TRUST, TECH, AND TREATMENT: HOW MIGHT NEW CLINICAL AI TECHNOLOGIES SHIFT THE FIDUCIARY RELATIONSHIP IN THE FUTURE OF WORK?

Discussion Paper

Marina Salis, Jennifer Bell, Paula Rowland, Mattea Welch, and Andrew Hope
Jennifer.Bell2@uhn.ca

Cite as:
Salis, M., Bell, J.A.H., Rowland, P., Welch, M., & Hope, A. Trust, Tech, and Treatment: How might new clinical AI Technologies shift the fiduciary relationship in the future of work? [Discussion Paper]. ETCHWork. Sponsored by AMS Healthcare

Disclaimer: This paper represents the personal views and interpretations of the authors and does not necessarily reflect the positions or opinions of any organization, institution, or individual with which the authors are affiliated.

Trust, Tech, and Treatment: How might new clinical AI Technologies shift the fiduciary relationship in the future of work?

By: Marina S. Salis, Jennifer Bell, Paula Rowland, Mattea Welch, and Andrew Hope

The field of bioethics has typically characterized the clinician-patient relationship as fiduciary, predicated on the notion of *trust,* which allows both actors to navigate complex information and make decisions in the interests of good patient care. With technology-driven trends marshaling in a new era, future-of-work (FoW) literature proposes an interesting change to the traditional therapeutic interaction brought about by AI and other new technologies: rather than a bilateral relationship between clinician and patient, technologies are introduced as a *third* actor, described as a *triadic* or *clinician-patient-technology* relationship. This has consequences for not only the arrangement of the therapeutic relationship, but more interestingly to the configuration and perceived moral import of *trust*.

This discussion paper engages critically with the bioethical and FoW literature to explore the notion of trust in the context of the triadic relationship, and will seek to advance discussion of the following questions:

1. Given the potential technological disruption of morally salient relationships and decision-making processes, how might AI and new technology transform the fiduciary relationship in the future of work?

2. In light of challenges in applying the traditional notion and moral requirements of trust to a non-human agent, in what ways, if any, may these technologies engender a sense of "trust" with their collaborators, and how essential is trust to the clinical adoption of AI and new technologies?

3. What are the ethical sensitivities entangled with technologies and trust, and how may they best be managed in the future of healthcare work?

The Notion of Trust in Care Relationships

Philosopher Annette Baier provides a moral understanding of trust that has traditionally undergirded care relations, whereby trust, as Baier states simply, is "reliance on another's good will." This reliance to act in service of another presumes an unequal power relation that exposes a dependency and vulnerability, whereby one could be betrayed or harmed by ill will or lack of good will, and not merely let down [3]. Baier identifies that the need for trust in this situation is not always voluntary. This is especially the case in health care, where illness besets health and patients need to rely on others for care.

In bioethics discourse, the relationship between clinician and patient is recognized as *fiduciary,* possessing a deeper moral quality predicated on a notion of *trust* and a *duty of loyalty* [7]. In the therapeutic context, both clinician and patient often operate under conditions of uncertainty [21]. In one respect, the practice of medicine is not an exact science: while clinicians are equipped with an expert set of skills and knowledge, the diagnoses, prognoses, treatment formulation and treatment response are complex, imperfect processes, requiring reliance on scientific

evidence, practical reasoning and clinical judgment. Furthermore, these therapeutic interactions are intimate, but in hospital context often occur between two ill-acquainted parties. Healthcare providers, as clinical experts, and patients, as experts on the self, come together to exchange knowledge not only to seek a solution to a clinical problem or need, but to do so in alignment with the wishes or best interests of an individual patient [11].

Additionally, this exchange of knowledge between clinician and patient requires bi-directionality, but is often encumbered by a power dynamic that privileges the clinician. This is particularly the case for doctors, historically a learned profession with elitist cultural and legal status conferring them power to provide service for the public good [26]. Patients, occupying a vulnerable position due to illness and other compounding factors, rely on doctors and other healthcare providers as gatekeepers to complex clinical knowledge and resources, to receive response to their health-related needs with potential impacts beyond just the physical person at stake.

Taken together, trust, then, acts as an ultimate foundation of the therapeutic relationship in which patients entrust their vulnerability, health and well-being, and important values to an individual with the knowledge, skills, and role-related duties to respond to the best of their abilities for the ultimate goal of shared decision-making and quality patient care [15]

Shifting Trust Dynamics in the Future of Work: A Triad of Trust?

