Institutional Trust in Medicine in the Age of Artificial Intelligence

Michał Klincewicz

1 Introduction

It is easier to talk frankly to a person whom one trusts. It is also easier to agree with a scientist whom one trusts. Even though in both cases the psychological state that underlies the behavior is called 'trust', it is controversial whether it is a token of the same psychological type. Trust can serve an affective, epistemic, or other social function, and comes to interact with other psychological states in a variety of ways. The way that the functional role of trust changes across contexts and objects is further complicated when communities and individuals mediate it through technologies, and even more so when that mediation involves artificial intelligence (AI) and machine learning (ML).

In this chapter I look at the ways in which trust in institutions, and specifically the medical profession, is affected by the use of AI and ML. There are two key elements of this analysis. The first is a disanalogy between institutional trust in medicine and institutional trust in science (Irzik and Kurtulmus 2021, 2019; Kitcher 2001). I note that as AI and ML become a more prominent part of medicine, trust in a medical institution becomes more like trust in a scientific institution. This is problematic for institutional trust in medicine and the practice of medicine, since institutional trust in science has been undermined by, among other things, the spread of misinformation online and the replication crisis (Romero 2019).

There is also a strong analogy between the psychological state of the person who trusts a scientific report or testimony and the psychological state of a patient who trusts individual recommendations made by a medical professional in a clinical setting. In both cases, institutional trust makes it less likely that a mistake or malfeasance will result in reactive attitudes, such as blame or anger, directed at other individual members of that institution. However, it also renders people vulnerable enough to blame the institution itself. This, with time, can erode trust in the institution and naturally leads to policy recommendations that aim to preserve institutional trust. I survey two ways in which that can be done with institutional trust in medicine in the age of AI and ML.

2 Institutional Trust in Science

We routinely appeal to the epistemic authority of others to justify our beliefs. This, to some extent, appears to undermine individual epistemic authority in the sense that when we rely on testimony of an expert we do not know ourselves; the expert knows *for us*. Indeed, when we appeal to the knowledge of experts in our reasoning instead of our own knowledge, we relinquish a portion of our rational control over what we believe. We do this, typically, because we believe the testimony of the expert to be true and we trust them in some relevant way (Faulkner 2011, Jones 1996).

John Hardwig points out that the key to preserving our epistemic authority when we appeal to experts is to know who the experts are, what gives them their status as experts, and thus be in a position to rationally accept their epistemic authority over us (Hardwig 1985, p. 336). When we appeal to the testimony of a scientist to back up our claim about, say, planets, we typically appeal to not just any scientist, but a scientist that we know to be an expert about planets. Given this, we should have reason to believe them to be an authority about planets, perhaps by knowing that they published in peer-reviewed journals or got a PhD in planetology from a reputable university.

If we do not know enough about planets, or whatever topic, to be able to assess whether the experts we appeal to have sufficient competence (Goldman 2006), we have to rely on the collective wisdom of communities. For example, we rely on the wisdom of the community of scientists to select the right person to appear to the general public as an expert about planets. Hardwig takes this to suggest that at least in some cases of appeal to the authority of experts, communities are the bearers of our rational beliefs, rather than individual experts.

Importantly, this is true not only of the general public. Appeals to communal wisdom are an important element of the social order writ large (Alfano 2016) and especially important to the functioning of the scientific community. This is because "scientists, researchers, and scholars are, sometimes at least, knowers, and all of these knowers stand on each other's shoulders in the way expressed by the formula: *B* knows that *A* knows that *p*." (Hardwig 1985, p 345). Scientists have to keep themselves informed about who the experts they appeal to are in hopes of preserving their own epistemic authority.

Given all this, when a rational person—scientist or lay person—defers to an expert in some field, this act crucially depends on a specific sort of trust (Irzik and Kurtulmus 2019; Zagzebski 2012). This mental state is functionally distinct from trust individuals put in each other in social contexts, so for example towards partners or friends (Baier 1986). Trust in testimony of an expert crucially depends on the non-expert having a specific sort of *warrant* for that trust. While we should also have some warrant to trust people socially close to us, this warrant is typically mediated, or to some extent even substituted by the emotional bond that we may have with them and expectations about respect for our personhood. For better or worse, we typically (affectively) trust the people we like and who respect us. In ideal conditions, we (epistemically) trust experts not because we like them or because we believe them to respect us, but because we think they are reliable sources of facts (Lackey 2008) or because of some other epistemic function that they fulfil (look: Faulkner 2020 for overview). Trust in science is epistemic, rather than affective (Rolin 2020).

