Society of Medical Education & Research

2023, Volume 3, Page No: 26-39
ISSN: 3108-5059

Asian Journal of Ethics in Health and Medicine

Building Trust in the Application of Machine Learning Algorithms for Rare Disease Diagnosis

Lukas Lembo[1], Marcello Barra[2], Alexander Iriti[3]*

[1] University Of Salerno, Via Giovanni Paolo II, 132, Fisciano, 84084, Campania, Italy.
[2] University of Naples Parthenope, Via Ammiraglio Ferdinando Acton, 38, Naples, 80133, Campania, Italy.
[3] La Rochelle Université, 23 avenue Albert Einstein, La Rochelle, 17031, France.

***E-mail** ✉ Iritialex67@gmail.com

Abstract

As AI becomes increasingly integrated into healthcare and computerised systems influence clinical decision-making, addressing both trust in and the trustworthiness of AI tools is critical. Focusing on computational phenotyping (CP) for diagnosing rare diseases in dysmorphology, this paper investigates the conditions under which medical AI tools employing machine learning can be trusted. Semi-structured qualitative interviews (n = 20) were conducted with stakeholders involved in designing or using CP systems, including clinical geneticists, data scientists, bioinformaticians, industry representatives, and patient support group spokespersons. Interview data were analysed using the method of constant comparison. Participants highlighted the centrality of trust in the deployment of CP technology for rare disease diagnosis. Trust was conceptualized in two interconnected ways. First, they emphasized the importance of trust relationships: patients must trust the clinicians using AI tools, and clinicians must trust AI developers, in order to facilitate adoption. Second, participants stressed the need for trust in the technology itself, or epistemic trust in the knowledge it generates. CP tools may be viewed as more trustworthy if their reliability and accuracy are verifiable and if the users or developers themselves are trusted. The findings indicate the necessity of intentionally designing AI systems that are reliable and confidence-worthy for healthcare applications. Moreover, establishing robust processes—such as randomized controlled trials (RCTs) or frameworks ensuring accountability, transparency, and responsibility—can help affirm the epistemic trustworthiness of these tools.

Keywords: Qualitative research, Epistemic and relational trust, AI, Electronic phenotyping, Trustworthiness, Computational phenotyping, Rare diseases, Digital phenotyping

Introduction

Ways of looking at trust

Trust can be understood as a relational phenomenon—an intentional attitude or disposition—arising in contexts of uncertainty, dependence, and expectations about future actions or intentions [1]. When framed as a property of relationships, trust involves A (the truster) placing confidence in B (the trusted) to X [2, 3], whereas trustworthiness refers to qualities of B, described as "…the commitments, virtues, traits or features [of B] that ground justified or well-placed trust" [4: 24]. Accordingly, if B is viewed as dishonest, unreliable with confidential information, or inconsistent and incompetent, they will be judged untrustworthy and avoided in trust-based interactions.

Trust and trustworthiness are deeply intertwined with risk and uncertainty. Bauer [5] defines trust as ".a belief formed as a result of probabilistic reasoning…a probability that quantifies a belief that the trusted person will in fact do what one is expecting her to do." [p4], meaning that A trusts B because A considers it likely that B will fulfill the expected action. Similarly, Starke *et al.*

Access this article online https://smerpub.com/

Received: 07 January 2023; **Accepted:** 11 March 2023

How to cite this article: Lembo L, Barra M, Iriti A. Building Trust in the Application of Machine Learning Algorithms for Rare Disease Diagnosis. Asian J Ethics Health Med. 2023;3:26-39. https://doi.org/10.51847/Mo7NXmiBnA

[3] emphasize that trust helps manage uncertainty, not only concerning whether B will perform X but also when A is vulnerable or reliant on B's goodwill to X [2].

Trust is context-sensitive: Starke *et al.* [3] suggest that whether A decides to trust B depends on A's readiness to trust, perceptions of B, past experiences with B or similar situations, and the broader circumstances surrounding the task. These considerations influence both B's perceived trustworthiness and A's willingness to place trust, illustrating that trust is conditional and fluctuates over time. As O'Neill observes: "Our aim—everybody's aim—surely is to trust the trustworthy, but not the untrustworthy." [6: 293]. The challenge lies in identifying who is deserving of trust and determining the direction in which trust should be allocated.

Trust in healthcare

Clinical relationships between patients and clinicians exemplify trust-dependent interactions. These relationships inherently involve uncertainty regarding prognosis, diagnosis, and treatment, as well as dependence, with patients occupying a vulnerable position relative to clinicians. Expectations are that both parties will act appropriately, particularly that clinicians will prioritize patients' best interests. These relationships depend on the sharing of information and feelings, operating on the assumption that confidences are respected and that both parties are trustworthy.

Like other trust-based relationships, clinical trust is dynamic and constantly renegotiated [7]. Research highlights that ongoing care is crucial for forming perceptions of general practitioners' (GPs') trustworthiness [7]. Key factors sustaining trust include effective information exchange, respecting patients' values, and ensuring patients feel acknowledged and taken seriously [1, 8]. Robb and Greenhalgh [9] note that competent clinicians are also expected to demonstrate empathy, care, and respect. Ward [7] observes that patients often mistrust locum GPs due to their lack of familiarity with individual patients, limited social knowledge, and weaker interpersonal skills. Such distrust can lead patients to disregard advice, question diagnoses, and fail to adhere to prescribed treatments or even fill prescriptions.

Ways of looking at trust in AI (in healthcare)

Recent discussions have emphasized that AI technologies, particularly those utilizing machine learning algorithms, are well-equipped to perform specific healthcare functions. For example, these tools can interpret pathological and radiological images with considerable accuracy and reliability, thereby enabling clinicians to allocate more time to other aspects of patient care [10]. Several authors have argued that as AI adoption grows and black box computerized systems increasingly inform clinical decisions, attention must be paid to both the concept of trust in AI and the assessment of AI tool trustworthiness [3, 11, 12].

It has been argued that although humans may depend on AI, they cannot establish a trust relationship with it [13]. Trust inherently requires recognizing another's goodwill toward oneself, a property unique to human actors, as "…one can only trust things that have wills" [15:14]. In this paper, we do not engage in the debate over whether inanimate entities, such as AI tools, can participate in trust relationships (see [15]), nor whether non-human agents—such as technologies, institutions, or practices—can be deemed trustworthy (but see [16]). While we acknowledge Metzinger's [14] concern that labeling technology as trustworthy or untrustworthy may be conceptually flawed, we agree with Starke *et al.* [3] that everyday language often allows people to claim trust (or distrust) in inanimate objects or institutions. Often, this trust is indirect; for instance, describing a bridge as trustworthy [3] and relying on it to carry weight from point A to B implicitly reflects trust in the bridge's designers or builders, who are believed to have endowed it with features perceived as reliable. The focus of this paper is not whether AI itself can be trusted or judged as (un)trustworthy, but rather the conditions under which medical AI should be trusted [3].