Future of Work (FoW) literature proposes an interesting change to the traditional understanding of the therapeutic relationship with the application and adoption of AI and clinical technologies: rather than a bilateral relationship between clinician and patient, these technologies are introduced as a *third* actor, described as a *triadic* or clinician-patient-AI relationship [17]. More than a change in dynamic, this proposal would impact the way in which trust is conventionally engendered in the therapeutic interaction.

Previous bioethics research and literature identify factors such as competence, compassion, privacy and confidentiality, reliability, dependability, and communication as important for engendering trust in one's clinician [15]. However, the extent to which these traditional dimensions of trust are maintained in this triadic notion, and whether new dimensions or characteristics are raised as a result of increasing reliance on new technologies is worthy of additional attention given its ethical relevance.

For example, technologies such as computerized clinical decision support systems (CDSS) are engaged in morally salient tasks and decision-making processes that have implications for care planning and outcomes [23]. AI-based CDSS can simplify the complex task of integrating mass amounts of information required for prognostication, including probabilities of future events and time intervals (e.g. death and survival predictions), as well as quality of life trajectories (i.e. ability to perform activities of daily living, frailty, and cognitive capacity) [11]. When CDSS provides, for example, alerts about the need for patient testing or medication safety, it inserts knowledge relevant to patient care that clinicians are forced to contend with, even if that means silencing the alert in favour of clinical judgement. Much like in clinician-patient interactions, the kinds of information that such technologies can offer, how healthcare providers use this information to inform their clinical judgment, and the way this information is communicated to a patient for shared decision-making can impact care recommendations and treatment decisions [23]. These decisions are imbued with ethical and moral significance, not only related to the consequences of the decision (i.e., the treatment/care recommendation) but also the decision-making process to the extent they impact the fiduciary relationship [23]. For example, the moral salience of technologies

informing life or death and the ethical principles, values, and criteria justifying decision making feature prominently. Clinicians and patients may raise concern about whether it is ever appropriate to place trust in AI or other technologies to ethically perform morally relevant tasks, or at the very least, they may question what values are underlying the decision making.

An auxiliary challenge is the difficulty in translating the components of human-human trust to human-technology trust. While we can hold *individuals* accountable for the ethics of their choices and actions, it is not as straightforward as to how we may expect, or whether it is possible, to hold AI and clinical technologies accountable for moral deliberation and consequences absent the same level of human agency [20, 22]. For example, a foundational understanding of trust according to Baier assumes sapience in terms of moral intentionality to properly theorize trust relations [3]. AI *may*[1] take on a shallow sense of moral accountability to the extent that the clinician or patient has direct control or oversight over the technology, thus extending the human components of trust to the technology. However, how this maps onto a human versus technological agent, and to what degree that impacts the respective assignments of responsibility remains a challenge for healthcare leaders and regulators.

Potentially addressing these considerations, the FoW literature expands the notion of trust to include both a functional construct, where the AI/technology/human acts as intended, as well as an interpersonal notion which refers to the relationship between the AI, the clinician, and the patient [7, 8, 9].

Factors of Influence: Conditions of Trust in the Triadic Relationship

Compared to the aforementioned traditional understanding, the notion of trust within the FoW triadic relationship is much more complex and nuanced than how trust is characterized in the dyadic relationship. One way we may begin to understand the difference is to examine what factors and characteristics have been identified as influencing the uptake of these technologies within the triadic therapeutic relationship. FoW literature identifies many of these features, which we group into the categories “functional” trust and “interpersonal” trust based on the ways in which they are defined or described.[2] By charting how trust is conceptualized in the literature and positioned in relation to the clinician and the patient in the emerging context of new technologies, the nuances, moral relevance, and impact to the therapeutic interaction can be exposed.

Functional Trust

Some factors discussed in the literature relate to a concept of *functional* trust or the technical operations of a human, AI or other clinical technology in the triadic therapeutic interaction. These factors are often positioned as amoral, as they relate primarily to the internal workings of the agent, contributing to the understanding or

[1] The use of italics is to identify that while this paper does propose some common ethical/moral challenges in the use of AI such as questions of moral accountability, we resist engaging in phenomenological debate surrounding moral accountability and agency for the sake of scope, and operate off the taken for granted assumption that these technologies do *not* possess the same level of agency required for an equivalent account of moral responsibility. Rather, we propose a tentative, but non-argumentative alternative interpretation.