The burden of providing this special sort of warrant lies with the experts themselves, both as individuals and collectively as a community. This means that to have the general public trust expert testimony in the relevant way, scientists have to do more than offer methodologically proper science and have the requisite recognition of competence. Scientists also render themselves trustworthy. But, unlike a person whom we trust because of their social relationship with us, a trustworthy scientist is one who provides the necessary sort of warrant for epistemic trust irrespective of whether they respect us or whether we like them.

That said, a non-expert is rational in believing testimony of an expert only if the latter has the right state of character, including being honest (Anderson 2011). This special sort of character trait "amounts to their commitment to the ethical norms of their trade and their sense of obligation to truthfully and accurately share significant knowledge with the public" (Irzik and Kurtulmus 2018, p 5). For example, if a scientist is known to be (a) trustworthy on account of their belonging to a professional scientific body that concerns itself with planets that has a strict code of conduct and (b) to have competence about planets, then they are worthy of trust when they make claims about planets. If (a) and (b) are met, they are good source of warrant needed to preserve epistemic authority, should a member of the general public or another scientist choose to appeal to their testimony about planets.

Trustworthiness is not just there to provide individuals with warrant—it also insulates science, as an institution, from reactive attitudes (Strawson 1962) that people are prone to have when expert testimony is proven wrong. Scientific claims are often characterized by being falsifiable (Popper 2005/1959) and, indeed, are often falsified by more science. With enough institutional trust, a falsified theory or finding can be understood to be a necessary

part of scientific progress, rather than evidence of malfeasance or incompetence. It becomes just further proof that individual scientists and science as an institution are doing their job. The take-away point is that the other important function of being trustworthy, as a scientist or scientific community, is to mitigate the negative effect that incremental and even revolutionary scientific discoveries have on individual and institutional trust.

From the perspective of the non-expert, institutional trust in science can also dampen the psychological effect of having one's own previously endorsed claims prove to be false. When the public is informed that, for example, some commonly used product causes cancer, they are in a position to accept it, if they already trusted the institution that communicated that information. They are also less likely to become disillusioned, angry, or feel betrayed by endorsing this new claim (even though they may feel so betrayed by the companies that sold that product). The same applies to experts who trust other experts in their community when they hear that their own scientific claims have been proven wrong.

Finally, institutional trust helps science and scientists maintain credibility in light of misconduct of other scientists. Scientists, as a group, know this and benefit from institutional trust as much as the public. As Hardwig himself observes, "the horror that sweeps through the scientific community when a fraudulent researcher is uncovered is instructive, for what is at stake is not only public confidence. Rather, each researcher is forced to acknowledge the extent to which his own work rests on the work of others—work which he has not and could not (if only for reasons of time and expertise) verify for himself" (Hardwig 1985 p 348). Institutional trust is, therefore, also an important part of scientific practice. Scientists depend on it to feel comfortable with standing on the shoulders of their predecessors, instead of spending time validating previous research.

Once institutional trust is gained, it takes effort to maintain it. This burden is, again, primarily on individual scientists, but also on the scientific community as a group. Individual scientists have to consistently conduct themselves in accordance with the ethical and legal norms that govern their profession. The scientific community, on the other hand, has to consistently organize itself to facilitate trust-generating conduct (Kitcher 1983, 2003), create and update such codes of conduct, and distribute rewards and punishments to its members in accordance with merit and role (Strevens 2006). They should also share information about all of this activity. Research, organization, and cases of misconduct need to be appropriately framed, delivered, and disseminated to the public. An important part of this framing, at least in democratic societies, is that science plays a public role and that it deserves to be maintained and funded by taxpayers.

Irzik and Kurtulmus (2021, p. S4733) provide a useful formalism for *basic trust*[1] that summarizes these insights in scientific testimony. Where M is a member of the public, S a scientist or scientific body, and P a proposition, M can place warranted basic trust in S as a provider of P when:

(C1) S believes that P and communicates it to M honestly,
(C2) M takes the fact that S believes and has communicated that P to be a (strong but defeasible) reason to believe that P,
(C3) P is the output of reliable scientific research carried out by S, and
(C4) M relies on S because she has good reasons to believe that P is the output of such research and that S has communicated P honestly.