Empirical studies of trust in AI (in healthcare)

Despite extensive theoretical discourse on the importance of trust in AI and its conceptualization, empirical investigations into trust regarding AI tools in healthcare remain scarce, particularly concerning the perspectives of AI developers and clinicians. Capturing these viewpoints is critical since developers influence the design of AI tools, while clinicians apply these tools within existing trust frameworks. From a developer standpoint, it is important to assess how they value trust and how trustworthiness could manifest in the AI systems they create. From the clinician perspective, understanding how AI may alter established trust

relationships and how the trustworthiness of AI tools is evaluated is essential.

A recent physician survey explored this by presenting a hypothetical clinical scenario involving a machine learning risk calculator for pulmonary embolism prediction, investigating the relationship between physicians' comprehension of algorithmic outputs, their ability to communicate these outputs to patients, and their intended clinical actions [17]. Intended actions were considered a proxy measure for trust in the algorithm. In the scenario used by Diprose *et al.* [17], the risk estimate was presented in varying formats alongside treatment recommendations—either discharging the patient or referring for an angiogram. Findings indicated that physicians who better understood the model's output and could explain it to patients were more likely to follow the algorithm's recommendation. The authors concluded that a physician's ability to interpret and explain model output correlates with higher trust, as evidenced by a greater willingness to adhere to the model's guidance.

While several issues can be identified in this study, two warrant particular attention. First, the scenario provided only limited clinical information about the individual case, intentionally designed to prevent clinicians from disagreeing with the ML output, which may have prompted them to align with and follow the recommendation. More critically, the lack of detailed clinical information indicates that the scenario does not accurately reflect real-world diagnostic encounters in which such tools are likely to be implemented. Second, it remains uncertain whether (hypothetical) intended behavior—choosing to follow or not follow the algorithm's recommendation—serves as a valid proxy for trust. It is possible for clinicians to follow AI recommendations even if they perceive the tools as untrustworthy, particularly in forced-choice hypothetical situations.

A recent qualitative study conducted in France [18], described as the first to investigate AI researchers', clinicians', and other stakeholders' (e.g., ethicists, lawyers) perceptions of AI and its implications for healthcare practice, found that insufficient explainability and limited understanding of AI were seen as potentially undermining the doctor-patient relationship and disrupting healthcare organisation. However, Lai *et al.* [18] also observed that clinicians were optimistic about AI's potential, noting that it could enhance patient care if the healthcare community receives proper training and understands how these tools function. Similarly, researchers expressed support for AI use in healthcare but emphasized that, currently, these tools should be assistive rather than replacement technologies. Other stakeholders highlighted the importance of informing patients about AI's potential harms and benefits and stressed the need for regulatory frameworks to bolster trust in AI. Ultimately, Lai *et al*.'s study underscores the necessity of collaboration among different stakeholder groups for successful clinical AI implementation: researchers need access to clinical data and feedback on usability, while clinicians require understanding of AI design and feasible applications.

Although these studies [17, 18] provide valuable insights into stakeholder perspectives on AI in healthcare, their limitation lies in the reliance on hypothetical scenarios or generalized discussions about AI with participants who may have limited familiarity with the technology—though Lai *et al.* [18] did include a small number of radiologists with more experience using these tools.

Hui *et al.* [19] partially address this limitation by examining patients' and clinicians' trust in the internet of things to support asthma self-management. Semi-structured interviews involved patients with asthma, some already using digital devices for management, alongside primary, secondary, and community clinicians caring for this patient group. Participants were asked to conceptualize an AI system to assist asthma management. The study found that while patients were receptive to hypothetical AI tools and perceived them as helpful for self-management, they were hesitant to allow AI systems to operate autonomously in certain areas, insisting that new management advice or diagnoses be approved by clinicians. Clinicians similarly acknowledged that AI systems could support self-management, for example, by uploading patient data for clinician review, but expressed skepticism regarding their capacity to drive behavior change. Both patients and clinicians emphasized the need for clinician oversight of AI-generated advice and diagnoses and highlighted that trust would depend on access to the underlying evidence base.

Finally, a recent study focused on researchers and clinicians directly involved in developing AI tools explored how AI algorithms come to be trusted in clinical decision-making [20]. Winter and Carusi [20] argue that conventional elements of AI trust, such as algorithmic explainability, transparency, and the ability to justify output, are insufficient on their own. Instead, trust emerges through negotiation and collaboration between

developers and clinical users during the validation process. Based on seven interviews with clinicians and AI researchers engaged in developing three AI tools (screening, imaging, and biomarker algorithms) for early detection of pulmonary hypertension, the authors suggest that validation processes are central to fostering trust in AI tools. They contend that involving clinicians alongside researchers throughout the design, development, and implementation phases is fundamental to validation and trust-building. Reflecting Lai *et al.* [18], Winter and Carusi emphasize that developing trustworthy AI requires collaboration among stakeholders, particularly during validation. In essence, they conceptualize epistemic trust in AI tools as rooted in the trust relationships formed between designers, developers, testers, and users, cultivated through collaborative practices in AI development. This paper aims to extend empirical research on trust in healthcare AI by examining the perspectives of those who design and work with computational phenotyping algorithms.

Computational phenotyping of rare disease

Computational phenotyping (CP) serves as a tool to enhance the diagnosis of rare disorders that present with dysmorphic facial characteristics [21, 22]. The method involves using facial photographs and additional biomedical data from individuals who have a confirmed clinical or molecular diagnosis of a rare (typically genetic) disease. These data are employed to train facial recognition (FR) algorithms to detect and classify the unique facial phenotypes associated with different rare conditions. Because these diseases are uncommon, assembling effective datasets requires global collaboration to obtain digitized patient images. The Minerva Initiative, a worldwide consortium comprising academic, clinical, and commercial researchers, focuses on advancing CP through FR algorithms [23]. This consortium provides a secure platform for sharing digitized facial images and related data, promoting research in the field. Within this framework, the Facing Ethics Project sought to understand how consortium members perceive the ethical challenges that emerge when CP technology is applied in both research and clinical contexts.