[2] It is important to note that the terms take on different meanings and interpretations within and across the literature. For example, in some literature, the term agency is used interchangeably with dependence, though while some use the term functionally, others use it interpersonally. Factors were grouped by description and purpose, and thereafter labeled with the most relatable heading.

evaluation of what the agent[3] is set out to do, how it does it, and whether it produces its intended outcomes [13]. The table below outlines the various factors perceived to fall under a functional notion of trust.

While some of these terms are traditionally used in reference to trust of *technology*, the literature applies them to both human and technological agents, though the ways in which they are interpreted, applied, and exemplified vary. For example, the factor of dependence generally describes how much input or assistance is needed to produce a particular output, such as a prognosis. For human agents such as clinicians, dependence may involve consulting with expert

Table 1. Functional Trust

Factor Name	Description(s)
Dependence	What degree of exterior input, assistance or additional oversight is required to ensure appropriate functioning [16, 24]? What is the level of responsibility delegated to it [17, 23, 24]?
Accuracy	To what degree does the estimation or calculation corroborate with the actual outcome [17]?
Fault Susceptibility	How often or likely, if at all, is the [human/technological] agent susceptible to mistake, fault, or failure? [16, 24]
Harm Exposure	Are there any tangible harms to the potential user/recipient or others involved and if so, what are they [16,20]
Interpretability	How easy is it to interpret and understand the operation/process and outputs (e.g. suggestions, recommendations) of the technology/human [16]? How often is it misinterpreted [24]?
Explainability	To what degree is an understanding of the operation/process or reasoning for outputs (e.g. suggestions, recommendations) accessible or transparent [16]?
Robustness	What resources/information/data points are referenced to produce the intended outputs? To what degree do the inputs used accurately reflect the complexity, nuance, or full picture of the context and circumstances impacting a decision [16]?
Maturity	How long has the agent been in practice? What is the level of experience in its use? To what degree is it engrained or a subject of education for practice [24]?
Orientation	Who has access to the outputs and operation of the agent? Who is the intended audience/user/recipient (e.g. a patient) [6]?
Reliability	Can the agent produce its intended outputs predictably [21]?
Calibration	To what degree does the perceived function and abilities map on to the actual function and abilities of the agent [12]?

[3] The term agent, though often used to describe a clinical technology, is used in this literature to describe any actor, human or technology, of the therapeutic encounter (e.g. clinician, patient, technology).

peers in order to produce or validate the requested information. For technological agents, dependence describes how much human input, programming, or oversight is needed to ensure that the outputs are produced appropriately [16, 24].

The significance of these factors and the way they interact to inform human judgment of functional trust may depend on the standard they are being evaluated against or what may be reasonable to expect in a given situation. In evaluating functional trust, one may question how one agent (i.e. human or technology) stands against another to produce the same output. Returning to the example of prognostication, human accuracy is an ongoing challenge [19]. How an agent is expected to perform may inform the degree to which accuracy must be met for trust to obtain within the triadic relationship. A technological agent may be presumed to perform with high accuracy to encourage trust whereas it may be more acceptable for the clinician to express greater uncertainty maintaining the same.

While each characteristic may be considered in isolation, their interplay informs a holistic picture of whether or not the human or technological agent can perform to a satisfactory degree, and whether one can rely on the outputs that a given agent produces. For example, while technological agents such as CDSS are more accurate at activities such as prognostication compared to their human counterparts, they are often far less explainable due to their black box nature [25]. Whether or not one values the ability to understand the inputs and reasoning behind a decision compared to its ability to produce an accurate response may influence the nature of trust one can engender in a human agent (such as their clinician) compared to technology. This may be compounded with factors such as the orientation of the technology to influence what is more or less relevant for a trust relationship: for patient-facing tech, we may be more inclined to value explainability, interpretability, and dependence due to the absence of an expert intermediary to ensure understanding for informed consent. For doctors, accuracy may be more relevant than explainability since the technological output would be one among many sources of information contributing to a holistic decision-making process wherein the clinician bears ultimate responsibility for the recommendation. Contrarily, the degree of explainability is a useful tool to influence how confident doctors feel in knowing when to accept or when to disregard CDSS output. This assessment may ultimately depend upon whether technology is intended to *augment* or *replace* clinician judgement.