[1] I am ignoring their notion of enhanced trust, since it introduces further conditions that do not straightforwardly apply to the medical domain.

Leaving aside all the complications that come from the diversity of criteria of competence and reliability across scientific fields, diversity of goals that individuals may have in trusting scientific testimony, and the stratification of competence in the scientific community, basic institutional trust in science as an institution is conditional on (C1-4). Only if these conditions are met can a non-expert rationally trust an expert to *know for them*.

3 Institutional Trust in Medicine

A parallel can be drawn between the conditions for warrant necessary for trust in scientific testimony and those necessary for trust in medical advice. Borrowing from the earlier formalism, where M is a member of the public, S a *medical professional* or a *medical community*, and P a proposition, M can place warranted basic trust in S as a provider of P when:

(C1) S believes that P and communicates it to M honestly,
(C2) M takes the fact that S believes and has communicated that P to be a (strong but defeasible) reason to believe that P,
(C3*) P is the output of reliable medical research or *practice* that S is in a position to trust,
(C4*) M relies on S because she has good reasons to believe that P is the output of such medical research or *practice* and that S has communicated P honestly.

The minor differences between (C3) and (C3*) on the one hand and (C4) and (C4*) on the other highlight the special role of *practice* for trust in a medical context.

Philip Nickel and Lily Frank (2020) point out that trust operative and necessary for fruitful patient-doctor relationships is not only important to medical ethics, but to how the general public thinks about medical practice. Practice here is understood to be all the ways in which a medical professional and the medical profession as an institution come to interact with members of the general public, e.g., in the clinic, hospital, announcements of public policy, recommendations, etc. So, the apparently minor difference between (C3-4) and (C3*-4*) has a potentially large impact on how institutional trust, understood as a mental state, functions in the medical context. There are also good reasons to think that (C3*) and (C4*) do more than provide warrant for trust: they play a central role in defining professional medicine as such.

This is, roughly, what Rosamond Rhodes argues for in *The Trusted Doctor* (2020), which aims to place medical ethics in its own category, distinct from everyday morality and other fields of professional ethics. On her view, the principles of medical ethics are derived from an implicit contract between medical professionals and the general public, which is made explicit in legal instruments and codes of conduct. The nature of this contract overlaps with the nature of the difference between the warrant necessary for institutional trust in medicine and warrant necessary for institutional trust in science. Codes of conduct for medical professionals and legal norms that govern the medical profession pay special attention to respect and dignity of patients, especially during treatment, so the practical side of medicine.

Through this contract the general public agrees to the medical professional having the right to cut bodies, administer poisons, etc., on the condition that medical professional *seek trust and be deserving of it* (Rhodes, 2020 p. 51). All the other principles of medical ethics, such as “loving thy patient” or “keeping oneself up-to-date with respect to medical advances” can be derived from that one foundational principle, Rhodes argues. Importantly, scientists do not enter a contract with the general public that involves their practice, since their practice, as scientists, is primarily confined to interactions with other scientists. The obligations they have to the general public are commensurate to the amount of financial support they receive

from taxpayers through publicly funded institutions. If scientists interact with the general public in their capacity as scientists at all, it is to communicate their results or to gather data. Medical practice, on the other hand, requires that at least some medical professionals intact with members of the general public directly, e.g., as patients.

One of the reasons Rhodes gives for thinking that this implicit contract exists is the observation that patients are not prone to ponder whether to make themselves vulnerable or not when they seek help from a medical professional (Rhodes 2020, p. 305). Instead, they trust that their vulnerability in a clinical situation will be handled with care. People trust the medical profession as a whole and this mediates the trust they put in individual medical professionals. Without this greater context of trust in the profession, these patients would have to engage in means-ends analysis at every doctor's visit, assessing the likelihood that that interaction with this particular medical professional and its associated risks will be of net benefit to them, all things considered.