Methods

Given the scarcity of knowledge regarding the perspectives of stakeholders involved in AI tool development, this study adopted a qualitative, exploratory design. In-depth interviews were conducted to probe participants' understanding of the ethical issues linked to the use of AI, particularly CP tools, in healthcare and research settings.

Recruitment of participants

Following ethical approval from the University of Oxford's OXTREC committee, participants were invited via email from the Minerva Consortium membership list, supplemented through snowball sampling using authors' and interviewees' professional networks, including industry representatives and patient advocacy groups. Of the forty-seven individuals approached, two declined participation, one offered partial feedback via email, and twenty (42%) consented to be interviewed. The remainder did not respond despite follow-up emails. The final cohort constitutes an opportunity sample, primarily consisting of Minerva Consortium members, alongside one industry representative and one spokesperson from a patient support group **(Table 1)**.

Table 1. Participant characteristics

Location	
Africa	1
Europe	5
Australia	5
US	6
UK	3
*Expertise**	
Clinical genetics	9
Paediatric genetics	2
Bioinformatics/data science/computational biology	5
Other clinical speciality	1
Other academic discipline	2
Commercial	2

*N = > 20 as some interviewees fall into two categories for example Academics who are involved in commercial spinouts of CP tools

Data collection

SB carried out the data gathering through detailed interviews, each lasting no longer than 60 minutes, conducted remotely via telephone or Skype during March and April 2019. With participants' permission, all sessions were digitally recorded. The interviews initially

explored participants' roles in relation to CP and their understanding of AI systems in healthcare. Following this, participants were invited to share their perspectives on the advantages and limitations of computational phenotyping. Additional questions emerged either from participants' responses or from recurring themes identified in the AI and ethics literature. These aimed to capture opinions on topics such as the use of CP in rare disease research and care, the distinction between facial images and other personal data types, considerations of privacy and consent, challenges surrounding data-sharing in public versus private initiatives, as well as the effects of data siloing, algorithmic bias, and incidental findings [24].

Data analysis

The transcribed interviews were carefully reviewed multiple times by NH to detect recurring patterns within and across participants' accounts. Using the method of constant comparison [25], a coding framework was developed in collaboration with SB and systematically applied to all transcripts. This process yielded four main themes: the influence of computational phenotyping on dysmorphology practice, managing expectations about AI technology, trust in AI, and the evaluation of costs and benefits associated with CP tools in dysmorphology diagnosis. Although trust in AI was identified as a standalone theme, it intersected with the others; for example, participants described that realizing the benefits of CP relies on trust, while expectations regarding AI are shaped by the presence or absence of trust. The findings indicate that trust is seen as a cornerstone for the acceptance of machine learning in both healthcare research and clinical practice, and the analysis below explores how such trust can be cultivated.

Findings

All participants highlighted the critical role of trust in the adoption of CP technology for rare diseases. Trust appeared in two closely linked forms. First, participants emphasized that CP tools must be integrated into a context of relational trust: patients need confidence in clinicians who use AI, and clinicians—and to a lesser extent patients—must trust the developers of AI systems for successful implementation. Second, there was a clear need to establish trust in the technology itself, specifically in the reliability and validity of the knowledge it generates—what is referred to as epistemic trust.

Trust in users and developers

Participants reflected on how CP tools function within clinical encounters. They noted that clinicians' familiarity with rare disorders varies, and objective CP tools can enhance diagnostic capabilities, particularly for those with limited experience. P007 explained that CP could reduce uncertainty in non-specialist diagnoses by offering more standardized, objective assessments:

P007 …the rest [non specialist clinicians] of us could benefit from a little science, I think. So I think that the facial phenotyping software and the development of artificial intelligence that the computer can help you think, "Go this way," I think is very useful. And I think that it takes out both the individual emphasis that a person puts on, like look at those eyes versus look at that nose, as well as you just standardise the measurements, it is what it is.

In diagnosing rare diseases, clinicians typically rely on a combination of phenotypic and genotypic information in addition to specific facial features. CP tools, therefore, offer only a fraction of the total evidence necessary for clinical decision-making. Reflecting this, most interviewees described AI as a complementary resource: a means of providing supporting evidence rather than a replacement for clinician judgment. P015, a data scientist, emphasized that clinicians' trust in CP tools often depends on whether the outputs align with their own assessments, noting that disagreement could lead to skepticism or dismissal of the technology.

P015: I don't think people [clinicians] are ready to trust these algorithms at face value. I think if it supports what they are saying then that makes sense, it's in line with other evidence that we have, I think that's also OK. Oh, we had a mutation in that gene and then the face algorithm is saying that it's also that gene – that makes sense. But the interesting situation comes when the face algorithm says, no, it's something completely different. And that, I think, would challenge clinicians, I think, when the face disagrees.

The interviewees also explored patients' perspectives, noting that trust in CP outputs alone is insufficient. While the technology may be scientifically accurate, diagnosis requires interpretation, explanation, and justification—tasks that demand clinician involvement. In other words,

patient trust hinges on the human element within the diagnostic process.

P014: I think a lot of people [patients] always trust the doctor more. And that, you know, let's say the machine learning goes to a point where someone can have their facial phenotypes analysed by a machine, by artificial intelligence, I think, people always, in my opinion, are going to trust the person's [clinician's] opinion more than artificial intelligence, which is probably a good thing. ... Because I think you will always have the person, if facial phenotyping at least always has a validation point at the end where it's done by a specialist.

Some participants highlighted that patients often fear machines taking over diagnostic responsibilities, and they reassured that such scenarios are unlikely. The technology is intended to assist, not replace, the physician, serving as an adjunct rather than a substitute.

P019: I think it's mainly that people [patients] are afraid of a computer taking it over… I think people should be aware that this is in addition to, it's not going to replace the physician, never going to replace your consultation. I think that's what people are afraid for, is that they go to the hospital and will go to a machine and they'll push in some buttons and say, "Rat's serious," or, "You have this and this disease." ... I think that's the point to make clear, that this is there to help the clinician.

For many interviewees, the cornerstone of healthcare remains the relationship between doctor and patient. They argued that even the most advanced AI cannot substitute the trust embedded in human interaction. As P010 put it:

P010: I don't want to blame the technology, because the technology is just the tool. It has always been just the tool. For me, it's all about what use you make of the technology. So it's all about the person…I believe that human to human interaction is a cornerstone of trust. So this is where I think AI will always come short. . but there are moments, and we need to define what these moments are, where I believe human interaction is what makes us human. So what makes a difference from dealing with a person like a number or dealing with a person like a person.