Not only may these factors be weighted and balanced against one another to inform a sense of trust with respect to an agent, but moreover, the functional abilities and sense of trust engendered in one form of technology may influence the expectations or notion of functional trust in another agent. Automation bias describes the propensity for individuals to favour the suggestions and recommendations of technological agents over humans, sometimes to a degree to which individuals may ignore errors or contradictory information [24]. Nevertheless, in these situations, it is interesting to consider whether the functional abilities of one may impact the expectation, and inversely the relationship of the other: does the increased accuracy of technology like CDSS in prognostication influence the perception of how a doctor should be in the practice of the same?

Ultimately, the goal of functional trust is to ensure that agents, human or technological, function and perform in the way they are intended, to ensure they operate safely, effectively, and in a way that can be understood by those utilizing and impacted by the agent inputs, outputs, or judgements when engaged in the decision-making process.

Interpersonal Trust

Other factors denoted in the literature take on a slightly different, interpersonal character, referring to the nature of the interaction or relationship between agents within the therapeutic context. Factors of interpersonal trust differ from functional in their relationship to ethics as they go further to speak to both intrinsic and instrumental moral characters and values present in the therapeutic interaction. Rather than focusing straightforwardly on the ability or function of the agents, which may or may not carry ethical consequences and considerations, interpersonal trust evaluates the quality and character of the interaction and decision-making process informed by familiar ethical principles and values. The table below outlines the factors of interpersonal trust described in the literature, whereby the term "agent' again refers to clinician, patient and technology.

Just as with functional trust, the application and interpretation of these factors may differ depending on the type of agent interaction they are describing. For example, the arrangement of accountability may look very different in the context of a clinician-patient interaction, versus a patient-clinician-technology interaction, with the limitation of the latter being that technological agents such as CDSS cannot be said to be responsible for the outcomes in the same way as a human agent [17,23, 25]. Given their limitations, it is likely that these interpersonal factors are identified in human oversight where technology is involved, rather than AI evaluating its interpersonal functions introspectively, as it cannot be assumed that AI understands or can weigh some of these more nuanced, contextual considerations outside of what it is programmed to do [22].

Table 2. Interpersonal Trust

Factor Name	Description(s)
Agency and Power	To what degree does the agent have influence or power over making a moral decision [23]?
Accountability	Who is entrusted to carry out the moral action/decision and who will be held accountable for the risks and benefits of that decision [17, 23, 25]?
Ethical Reasoning	Does the agent make decisions and provide justification based on ethical deliberation and reasoning? Does the agent consider all relevant factors and perspectives necessary for informed choice (i.e. contextual factors, moral values, culture, socioeconomic factors, emotions and nonverbal communication) [13, 16, 23, 24]?
Bias	Are there prejudices or unfair assumptions (including automation bias) made or created by the agent(s) intentionally or unintentionally that may impact inputs, outputs, judgment, or appropriateness of actions [24]?
Justice and Fairness	Are any individuals or groups unfairly/disproportionately advantaged intentionally or unintentionally by the inputs, outputs, or judgements of the agent [24]?
Honesty and Transparency	Are all actors influenced or impacted by the inputs, outputs, and/or judgment of the agent aware and informed of its use and potential impact [16)?
Proportionality	Are the relevant factors appropriately considered, weighted, and balanced in a way that is acceptable and justifiable to those impacted by the inputs, outputs, or judgment of the agent [17]?

The interpretation and moral import of some or all of these factors again interact with one another, as well are weighted and balanced against each other to paint the relationship of trust. The presence of bias, for instance, such as automation bias, may conflict with the quality of ethical reasoning: if an agent, such as a doctor or patient, prima facie favors or is inversely aversive towards technology to a degree that is preconceived or ill-justified, it may adversely degrade the appraisal of the ethical strengths and benefits of each option. This weighting and balancing may also occur *across* factors in that the manifestation of interpersonal factors may also be dependent upon the arrangement of functional characteristics. To illustrate, the ability of an agent, such as CDSS, to perform fulsome ethical reasoning may rely heavily upon factors such as robustness: a technological agent's ability to consider all factors relevant to an ethical decision-making process is dependent on whether it is programmed to take those data points into consideration. For human agents, such as a patient, robustness may require a degree of dependence to exercise power over decisions given aforementioned bi-directionality and often imbalanced knowledge and power dynamic that exists most often privileging the doctor or other healthcare provider [18].