The other important reason to believe that Rhodes is right about the implicit contract is that institutional trust in medicine was not always there; occultists and charlatans used to be common (Kang and Pedersen 2017; Castiglioni 2019/1947). Over time, trust in medicine and doctors developed because of the professionalization of medicine and the historical fact that medical professionals have consistently discharged their duties towards their patients. The way that doctors and patients interact today is historically unique and rooted in the professionalization of medicine that involved the eventual explicit endorsement of the implicit contract in codes of conduct that put the doctor-patient relationship front and center (Holsinger and Beaton 2006).

The third reason to accept Rhodes' claims is that accumulated trust in medicine can also be eroded by evidence of abuses, such as at the "doctor's trial" in Nuremberg, *United States of America v. Karl Brandt, et al.*, or the "Tuskegee Study of Untreated Syphilis in the Negro Male" cases. Involvement of medical professionals in these abuses has had a significant impact, in terms of subsequent legislation and regulations (for example, the Nuremberg Code, Declaration of Helsinki and United States Department of Health and Human Services Common Rule 45 CFR part 46), but also on the level of trust that specific populations have in doctors, medicine, and medical research (Jacobs et al. 2006; Halbert et al. 2006; Guffey & Yang 2012; Hanson et al. 2016). The disparity between the way that African Americans and American whites interact with doctors can be accounted for by a difference in the level of institutional trust in medicine these groups have respectively (Shavers, Lynch, and Burmeister 2000). It should be noted that this mistrust can be accounted for by social and historical factors that go beyond any individual case of medical abuse, such as the Tuskegee study (Brandon, Isaac, and LaVeist 2005).

4 Technologically Mediated Institutional Trust

In context of science:

> When we take what a scientist has told us as a reason to believe the claim they are making, we take them to be honest, thus displaying some level of good will towards us. It is not the case that we only assume the incentive to tell the truth is what guides their action. If that were the case, we would not feel betrayed if we found out they had lied to us (Irzik and Kurtulmus 2021, p S4734).

By analogy, any member of the public would feel betrayed by a doctor's testimony or advice, if they thought they were telling the truth and believed at the time that the doctor was honest. It is hardly controversial that feelings of blame, betrayal, and other related mental states often have their origin in misplaced trust.

Furthermore, reactive attitudes of people that were wrong to trust testimony coming from a member of an institution are likely to have the institution itself as their object, not just its individual members. After all, it is evidence that the institution is not organizing itself properly—the bad apples are put in a position that signals they can *know for us*. We trust a Professor of planetology at a reputable university to tell us how many planets there are in the solar system at least in part because we assume that they got to be a Professor after they passed scrutiny of other planetologists. If we find out that they got it wrong about the number of planets, our trust in that community's ability to scrutinize its members should diminish. We would be irrational to not calibrate our future attitudes towards planetologists accordingly.

What mitigates the deterioration of institutional trust *in medicine* in similar cases is the assumption that, besides honesty and some level of good will, the medical professional cares about us getting better. In practice, this typically means that they act in accordance with the standards of their profession and codes of conduct. Patients can rationally accept that "there was nothing to be done" or blame chance or themselves when things go badly with treatment, only when they have reason to believe the doctor to care for their well being. Rational patients should assume that medical professionals are professional care-givers, but not professional body-fixers, on a model of a car mechanic. We should not expect of our car mechanics that they care about cars in the way that doctors care about patients.

Nickel and Frank (2020) prognosticate that "through a range of institutional, scientific and technical, and value changes to the practice of medicine" (p 374-5) the nature of trust in medicine may change. They think that, as medicine becomes more reliant on depersonalizing technologies, trust in doctors may play a smaller role. One recent development that makes Nickel and Frank's forecast probable is the growth of internet-mediated medicine (Dyer 2001), health apps, social robots, and the growing use of AI and ML technologies in diagnostics and care (Dilsizian and Siegel 2014). This includes the digitization of records, patient data, and the move to 'personalized' medicine, which puts the patient in control of decision-making with the help of technologies that take advantage of all these other things (Mathur and Mathur 2017). As many have noted, this takes the decision-making process out of the doctors' hands and puts it into AI- or ML-powered software (Nickel 2022).