In summary, our interviewees emphasized that CP technology should not function as an independent diagnostic tool, but rather as a support to trusted healthcare professionals, enhancing their clinical expertise. They highlighted that human involvement is essential for fostering patient confidence in algorithmic outputs, since effective healthcare goes beyond assigning diagnostic labels. As P010 noted, caring for patients requires the ability to "…deal with a person like a person," something that machines cannot accomplish. Consequently, CP tools are best applied within the framework of an established clinician–patient relationship, given that diagnosis constitutes only one part of comprehensive medical care.

While interviewees viewed relational trust between clinicians and patients as a prerequisite for CP tool use, they also acknowledged that trust in the technology—on the part of both clinicians and patients—could be challenged when commercial companies are involved in its development. Several clinicians recognized that commercial investment is often necessary for healthcare innovation, including AI, without which certain technologies might not be developed.

P004: I think we have to be mature about this and think about how commercial partnerships are entered into and engaged in. I mean many good things have happened with public [private] partnerships or that have flowed from commercial operations. I think we just need to acknowledge the real opportunities and having a way of managing that.

Nonetheless, some participants were cautious about commercial involvement, suggesting that it might compromise the trust central to the clinician–patient relationship. P006 expressed concern that families might perceive advice as biased or influenced by commercial interests:

P006: I'm sort of nervous of it. I think [patients'] families are not quite at that level yet. So I don't tend to get too involved with commercial organisations, simply because, when you are speaking to families, you then have to declare an interest, and I think that does impinge on trust a little bit. ... I suppose I'm very used to people just feeling that my agenda is their child. And that's how I want them to feel. I don't want them to think my agenda is a company that I am working with or something…you just have to be cleaner than clean.

A number of interviewees expressed skepticism toward commercial motives, particularly in relation to designing AI for healthcare, raising concerns that this might hinder AI development:

P008: Re other issue is, if this were theoretically possible, what I've just described, then I simply wouldn't have enough faith in the big tech companies these days to not think that it wasn't going to go awry. And that would fundamentally undermine the whole point of the project. Because, if you didn't have trust, then people

[clinicians] wouldn't do it. And if people wouldn't do it, the whole thing is pointless.

Several participants referenced recent controversies, such as the unconsented use of data by Cambridge Analytica and Google DeepMind, highlighting concerns over the lack of transparency in who manages and protects patient data. They suggested that there is a broader crisis of relational trust between commercial data-handling entities and the public. P008 elaborated on this point:

P008: I think it's very difficult because you've got a handful of very, very large companies which seem to have a lot of control over this world… they are so significantly ahead of the curve in terms of what they are doing, and they are so opaque in terms of what they are doing, that I can't see people really having genuine trust for a very, very, very long time, and it would only come with massive changes in the way they operated. And I think they get away with it because they have made things which are just so useful and helpful that we just can't be bothered to do without them. If I fundamentally distrusted Facebook I would have cancelled my account, and I haven't. So I think there is distrust and dissatisfaction and grumbling, but nevertheless people continue to use what's on offer. So even in the context of lack of trust, they'll still exist and they'll still keep going.

In certain situations, when individuals have no alternative means to access a desired service—meaning that refusing to 'trust' would prevent them from obtaining it—they may act "as if" they trust due to their dependence on the service. This behavior, however, does not necessarily reflect a genuine trust relationship. As P008 points out, there is still "distrust or a lack of trust here"; it is the absence of feasible alternatives that produces what appears to be trust-like behavior.

Some participants suggested that distrust toward commercial developers could be reduced or offset if the technology they produce is perceived to generate broader societal benefits. In particular, they argued that companies should engage in benefit sharing, especially when AI is developed using publicly sourced healthcare data.

P006: Not really. I think it is important that the NHS or whatever healthcare entity, like the third world, if they're getting involved with that, they see some benefit. If they are giving data that's going to be used for commercial purposes then I think they should benefit in some ways from that. I think the NHS should benefit from IP that's generated.

Overall, many interviewees acknowledged the increasing importance of public-private partnerships in technology development, particularly within publicly funded healthcare systems such as the NHS. However, they remained skeptical of large technology companies, suggesting that public benefit sharing could be a strategy to foster public trust. Participants also emphasized that trust is needed not only in the organizations handling data but also in the AI tools themselves. As P003 explained, ensuring epistemic trust requires confidence that AI tools are reliable, rigorously tested, and supported by robust data governance systems.

P003: Re NHS is having lots of funding problems and I think taking commercial stuff out of all of it is closing the door to so much available resources and money. So I think it's sensible that we explore all of these options, but I think it's important that we do our due diligence and we trust that they – that the tests they are doing are accurate, that they are reproducible, that the quality is high enough, that they look after the data properly, they analyse the data fully, you know, all of those sorts of things.

Trusting technology: epistemic trust in AI

While trusting the individuals who use or develop AI is crucial, our interviewees also stressed that clinicians and patients need to place trust in the technology itself, or in the validity of the knowledge it generates. Several participants highlighted that CP technology is still new, making it difficult for people to trust at this stage.

P011: And will people trust those results? Rey probably will. I think it's just people have a trouble trusting something that's new, but once it becomes part of like, …I think it's that people find these new technologies a bit scary at first but they just become normal. So [CP technology] will become just what everybody gets used to using.

Epistemic trust in CP tools was viewed as something that would develop over time. Many participants suggested that trust in algorithmic outputs could be strengthened through the accumulation of evidence demonstrating that these tools work reliably and produce trustworthy knowledge.

P007: Well, I think it's proof, accruing evidence is what will lead to trust. I think there's certainly potential for harm, in that it really depends on who is programming the machine, what it's being told to do… I don't think it's appropriate to have [a ton of] trust in machine learning

yet. I don't know that it's proven its point and worth completely.

Interviewees indicated that clinicians' limited epistemic trust in CP tools stems both from the novelty of the technology and their limited experience in applying it for diagnostic purposes. P006 suggested that, over time, CP tools—like other medical instruments—could come to be regarded as reliable sources of clinical knowledge:

P006: I think the trust is really in the scientific method and actually saying, well, what are you actually trying to do? ...So I think it's just a matter of working out what you want a machine to do and how are you going to train it to do that and how are you then going to monitor its performance. And it's the same as any piece of machinery that you'd use in the clinic, you've got to have ways of monitoring how it's performing….So trust, to me, is really, I mean do I trust my stethoscope? Yes, because I've been using it for nearly 40 years. So do I trust an ECG? Yes, I do, if it's done by someone who is experienced and knows where to place the leads and knows how to calm the child down. …. And the precautionary principle is, don't do anything in medicine unless you know what you're doing, and you don't trust things you don't know. And have a reasonable idea, if you've got a machine there, have a reasonable idea how it's working.