Ultimately, not only do we want to ensure agents perform in the way they are intended, in a way that is predictable and explainable, but moreover, their proximity and relationship with *ethical* decision-making adds the additional concern that agents bolster rather than hinder the ability to deliberate and act on what is right, with robust consideration towards the potential consequences of a decision or lack of a decision. That agents have the capability to directly impact the moral environment underscores the importance of recognizing that agents are not value-neutral, objective machines, independent algorithms, or unbiased reasoners. The urge to trust explicitly recognizes value non-neutrality and objectivity, since there is recognition of an agent's intention (often technically-driven in the case of technology-agents) to act in a good or acceptable way and that there is the possibility of untoward consequences.

Thinking Deeper: The Ethical Stakes

It is important to recognize that despite the very deterministic narrative in the FoW discourse that the adoption of technologies in health care is seemingly inevitable, the ways in which they *ought* to be used remains a poignant piece of conversation. This is especially the case in light of the fact that the application of technologies can have serious impacts on decisions, and consequently on the care that individuals receive [11].

For example, in the context of prognostication towards end of life, predictions surrounding disease trajectory and outcome often can have significant influence on the recommendations or decisions made, and the actions taken for both the care decided upon and received, and quality of life related choices as a result of the prediction for both the doctor and patient [7, 19]. Prognosis, though not without challenges,[4] can and often does inform part of the foundation for what treatment options are reasonable and available to propose, and additionally can inform the analysis of what treatment options *should* be pursued in light of the anticipated outcomes of the relevant options, the costs, and the anticipated benefits [19]. While the inclusion of probabilistic information and risk prediction such as prognosis contributes to the practice of informed consent and truth telling, disclosures of this nature

[4] Prognostication is discussed as a complicated and imperfect skill. Some of the challenges discussed in the literature include the inclusion of prognosis for life expectancy can cause patients to make decisions on the basis of false hope when the communication is optimistic for the circumstance, or false despair in the inverse [8, 19, 20].

especially in the context of high-stakes decisions may be harmful or impede decision-making by altering the patient's perspective of their future and the reasonableness of pursuing treatment or non-treatment [7].

Probabilities, however, only inform *part* of the decision-making process. Rather, the decision-making process requires attentiveness beyond quantifiable variables, reflecting the intersectional quantitative and qualitative clinical, social, contextual, and cultural milieu that inform what is acceptable and justifiable to those most impacted by the inputs, outputs, or judgment of the decision [17, 19, 22, 25]. Factors such as the robustness of the data (for technology-agents) or information (for human-agents), agency and dependence, ethical reasoning, and proportionality play an important role in ensuring the decision-making process reflects the complexity and circumstances surrounding and impacting a decision [16]. As a challenge, technologies are often discussed as being reductionist, utilizing datasets that often ignore circumstance, vulnerability, and specific characteristics introducing an impartiality that may not reflect what holistic and informed decision-making requires [22, 25]. Depending on how much agency is delegated to the technology-agent, the predictions and recommendations generated may not accurately reflect the personalization and accuracy required for such highly sensitive circumstances.

Additional complications may be present in consideration of the requirements of informed consent [6]. In a positive vein, for example, one may consider the possibility of technology as a mechanism to empower patients and thereby shift traditional power dynamics by patients accessing AI directly (e.g., ChatGPT) to inform, better educate or advocate for themselves [10]. Yet, that a patient-oriented technology lacks interpretability and/or explainability may make it difficult for patients to accurately understand the information and apply it to their decision-making process, especially in uncertain, risky, and emotionally fraught circumstances where patients are often disadvantaged in knowledge and power in the therapeutic context [6, 17, 22].The same can be said for doctors and other clinicians tasked with interpreting the outputs of these technologies to inform their recommendations: while predictions may be of better accuracy, a lack of interpretability or explainability, for example, may inhibit the exercising of clinical judgment and reasoning to support treatment recommendations and decisions [6]. In this way, increased access to information does not necessarily equate to increased beneficence or clinician/patient autonomy [10].].