With AI and ML diagnostics, advice, and maybe even direct involvement in clinical situations, medicine will become more personalized in the sense of putting the patient in the center of the process but depersonalized by taking the doctor more out of it. Some patients will primarily interact with platforms powered by data and algorithms, perhaps through apps on their phone or personal computers. If doctors are involved at all, their role will be to facilitate, consult, and offer advice, as in the old days, but to a much lesser extent. A personal physician may be a luxury. In such a situation, the platform, algorithm, app, and institution of medicine are likely to bear the brunt of reactive attitudes when things go badly for the patient, but patients will also have reason to feel betrayed by the doctors, if they were at all involved. Doctors are not computer scientists or engineers that designed the platforms, so patients will have evidence that they took an "unquestioning attitude too easily, without considering the vulnerabilities and changes they're bringing inside their agency" (Nguyen 2022, p. xx) by the use of AI and ML.

This raises an important question about the nature of medicine, as an institution, in the age of personalized medicine and the nature of trust in that institution going forward. If we take Rhodes' arguments about the crucial role of personal trust to the professionalization of medicine on board, medicine is perhaps not even a distinct profession with its own ethics without it. Some argue that in the age of AI- and ML-aided medicine, instead of institutional

trust, what we will end up with is a return to 'lazy' trust in the Kantian sense of trust in "guardians, i.e. some authority or expert" without the requisite active cognitive attitude (Myskja and Steinsbekk 2020). Lazy trust does not challenge testimony and does not preserve autonomy needed for collaborative decision-making or epistemic authority when others *know for us*. It certainly does not require warrant, as specified by (C3*-C4*). At the least, the conditions for warrant needed to trust medical advice would now be identical to those needed to trust scientific testimony (C1-4). If that is true, trust in medicine and trust science would be psychologically and epistemically indistinguishable.[2]

This would mean that substituting patient-doctor trust with institutional trust, as Nickel and Frank (2020) prognosticate, will have profound effects on how the general public should deal with medical testimony and advice. Things that can potentially undermine the public's trust in testimony of scientists would also now equally undermine the public's trust in testimony of medical professionals. Sadly for medicine, the public's institutional trust in science is not in great shape. Some disciplines, especially those that rely on hypothesis-testing with inferential statistics, are in the midst of a replication crisis, which undermines the veracity of decades of research publications, established theories, and successful careers (Shrout and Rogers 2018).

The replication crisis is not a case of falsification typical of incremental scientific progress, fallout from a scientific revolution, or even pernicious assumptions at work. It has social, cultural, as well as methodological origins (Romero 2019). On the one hand, most scientific disciplines organize on a model that rewards a notable few at the expense of the many. On the other hand, individual scientists have found ways to engage in misconduct, ranging from questionable research practices, such as p-hacking, all the way to outright fraud. The confluence of these two things is made worse by an increasing scarcity of rewards and a publish-or-perish culture in academia. Scientists have a big incentive to become undeserving of trust.

In some disciplines, the replication crisis can be taken as evidence that their respective communities have failed to organize themselves properly (Romero 2017). This would mean that these communities are not a good source of warrant for epistemic trust and members of the public will *not* satisfy (C1-4) when they hear testimony from their members. Unfortunately, scientists that produce reproducible science or are from disciplines not so affected by the replication crisis end up being unfairly lumped together with the bad ones. As personalized trust is diminished by personalized medicine, this lumping would with time apply to medical professionals, too.

Arguably, the most troubling possibility for personalized AI- and ML-aided medicine is a potential replication crisis in computational modelling. Currently, there are few enforced standards of replicability and reproducibility in sciences that rely on the sort of techniques at the heart of medically-related AI and ML models (Miłkowski, Hensel, and Hohol 2018). While it is difficult to imagine what a replication crisis in AI- and ML-aided medicine would look like, we can probably safely assume that it would have a similar negative effect on institutional trust in medicine that the replication crisis has had on institutional trust in science.

Independently of this, there may already be an ongoing crisis of institutional trust in medicine. First, we should note that institutional trust in medicine has been strained by the replication and retraction crisis in medically important scientific disciplines, such as the

[2] This may be a reason to resist the idea that what would remain after the washing out prognosticated by Nickel and Frank will amount to institutional trust *in medicine* at all—but that is a separate issue, outside the scope of this chapter.

biological sciences (Steen 2011). Second, general trust in doctors is diminishing and not only because of COVID-19 (Zhao and Zhang 2019; Griffith et al 2021). Thirdly, we have evidence that testimony of medical professionals is significantly undermined by the online spread of conspiracy theories (Pertwee, Simas, and Larson 2022; Andrade 2020). De-personalizing trust in the age of personalized medicine is likely to amplify these already existing sources of public distrust in medicine.