Participants emphasized that trust in technology develops through hands-on experience. As P006 noted, trust is earned rather than demanded: "…you gain trust through experience. You can't demand trust, you can't demand that somebody trusts a computer." Technology itself can also acquire credibility over time as it becomes validated against other evidence that confirms the accuracy of its outputs. P015 observed that new technologies are often introduced in parallel with the tools they are intended to complement or replace, allowing epistemic trust to build gradually:

P015: Trust, yeah you have to gain it. You can't just put a computer in a room and say, "Right, you need to trust it now it's going to diagnose all your patients based on a face," that's not the way it works. I mean many of these clinicians have had very many years of experience as well, so they've seen a lot, so they think they also are right (laughs)…Yeah, the stereotypical way of getting a new technology into diagnostics is that you run it in parallel for a time, and then sometimes it agrees with the existing technology that provides results and then sometimes it gives new insights and then eventually the trust is there and you make the switch. … People have to be willing to take that leap of faith to run it in parallel for a while.

With respect to CP tools specifically, interviewees noted that molecular testing can serve as a backup or verification mechanism for AI-generated knowledge, and vice versa. In other words, clinicians' short-term assessments of trustworthiness are often grounded in some form of objective, external validation.

P002: To be honest, at this moment, it's [CP technology] not good enough. But I can say it's not good because I have the experience. But in 10–15 years' time, [when] people like me [specialist dysmorphologists] are not working any more if you blindly trust the outcome of the algorithm and stop thinking then you might make the wrong diagnosis. But if you use it, if you can actually still confirm by molecular testing the outcome of the algorithm, then it's helpful….

This perspective prompted some interviewees to question whether CP technology could ever be trusted to operate without human oversight, arguing that clinician involvement is essential for establishing epistemic trust in AI. Consequently, the majority speculated that patients are likely to perceive relational and epistemic trust as mutually dependent: epistemic trust in CP tools is influenced by relational trust in the clinicians using them. In other words, if a trusted clinician employs AI, patients are more inclined to trust the resulting algorithmic output.

P011: I think most people think that the algorithms are better, I would have thought. I think they are very good, all of the algorithms. But I think, at the end of the day, there is a doctor involved and … I think people are more likely to trust something where the doctor is involved, rather than just, "Ris algorithm said you've got this diagnosis." Whereas if the algorithm is helping the doctor make a decision or helping a lab find the variant, I think that's different. I think helps people have trust.

Our findings indicate that clinicians' trust in CP tools is closely linked to their perceptions of the tools' reliability and validity—essentially, whether they function correctly. Across interviews, participants consistently emphasized that AI technology must undergo rigorous reliability testing and cross-validation.

P008: So the AI itself, I think that trust would be established by illustrating that it worked, much like I would trust a new cardiovascular drug if there was a good RCT showing that it worked. I don't think I'd need any more evidence. Rat just good, high quality level of evidence would do the trick for me.

Several interviewees highlighted that epistemic trust depends on transparency and understanding of how algorithms operate, rather than uncritical or "blind faith" in their outputs. One data scientist explained their efforts to develop interpretable CP algorithms that clearly show the components informing their decisions, making the tools more transparent for clinicians. Others, however, noted that creating explainable CP tools is challenging, since current CP-based diagnostics are probabilistic and do not rely on anatomical features in the same way human diagnoses do. This means that algorithmic outputs are presented in a format unfamiliar to clinicians.

P015: Re computer analyses this face and it says, yes, they are the same. But it cannot say in a concrete way it's the nose or it's –, it's like it's generalising and it's seeing something. And that's then very hard to explain to clinicians, because clinicians are used to looking at a face and saying, oh, it's the ears, or it's the eyes. So translating, I think that's a problem with some of the deep learning methods, they are abstract and that makes it very hard for a clinician to understand why the computer is saying what it's saying. Because ultimately the computer just pumps out a number it says, yeah, there is a probability of 83.2% that this person has something or other.

Concerns about algorithmic bias were also raised, particularly its potential impact on different ethnic groups, which could undermine the perceived trustworthiness of CP tools.

P003: So I think really what you're asking is how will people have faith and trust that the algorithms are not biased and they are accurate? I guess by becoming more prevalent and proving that they work, is the obvious answer. Rat if they make diagnoses that are then confirmed by a molecular diagnosis, then that is proof in itself. I haven't used this software so I can't speak from personal experience, but I think the concern will always be, with rare diseases, has there been enough data?

Bias remains a significant challenge in CP research because the limited number of individuals with rare diseases raises questions about dataset representativeness and, consequently, the reliability and validity of algorithmic outputs. A data scientist highlighted the difficulty of this limitation:

P015: So I think, yeah, I think the biggest thing is that the methods need to be able to handle not having 50 patients. It's possible but it's, yeah, it's tough. You know, you're on the border of your power and that means that your confidence goes down a bit, which then makes it harder to accept what the computer is saying.

Many participants suggested that increasing transparency regarding training data composition and algorithm functionality could partially mitigate these issues. P020, an industry spokesperson, explained:

P020: Well, the first thing that you need to do is you need to be transparent. So you need to publish your work… In many senses it's easier to trust artificial intelligence if you have more information, right. So as long as you can make sure that it has seen enough cases, enough diverse cases and the network is up and running…So I think trust is going to get a different perspective in the future.

Finally, a small number of participants argued that not fully understanding how AI functions does not necessarily prevent its use. P012 noted that the emphasis on transparent and explainable AI may have been overstated, pointing out that even human clinicians, particularly those with less experience, often cannot fully explain how they arrive at a specific diagnosis:

P012: I mean it's definitely challenging to understand more about what these algorithms do. Rey are often referred to as black boxes, but still I think many clinical experts are black boxes too, all their trainees too, right, and they can hardly explain why they come up with a certain differential diagnosis. So it's not that different, I think. Ultimately you have to trust in the technology, in the knowledge or service it provides.