Preparing for the Future of Work: Lessons and Discussion for Healthcare Leaders

Overall, what the above suggests, at minimum, is that the arrangement and impact of trust within the therapeutic relationship is not as straightforward as may be prima facie assumed. Rather, the potential disruption to not only the arrangement of the therapeutic dyad, but to the arrangement, presence, and value of trust that AI and clinical technologies may bring to the clinical space give us pause to ask not only how the introduction of these technologies *will* change, but rather questioning the role that they *ought* to have in the therapeutic encounter. This is in light of the extensive list of considerations and ethical implications, as well as ensuring appropriate accountability for the adequacy, quality, and responsibility for these decisions [25]. Compared to the traditional notion of trust explored in the preceding sections of this paper, the introduction of these technologies into the therapeutic decision-making space transforms the fiduciary relationship in that it not only increases the complexity of the traditional notion and evaluation of trust, but so too requires a closer and more robust ethical analysis of not only the therapeutic decision-making process, but additionally of

decisions of whether or how to integrate these technologies into the therapeutic space.

The resources and preceding discussion offer a preliminary framework to understand the complex intersection of trust, the therapeutic relationship, and current and emerging technologies, and additionally draws upon the literature to elucidate the range of considerations that may inform an appraisal of the presence of trust. Furthermore, the literature uncovers the ways in which the degree of trust present influences the decision to use and/or the role various agents ought to have in the context of the therapeutic relationship. [17, 22, 23] These various factors and considerations act as a starting point for necessary engaged discussion: should, for instance, CDSS should be restricted to the role of assistive clinical decision making embedded under the control of the clinician as another clinical tool, rather than replacing clinician judgement? This may obviate the need for alternative trust models but can lead to other issues related to clinician interactions with AI, such as altered clinician cognition resulting from AI neglect or overreliance. In doing so, however, one removes the possibility of patient empowerment through direct engagement and access to these technologies. The need for greater discourse on the relationship of trust between clinician, patient, and technological agent is becoming more evident as the adoption and application of these technologies in clinical work and research continues to outpace regulation, education, and infrastructure necessary to support their ethical deployment [20]. This is especially pertinent in situations where clinical technologies such as CDSS are used in high-risk scenarios like prognostication, where clinical information and recommendations can and do have implications for morally salient tasks and decisions such as matters of life and death for patients [23].

A key part of this conversation, and beyond the scope of this paper, is the discussion of not only which trust factors identified ought to be considered, but moreover what weights and values *should* be assigned to each of the relevant factors identified to inform decisions of whether and how technologies ought to be brought into the decision-making space. The complexity and moral salience of this task as suggested in this paper, however, calls for careful evaluation and consideration of not only the presence and weight of these factors individually, but moreover, draws attention towards the intersection of these factors for trust and the therapeutic interaction. This also extends to lingering challenges surrounding questions of moral accountability. Ultimately, while we offer a potential foundation for understanding the different ways the components of trust may translate to various agents in the therapeutic interaction, questions still remain about where accountability ought to lie, and what skills, knowledge, and competencies individuals ought to be equipped with to ensure technologies in the decision-making space are appropriately and ethically employed.

A practical task for healthcare leaders and those engaged in FoW discourse resulting from this complex interplay involves the educational, professional, regulatory, and organizational response to define and optimize the emerging clinician-patient-technology relationship and the ethical boundaries these therapeutic interactions operate within [16, 17, 24]. For example, what accountabilities should or should not be delegated to a technological-agent? In light of this decision, how might the accountabilities of a clinician change, and what competencies or skills ought we equip clinicians with in order to optimize interactions with machines in the therapeutic interaction [5, 17, 20]? How might we equip, inform, and empower diverse patient populations around the use and involvement of technologies in decision-making [10]? Most importantly, how can we centre the needs of patients in the application and appraisal of technology given the complexity of the assessment and the necessary, yet fragile nature, of trust in healthcare delivery?

A potential response may be to develop tools, such as a trust or risk matrix, to assist and define how one might analyze and interpret these factors and relationships to ensure their ethical validity and soundness in a context specific way. Not only this, but given the moral implications of these decisions, further consideration may be given towards whether a *triad* is the appropriate model to understand trust within this new dynamic when looking towards the FoW. Ultimately, the task is not just to understand, but to respond with clear action and accountability to questions of how we *ought* to design, deploy, and adopt these technologies in the context of the therapeutic interaction in a way that is consistent with a thoughtful, ethical, and inclusive interpretation of enhanced patient care [1, 24].