5 How to Protect Institutional Trust in Medicine

To mitigate the potential crisis of institutional trust in the age of personalized AI- and ML-aided medicine, the institutions that organize medicine will likely have to change. This includes codes of conduct for medical professionals, legislature, insurance policy, and even medical education. The main question should be about the direction these changes take. In context of the arguments and observations made in this chapter, two possibilities are apparent. The first is that institutions that organize medicine should aim to *protect the trusted patient-doctor relationship* from personalized medicine powered by AI and ML. The second is that they should aim to *strengthen institutional trust*, including trust in science. These two possibilities are not mutually exclusive. Let us look at both in turn.

Frank Paquale in his *New Laws of Robotics* offers a set of principles that can be used to design legal instruments that protect the patient-doctor relationship. The second chapter of his book is dedicated to health care and the encroachment of AI-aided technology. There he observes that:

> For most healthy people, doctoring looks like a simple task of pattern recognition (diagnosis) resulting in procedure or prescription. (…) Were things so simple, robots could eventually stand in for doctors. But in the real world of medical practice, this image is long outdated. There is real and enduring uncertainty about the best course of action in many circumstances. And contemporary medicine demands the participation—or at least the understanding—of the patient with respect to a plan of care (Pasquale 2020, p 59).

The last sentence is an important signpost for speculation about what appropriately designed legal instruments, codes of conduct, and other institutional safeguards of the patient-doctor relationship would look like in practice. The focus should be on the role of the patient in collaboratively deciding with their doctor on a course of care. If we want to protect the doctor-patient relationship, any technology or innovation that encroaches on this activity should be put under scrutiny or not used at all.

So much for the patient in the patient-doctor relationship, but what about the doctor? The first of Pasquale's "New Laws" states that "Robotic systems and AI should complement professionals, not replace them" (Pasquale 2020, p. 3) and in our case legislation, codes of conduct, or standards of practice that would embody it would go far in preserving the sort of trust that matters most. The signpost here is the appeal to professionalism and in the case of medicine, the medical professional. Any technology or innovation that encroaches on the professional practice of doctors effectively replacing it, should be put under scrutiny or not used at all.

This two-pronged approach, with focus on the role of the patient in collaborative decision-making on the one hand and, on the other hand, on the role of the doctor as a professional expert, rests on the assumption that the practice of medicine is essentially tied to *human* expertise. Pasquale notes that "the bargain at the core of professionalism is to empower workers to have some say in the organization of production, while imposing duties upon them to advance the common good" (Pasquale 2020, p. 4). Putting AI- or ML-aided

technologies into that process effectively diminishes the burden of duties that medical professionals have to advance the common good. And this is precisely why patients would not be rational to trust them beyond whatever truth or diagnosis they would deliver in a AI-aided clinical setting.

The other strategy for mitigating the potential damage of AI-aided medicine on institutional trust in medicine would be to find ways to independently strengthen institutional trust. In response to fraud, questionable research practices, and violations of the Nuremberg Code and the Helsinki Declaration the medical community, with the help of legislation (at least in the United States and European Union), developed institutions whose primary function is to maintain ethical standards. Institutional responses to other ethically problematic consequences of technological advances include Institutional Review Boards, institutions like the US Food and Drug Administration (FDA), laws like the General Data Protection Regulation (GDPR), or EU's High-level expert group on artificial intelligence, among other things. A similar effort with similar scope and ambitions could be made for the preservation of institutional trust in medicine.

What this would mean in practice should of course be left to the experts, but the reaction of the scientific community to the replication crisis and the crisis of trust in science it portended may be an excellent model. In that effort the scientific community developed instruments such as pre-registration of empirical studies to ostensibly prevent p-hacking. These are repositories with code and hypotheses, methods of analysis, and expected effect sizes which are publicly available on websites, such as the Open Science Foundation (OSF). At the same time, the mandates of ethics committees in many universities were strengthened to include evaluation of compliance with GDPR, data handling protocols, and adequate informational transparency. Finally, the cutthroat competitive culture of academia has, at least nominally, been a focus of recent criticism.