The data indicate that having confidence in CP tools themselves is as critical as trusting the individuals who use or develop them. Interviewees emphasized that epistemic trust in CP tools is conditional and develops over time; clinicians are likely to build trust through repeated use, with reliability and performance validating the technology. In other words, AI tools gain credibility as they demonstrate consistent, externally validated results. Participants also highlighted that increasing transparency around how these tools function and are trained could enhance perceptions of trustworthiness.

Discussion

The findings from this study suggest that trust in AI tools in healthcare encompasses both relational trust—trust in those who use or develop AI—and epistemic trust—trust in the knowledge produced by the tools. Our interviewees proposed that CP tools used for diagnosing rare diseases are likely to be perceived as more trustworthy by clinicians and patients when the users can vouch for the

technology's reliability and accuracy, and when the individuals using or developing the tools are themselves trusted. The following discussion explores the interplay between relational and epistemic trust in the context of AI in healthcare.

Who should we trust?

As noted by previous authors [2, 3], trust is embedded within social interactions, with factors such as power, expertise, and expectations of both the truster and the trusted influencing how trust is established. In clinician–patient relationships, trust arises from both personal and structural dimensions; it is anchored in a clinician's expertise and knowledge [26], as well as in confidence in the institutions and regulatory systems supporting healthcare delivery [27]. Epistemic trust—relying on healthcare professionals' testimony about our health—forms the foundation of relational trust, which is grounded in belief in the clinician's expertise and goodwill toward the patient.

Despite the rise of internet-informed 'expert' patients and AI tools, clinicians still possess the most relevant expertise in this relationship. Our interviewees similarly suggested that from a patient's perspective, trust in CP tools is contingent on trust in the clinicians using them, who are regarded as the ultimate arbiters of AI's trustworthiness. Consistent with studies by Lai *et al*. [18] and Hui *et al*. [19], participants in this study anticipated that CP tools will not be deployed without human oversight [10]; rather, these tools are expected to augment human decision-making within the context of pre-existing trust relationships. Returning to O'Neil's question [6]—"how do we determine who is trustworthy or where/to whom we should direct our trust?"—the evidence suggests that in the context of AI in healthcare, patients should place their trust in clinicians who can demonstrate the requisite skills and expertise to use these tools effectively.

Although our interviewees believed that future patients would place confidence in them and other healthcare professionals to apply AI tools responsibly, they expressed reservations about the underlying motives of commercial developers, particularly large technology corporations. Several participants suggested that the credibility of AI developers could be strengthened in the eyes of both society at large and clinicians if these companies engaged in benefit sharing. Yet, as Kerasidou [12] points out, profit sharing on its own may not necessarily enhance perceptions of the trustworthiness of either the developers or the technology. She emphasizes that since trustworthiness is "self-motivated and self-regulated," the focus should be less on persuading commercial AI actors to behave more ethically and more on establishing clear regulatory frameworks and transparent accountability mechanisms. In short, external regulation should be implemented to ensure that commercial entities operate in acceptable ways.

What should qualify as trustworthy AI?

As highlighted by our interviewees, confidence in those who design and deploy CP tools is necessary, but it is equally essential to consider the dependability of the tools themselves. At the outset of this paper, we stated that our interest did not lie in debating whether non-human entities such as AI systems can be inherently trustworthy [14, 15]. Drawing on Starke *et al*. [3], we argued that indirect trust in developers and users could be sufficient to justify a trusting stance toward the technology itself.

"In its indirect, weaker sense, trust in AI does not require a fully independent agency of the program itself but rather ties trust to the intentions of its developers or those involved in its quality control, promoting 'indirect trust in the humans related to the technology'. For example, we may trust a sys- tem of medical AI because we trust the people who develop and regulate it. Even in this very limited sense, it may already be plausible to describe a potential attitude towards medical AI as 'trusting'." [3:157].

According to Starke *et al.* [3], if trust is to be treated as fundamental in human–AI interactions, then it must be understood within a three-dimensional framework:

"…the decision to trust an AI- based program depends on features of the trustor (e.g., overall willingness to trust), and the context (e.g., level of risk) in combination with the perceived reliability, competence, and intentions of the program." [3:158].

Within this framework, the trustworthiness of a system is judged by three factors: its reliability—whether it consistently functions across situations and conditions; its competence—whether it delivers accurate and valid outcomes aligned with what it claims to assess; and its (indirect) intentions—namely the potential conflicts of interest of its developers, the degree of transparency, and the representativeness of the training data. Thus, openness and explainability are regarded as critical to

cultivating trust in the system's intentions. While these dimensions collectively inform perceptions of trustworthiness, Starke *et al.* note that not all need to be fulfilled equally for users to consider a system trustworthy.

This tripartite model can also be interpreted through the concept of epistemic trust [28]. In this sense, what is being described are the qualities of epistemically reliable instruments that function properly within the contexts and purposes for which they are intended. The reliability of AI in medicine is reflected in its design and rigorous testing, which ensure the validity of its claims. Schwab argues that producing claims through reliable processes is precisely what is expected in medical practice [28]. Similarly, Wang *et al*. [29] reinforce this perspective in their review, recommending that medical AI should be assessed through randomised controlled trials conducted under strict reporting standards.

The findings presented above indicate that concerns linked to AI's epistemic features—what Starke *et al.* describe as its three dimensions of trust—were consistently emphasized in our study as significant. Interviewees often highlighted reliability and competence, focusing on whether the datasets used for training were representative and whether the outputs of CP tools had been validated against alternative forms of evidence, including clinical judgment.

In line with Starke *et al.* [3], participants also stressed that the development of transparent and explainable AI systems is central to enhancing perceptions of trustworthiness. They further argued that it is necessary to account for both the motivations and intentions of AI designers and developers. However, neither Starke *et al.* nor our interviewees clearly distinguished the relative weight of the trustworthiness of developers' intentions compared with the trustworthiness of the intentions embedded within the systems themselves (i.e., transparency and explainability). Moreover, the precise relationship between a system's intentions and those of its designers remains ambiguous. What was clear, however, is that our interviewees viewed developers' conflicts of interest as a significant obstacle to perceiving AI systems as trustworthy (In relation to this issue, we emphasize that untrustworthy AI systems (model behaviours) do not necessarily stem from malicious intent on the part of designers or developers. Instead, such problems may result from human error—for instance, limited foresight, insufficient effort in ensuring training datasets are representative, or neglecting to incorporate transparency into the system. While we recognize that conflicts of interest may (indirectly) contribute to these shortcomings, it is equally important to note that such detrimental outcomes can also occur unintentionally unless intentional and careful measures are implemented to prevent, assess, and mitigate them.).