References

1. Alexander GL. Issues of Trust and Ethics in Computerized Clinical Decision Support Systems: Nursing Administration Quarterly. 2006 Jan;30(1):21–9.5

2. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res. 2020 Jun 19;22(6):e15154.

3. Baier, Annette. “Trust and Antitrust.” Ethics 96, no. 2 (January 1986): 231–60. https://doi.org/10.1086/292745.

4. Beil M, Proft I, Van Heerden D, Sviri S, Van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. ICMx. 2019 Dec;7(1):70.

5. Chin-Yee B, Upshur R. (2019). The Impact of Artificial Intelligence on Clinical Judgment: A Briefing Document. Toronto, ON, AMS Healthcare. Click here to enter text.https://www.ams-inc.on.ca/wp-content/uploads/2020/02/The-Impact-of-AI-on-clinical-judgement.pdfhttps://www.ams-inc.on.ca/wp-content/uploads/2020/02/The-Impact-of-AI-on-clinical-judgement.pdf

6. Dalton-Brown S. The Ethics of Medical AI and the Physician-Patient Relationship. Camb Q Healthc Ethics. 2020 Jan;29(1):115–21.

7. Ferrand, Amaryllis, Jelena Poleksic, and Eric Racine. “Factors Influencing Physician Prognosis: A Scoping.” MDM Policy & Practice 7, no. 2 (December 22, 2022): 23814683221145160. https://doi.org/10.1177/23814683221145158.

8. Ferrario A, Loi M, Viganò E. Trust does not need to be human: it is possible to trust medical AI. J Med Ethics. 2021 Jun;47(6):437–8.

9. Hatherley JJ. Limits of trust in medical AI. J Med Ethics. 2020 Jul;46(7):478–81.

10. Hryciw BN, Fortin Z, Ghossein J, Kyeremanteng K. Doctor-patient interactions in the age of AI: navigating innovation and expertise. Front Med. 2023 Aug 30;10:1241508.

11. Lorenzini G, Arbelaez Ossa L, Shaw DM, Elger BS. Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making. Bioethics. 2023 Jun;37(5):424–9.

12. Madhavan P, Wiegmann DA. Similarities and differences between human–human and human–automation trust: an integrative review. Theoretical Issues in Ergonomics Science. 2007 Jul;8(4):277–301.

13. Mainz JT. Medical AI: is trust really the issue? J Med Ethics. 2023 Jul 26;jme-2023-109414.

14. Pearson SD, Raeke LH. Patients' trust in physicians: many theories, few measures, and little data. J Gen Intern Med. 2000 Jul;15(7):509-13. doi: 10.1046/j.1525-1497.2000.11002.x. PMID: 10940139; PMCID: PMC1495476.

15. Pellegrino, Edmund D. “Toward a Reconstruction of Medical Morality.” The American Journal of Bioethics 6, no. 2 (March 2006): 65–71. https://doi.org/10.1080/15265160500508601.

16. Radclyffe C, Ribeiro M, Wortham RH. The assessment list for trustworthy artificial intelligence: A review and recommendations. Front Artif Intell. 2023 Mar 9;6:1020592.

17. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak. 2023 Apr 20;23(1):73.

18. Sherwin, Susan. No Longer Patient: Feminist Ethics and Health Care. Philadelphia: Temple University Press, 1992.

19. Smith, Alexander K., and Paul Glare. Ethical Issues in Prognosis and Prognostication. Edited by Stuart J. Youngner and Robert M. Arnold. Vol. 1. Oxford University Press, 2014. https://doi.org/10.1093/oxfordhb/9780199974412.013.9.

20. Smith H, Downer J, Ives J. Clinicians and AI use: where is the professional guidance? J Med Ethics. 2023 Aug 22;jme-2022-108831.

21. Starke G, Ienca M. Misplaced Trust and Distrust: How Not to Engage with Medical Artificial Intelligence. Camb Q Healthc Ethics. 2022 Oct 20;1–10

***Acknowledgements**: Special thanks to our supporters at The Institute for Education Research (TIER) at UHN and AMS Healthcare.*