identifying metastatic breast cancer. arXiv:1606.05718.
Watson G (2004) Agency and answerability. Oxford University Press, Oxford
Wiggins A, Wilbanks J (2019) The rise of citizen science in health and biomedical research. Am J Bioeth 19(8):3–14
Williams B (1981) Moral Luck. Cambridge University Press, Cambridge
Winner L (1980) Do artifacts have politics? Daedalus 109(1):121–136