How then do reliability, competence, and the system's intentions connect to trust? These features may help mitigate uncertainty, which is central to trust relationships [3, 15], yet we must ask, as Starke *et al.* suggest, whether recognizing an enhanced ability to deliver accurate and reliable outcomes is equivalent to trusting. Put differently, is there a meaningful difference between receiving a reliable, competent, or well-intentioned diagnosis and receiving a trustworthy one? C. Kerasidou *et al.* [13] argue that in the context of commercial AI developers (and by extension AI technologies), fostering reliance is a more fitting aim than cultivating trust. The reason is that reliance can be reinforced through legal and regulatory frameworks, whereas trust is an attitudinal stance that cannot be legislated. Reliance is tied to predictability, while trust depends on the perception of goodwill from the trusted party toward the truster [2, 15]. Kerasidou and colleagues note that regulatory mechanisms that strengthen reliance on AI technologies create the circumstances in which trust might naturally arise as an end in itself rather than as a means to an end. In this way, being able to rely on AI developers or on the technology itself could, over time, foster trust.

Building on this perspective, Graham [30] proposes a model grounded in confidence instead of trust in his work on ethical data-sharing practices. He distinguishes confidence from reliance by suggesting that while reliance rests on predictability, confidence represents "assured reliance," meaning that B doing X is not only predictable but also guaranteed. Although Graham does not fully clarify the line between confidence and reliance [13], he offers specific conditions for establishing confidence-worthy systems in data sharing (which we can apply to AI systems). These include: transparency that is meaningful and verifiable, accountability structures, evidence that data are representative rather than biased, and assurances of a demonstrable social purpose (i.e., serving the public good). For Graham [30], confidence differs from trust in that it does not hinge on goodwill [2, 15]; like reliance [13], it imposes enforceable obligations. Thus, if AI developers fail to

uphold these standards—for instance, if they neglect transparency requirements—sanctions would follow.

Our findings indicate that although good intentions, competence, and reliability can render computational phenotyping (CP) tools more predictable [13, 30] and therefore "confidence-worthy" or reliable, these qualities alone do not make them trustworthy. As many participants emphasized, diagnosis involves more than correct categorization; it also requires acknowledging patients as persons and treating them respectfully [8]. Consequently, interviewees considered CP technology to be an expert, reliable, and competent instrument, but not a replacement for trusted human experts.

In conclusion, the introduction of AI tools into clinical diagnostic settings has sparked debate on the significance of trust [12, 31, 32]. While most agree that patients should trust clinicians who employ AI, some argue that relationships with AI developers or systems should be framed in terms of reliance [13] or confidence [30], rather than trust. It is also worth noting that although transparent AI may appear to embody more trustworthy intentions than opaque systems, transparency and explainability alone cannot generate trust. As Sand *et al.* [33] contend, what matters is how these systems are used responsibly: "focus on forward-looking responsibilities of physicians using such systems." [33: 163]. This involves cultivating skills and virtues such as evaluating AI outputs and inputs, maintaining reflexive awareness of one's expertise in using these tools, tracking their performance over time, and communicating uncertainty to patients. In this sense, Sand *et al.* argue that trustworthiness should not be attributed to AI systems themselves, but rather to the relational dynamic among the tool, the clinician, and the patient. Similar views were also advanced by Lai *et al.* [18] and Winter and Carusi [20], who stressed that continuous collaboration between AI developers and clinical users is crucial for trustworthy AI. Put simply, epistemic trust emerges from relational trust.

Before drawing conclusions, it is necessary to acknowledge the limitations of this research. The study concentrated exclusively on the perspectives of developers and clinicians, most of whom were affiliated with the Minerva Consortium and directly engaged in either creating or applying CP. As such, the findings primarily reflect their perceptions of CP tools and their expectations about how patients and other clinicians might respond once these tools are introduced in clinical practice. Since the majority of participants were drawn from the Minerva Consortium and were involved in supplying data for CP development or in designing the technology itself, it is possible that they had an incentive to portray CP as potentially trustworthy. Conversely, one could argue that, for clinicians in particular, it would not serve their professional interests to suggest that CP systems are—or will be—sufficiently trustworthy to perform diagnoses independently of human expertise, as doing so might imply diminishing their own clinical role [34]. With this in mind, we propose that future research should extend beyond the perspectives of developers and clinical users to include those of patients, whose diagnostic experiences are, or will be, directly shaped by CP technologies.

It is also important to note that although the interviews revealed a broad consistency of responses, participants represented a range of geographical settings—each with distinct healthcare systems and priorities—as well as different areas of expertise, spanning both users and designers of CP. This diversity, far from being a limitation, can be regarded as a strength, since it enabled examination of how a varied group conceptualize the issue of trust in CP from multiple standpoints.

A further consideration is whether the arguments advanced here apply solely to AI applications in rare disease diagnosis. Given the pronounced uncertainty in diagnosing rare conditions, one might ask whether establishing trust in CP-assisted diagnoses is more challenging than in other fields of medicine where AI is deployed, such as digital pathology or radiology. While the heightened uncertainty surrounding rare diseases and the lack of extensive datasets for CP training may indeed affect the requirements for fostering trust in this context, we maintain that the insights from this study have broader relevance for AI use across medicine. Although empirical work on this topic remains limited, the concerns voiced by our participants are consistent with other studies that emphasize AI should function within trusted relationships [18, 19] and that epistemic trust is built through collaborative interactions [20].

Conclusion

Trust lies at the heart of the clinician–patient relationship [35], with patients placing confidence in their clinician's expertise and in their commitment to act in the patient's best interests. Although trust is essential in all diagnostic contexts, it faces particular strain in rare disease cases, where clinicians often encounter gaps in knowledge or limited experience, since such conditions are seldom seen

in everyday practice. CP tools provide an opportunity to strengthen diagnostic outcomes in rare diseases by complementing clinicians' existing diagnostic capabilities. Our findings revealed that both clinical users and AI developers recognize a close connection between relational trust and epistemic trust in relation to the deployment of AI in healthcare. They emphasized that AI systems must be designed to achieve epistemic reliability, while also noting that perceptions of reliability emerge gradually, through repeated use by trusted practitioners. In this respect, clinicians' positive interactions with CP tools will shape greater trust in their diagnostic contributions. Thus, trust in AI systems is both conditional and contingent—dependent on mechanisms for evaluating and proving their reliability, as well as on their consistent and appropriate performance within established clinical relationships. Many of our interviewees further suggested that patients in the future would be inclined to reason, "if it is good enough for my clinician to use it, then it is good enough for me," thereby linking relational and epistemic trust in this setting.