None of these things is likely to fix damage to institutional trust in science any time soon, but they are, all things considered, the best and perhaps only long-term strategy to do so. A similar approach may be needed in sciences directly connected to medicine or perhaps even to medical practice itself. However, there is an important caveat. Legislative and community-based approaches in the context of science are not straightforwardly applicable in the context of medicine. First, the idea of open-medicine is unlikely to sit well with duties towards patients as it is likely to compromise their privacy and perhaps even undermine the patient-doctor relationship. Second, it is not clear what mandate institutions in charge of protecting institutional trust in AI-aided medicine should have. Should we empower these institutions to effectively veto or put in question the development or deployment of technologies that are likely to be effective in saving people's lives, because they could undermine institutional trust? This is a difficult question.

What may be a more promising approach is to focus on explainability of AI and ML models and empower medical professionals to engage with these technologies more directly in a clinical encounter with a patient. A particularly promising avenue here would be the notion of contestability, which is the idea that a patient has the right to contest any medical decision that concerns them by demanding an adequate explanation (Poug and Holm 2020, Ploug and Holm 2022). In cases where such an explanation is not forthcoming, the decision procedure, diagnosis, or other clinical encounter could be legally and ethically suspect. So, a perhaps better strategy than open medicine would be to create a hierarchy of institutions whose sole purpose would be to ensure that clinical encounters with doctors in the age of AI-aided personalized medicine would ensure the right to contestability.

1. Alfano, Mark (2016). The Topology of Communities of Trust. Russian Sociological Review 15 (4):30-56.
2. Anderson, E (2011). “Democracy, Public Policy, and Lay Assessment of Scientific Testimony,” Episteme 8(2), p 144-164.
3. Andrade, Gabriel. "Medical conspiracy theories: cognitive science and implications for ethics." *Medicine, Health Care and Philosophy* 23.3 (2020): 505-518.
4. Brandon, D. T., Isaac, L. A., & LaVeist, T. A. (2005). The legacy of Tuskegee and trust in medical care: is Tuskegee responsible for race differences in mistrust of medical care?. *Journal of the National Medical Association*, *97*(7), 951–956.
5. Castiglioni, Arturo. *A history of medicine*. Routledge, 2019/1947.
6. Dilsizian, S.E., Siegel, E.L. Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. Curr Cardiol Rep 16, 441 (2014). https://doi-org.tilburguniversity.idm.oclc.org/10.1007/s11886-013-0441-8
7. Dyer, K A. “Ethical challenges of medicine and health on the Internet: a review.” Journal of medical Internet research vol. 3,2 (2001): E23. doi:10.2196/jmir.3.2.e23
8. Faulkner, Paul (2020). “Trust and Testimony” in The Routledge Handbook of Trust and Philosophy Ed Judith Simon, New York: Taylor and Francis
9. Faulkner, Paul. (2011). Knowledge on Trust. Oxford: Oxford University Press.
10. Guffey, T., & Yang, P. Q. (2012). Trust in doctors: are African Americans less likely to trust their doctors than white Americans?. *Sage Open*, *2*(4), 2158244012466092.
11. Goldman, Avin, 1986. Experts: Which Ones Should You Trust?, Philosophy and Phenomenological Researh, 63, 85-110.
12. Griffith, Derek M., et al. "Using mistrust, distrust, and low trust precisely in medical care and medical research advances health equity." *American Journal of Preventive Medicine* 60.3 (2021): 442-445.
13. Halbert, C. H., Armstrong, K., Gandy, O. H., & Shaker, L. (2006). Racial differences in trust in health care providers. *Archives of Internal Medicine*, *166*(8), 896-901.
14. Hansen, B. R., Hodgson, N. A., & Gitlin, L. N. (2016). It’s a matter of trust: Older African Americans speak about their health care encounters. *Journal of Applied Gerontology*, *35*(10), 1058-1076.
15. Hafferty F, Salloway JC. The evolution of medicine as a profession. A 75-year perspective. Minn Med. 1993 Jan;76(1):26-35. PMID: 8426585.
16. Hardwig, John, 1985. Epistemic Dependence, The Journal of Philosophy, 82(7), 335-349.
17. Holsinger Jr, James W., and Benjamin Beaton. "Physician professionalism for a new century." *Clinical anatomy* 19.5 (2006): 473-479.
18. Irzik, Gürol and Kurtulmus, Faik, 2019. “What Is Epistemic Public Trust in Science?,” The British Journal for the Philosophy of Science, 70(4), 1145-1166.
19. Jacobs, E. A., Rolle, I., Ferrans, C. E., Whitaker, E. E., & Warnecke, R. B. (2006). Understanding African Americans’ views of the trustworthiness of physicians. *Journal of general internal medicine*, *21*(6), 642-647.
20. Jones, Karen “Trust as an Affective Attitude,” Ethics 107(1), 4-25
21. Kang, Lydia, and Nate Pedersen. *Quackery: a brief history of the worst ways to cure everything*. Workman Publishing, 2017.
22. Kitcher, Philip, 1993. The Advancement of Science: Science without Legend, Objectivity without Illusions. Oxford: Oxford University Press.
23. Kitcher, Philip. *Science, truth, and democracy*. Oxford University Press, 2003.