To conclude, our study underscores the need for careful and deliberate efforts to design AI systems that are reliable or confidence-worthy for healthcare. Equally important is the establishment of reliable or confidence-worthy processes—such as RCTs, frameworks of accountability, transparency, and responsibility—that can demonstrate the epistemic trustworthiness of these technologies. With such systems and safeguards in place, AI tools hold the potential to reinforce, rather than undermine, the trusting relationship between clinicians and their patients.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Calnan M, Rowe R. Researching trust relations in health care: conceptual and methodological challenges–introduction. J Health Organ Manag. 2006;20:349–58.
2. Baier A. Trust and antitrust. Ethics. 1986;96:231–60.
3. Starke G, van den Brule R, Elger BS, Haselager P. Intentional machines: a defence of trust in medical artificial intelligence. Bioethics. 2021;36:154–61.
4. Meagher KM, Juengst ET. Henderson GE Grudging trust and the limits of trustworthy biorepository curation. AJOB. 2018;18:23–5.
5. Bauer PC. Conceptualizing trust and trustworthiness project trust research. https://www.researchgate.net/publication/262258778_Conce ptualizing_Trust_and_Trustworthiness 2019. Accessed 20 Mar 2022.
6. O'Neill O. Linking trust to trustworthiness. Int J Philos Stud 2018;26:293–300.
7. Ward P. Trust, reflexivity and dependence: A "social systems theory" analysis in/of medicine. Eur J Soc Qual. 2006;6:143–58.
8. Hallowell N. Encounters with medical professionals: A crisis of trust or matter of respect? Med Health Care Philos. 2008;11:427–37.
9. Robb N, Greenhalgh T. "You have to cover up the words of the doctor": the mediation of trust in interpreted consultations in primary care. J Health Organ Manag. 2006;20:434–55.
10. Topol E. Deep Medicine: How artificial intelligence can make healthcare human again. New York: Basic Books; 2019.
11. AI HLEG. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019.
12. Kerasidou A. Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. J Oral Biol Craniofac Res. 2021. https://doi.org/ 10.1016/j.jobcr.2021.09.004.
13. Kerasidou CX, Kerasidou A, Buscher M, Wilkinson S. Before and beyond trust: reliance in medical AI. J Med Ethics. 2022 Nov;48(11):852-856. doi: 10.1136/medethics-2020-107095. Epub 2021 Aug 23. PMID: 34426519; PMCID: PMC9626908.
14. Metzinger T. Ethics washing made in Europe. Der Tagesspiegel. https:// www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in- europe/24195496.html 2019. Accessed 17 Feb 2021.
15. Jones K. Trust as an affective attitude. Ethics. 1996;107(1):4–25.
16. Kerasidou AT. Institutions in global health research collaborations. In: Laurie G, Dove E, Ganguli-Mitra A, McMillan C, Postan E, Sethi N et al, editors. The Cambridge handbook of health research regulation.

Cambridge: Cambridge University Press: Cambridge Law Handbooks; 2021. pp. 81–9.
17. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. J Am Med Inform Assoc. 2020;27(4):592–600. https://doi.org/10.1093/jamia/ocz229.
18. Laï MC, Brian M, Mamzer MF. Perceptions of artificial intelligence in healthcare: Findings from a qualitative survey study among actors in France. J Transl Med. 2020;18:14. https://doi.org/10.1186/s12967-019-02204-y.
19. Hui CY, McKinstry B, Fulton O, Buchner M, Pinnock H. Patients' and clinicians' perceived trust in internet-of-things systems to support asthma self-management: Qualitative interview study. JMIR Mhealth Uhealth. 2021;9(7):e24127. https://doi.org/10.2196/24127.
20. Winter P, Carusi A-M.'If you're going to trust the machine, then that Trust has got to be based on something': Validation and the co-constitution of trust in developing artificial intelligence (AI) for the early diagnosis of pulmonary hypertension (PH). Science & Technology Studies (2022).
21. Ferry Q, Steinberg J, Webber C, FitzPatrick DR, Ponting CP, Zisserman A, Nellåker C. Diagnostically relevant facial gestalt information from ordinary photos. Elife. 2014. https://doi.org/10.7554/eLife.02020.
22. Hsieh TC, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet. 2022 Mar;54(3):349-357. doi: 10.1038/s41588-021-01010-x. Epub 2022 Feb 10. PMID: 35145301; PMCID: PMC9272356.
23. Nellåker C, Alkuraya FS. The Minerva Consortium Enabling global clinical collaborations on identifiable patient data. The Minerva Initiative Frontiers in Genetics. 2019. https://doi.org/10.3389/fgene.2019.00611.
24. Hallowell N, Parker M, Nelllåker C. Big data phenotyping in rare dis- eases: Some ethical issues. Genet Med. 2018. https://doi.org/10.1038/ s41436-018-0067-8.
25. Strauss A, Corbin J. Basics of qualitiatve research techniques and procedures for developing grounded theory. London: Sage; 1990.
26. Giddens A. The consequences of modernity. Cambridge: Polity Press; 1990.
27. Pearson SD, Raeke LH. Patients' trust in physicians: Many theories, few measures, and little data. J Gen Intern Med. 2000;15(7):509–13.
28. Schwab A. Epistemic trust, epistemic responsibility and medical practice. J Med Philos. 2008;33:302–20.
29. Wang J, Wu S, Guo Q. Investigation and evaluation of randomized con- trolled trials for interventions involving artificial intelligence. Intell Med. 2021;1:61–9.
30. Graham M. Data for sale: Trust, confidence and sharing health data with commercial companies. J Med Ethics. 2021. https://doi.org/10.1136/medethics-2021-107464.
31. Future A. Ethical, social, and political challenges of artificial intelligence in health. London: Future Advocacy; 2018.
32. Nuffield Council of Bioethics. Bioethics briefing note artificial intelligence (AI) in healthcare and research. London: Nuffield Council of Bioethics; 2018.
33. Sand M, Duran JM, Jongsma KR. Responsibility beyond design: Physicians' requirements for ethical medical AI. Bioethics. 2021. https://doi.org/10.1111/bioe.12887.
34. Susskind D, Susskind R. The future of the professions: How technology will transform the work of human experts. Oxford: Oxford Universiry Press; 2015.
35. O'Neill O. Autonomy and trust in bioethics. Cambridge: Cambridge Universiry Press; 2002.