24. Lackey, J. (2008). Learning from Words – Testimony as a Source of Knowledge. Oxford: Oxford University Press.
25. Mathur, S., & Mathur, S. (2017). Personalized medicine could transform healthcare (Review). Biomedical Reports, 7, 3-5. https://doi.org/10.3892/br.2017.922
26. Miłkowski, M., Hensel, W.M. & Hohol, M. Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. *J Comput Neurosci* **45,** 163–172 (2018). https://doi.org/10.1007/s10827-018-0702-z
27. Myskja, B.K., Steinsbekk, K.S. Personalized medicine, digital technology and trust: a Kantian account. Med Health Care and Philos 23, 577–587 (2020).
28. Nickel, P.J. Trust in medical artificial intelligence: a discretionary account. Ethics Inf Technol 24, 7 (2022).
29. Nguyen, C. Thi (forthcoming). "Trust as an unquestioning attitude," in *Oxford Studies in Epistemology*.
30. Pasquale, Frank. "New laws of robotics." *New Laws of Robotics*. Harvard University Press, 2020.
31. Ploug, Thomas, and Søren Holm. "The four dimensions of contestable AI diagnostics-A patient-centric approach to explainable AI." *Artificial Intelligence in Medicine* 107 (2020): 101901.
32. Ploug, Thomas, and Søren Holm. "Right to Contest AI Diagnostics: Defining Transparency and Explainability Requirements from a Patient's Perspective." *Artificial Intelligence in Medicine*. Cham: Springer International Publishing, 2022. 227-238.
33. Popper, Karl. *The logic of scientific discovery*. Routledge, 2005.
34. Rolin, Kristin (2020) "Trust in Science" in The Routledge Handbook of Trust and Philosophy Ed Judith Simon, New York: Taylor and Francis
35. Romero, Felipe. (2017). Novelty versus Replicability: Virtues and Vices in the Reward System of Science. Philosophy of Science, 84(5), 1031-1043. doi:10.1086/694005
36. Romero, Felipe. "Philosophy of science and the replicability crisis." *Philosophy Compass* 14.11 (2019): e12633.
37. Shavers, V. L., Lynch, C. F., & Burmeister, L. F. (2000). Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. Journal of the National Medical Association, 92(12), 563–572.
38. Shrout, Patrick E., and Joseph L. Rodgers. "Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis." *Annual review of psychology* 69 (2018): 487-510.
39. Steen, R. Grant. "Retractions in the medical literature: how many patients are put at risk by flawed research?." *Journal of medical ethics* 37.11 (2011): 688-692.
40. Strawson, P. F. 2003 (1962). "Freedom and Resentment," in G. Watson (ed.), *Free Will*. Oxford: Oxford University Press, pp. 72–93.
41. Strevens, Michael (2006). The role of the Matthew effect in science. *Studies in History and Philosophy of Science Part A*, *37*(2), 159-170.
42. Zagzebski, Lisa (2012) Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief, Oxford: Oxford University Press.
43. Zhao, D., Zhang, Z. Changes in public trust in physicians: empirical evidence from China. *Front. Med.* **13,** 504–510 (2019). https://doi-org.tilburguniversity.idm.oclc.org/10.1007/s11684-018-0666-4