Role of trust in AI-driven healthcare systems: Discussion from the perspective of patient safety

Onur Asan[1] and Mehmet Bilal Unver [2]

[1] Stevens Institute of Technology, Hoboken, USA

[2] University of Hertfordshire, Hatfield, UK

oasan@stevens.edu, m.unver@herts.ac.uk

Abstract: In the field of healthcare, enhancing patient safety depends on several factors (e.g., regulation, technology, care quality, physical environment, human factors) that are interconnected (Choudhury & Asan 2020). Artificial Intelligence (AI), along with an increasing realm of use, functions as a component of the overall healthcare system from a multi-agent systems viewpoint (Choudhury, Asan & Mansouri, 2020). Far from a stand-alone agent, AI cannot be held liable for the flawed decisions in healthcare. Also, AI does not have the capacity to be trusted according to the most prevalent definitions of trust because it does not possess emotive states or cannot be held responsible for their actions (Ryan, 2020). A positive experience of AI reliance come to be indicative of 'trustworthiness' rather than 'trust', implying further consequences related to the patient safety. From a multi-agent systems viewpoint, 'trust' requires all the environmental, psychological and technical conditions being responsive to patient safety. It is fertilized for the overall system in which 'responsibility', 'accountability', 'privacy', 'transparency; and 'fairness' need to be secured for all the parties involved in AI-driven healthcare, given the ethical and legal concerns and their threat to the trust.

Key Words: Trust, Artificial Intelligence, ethical and legal concerns, healthcare systems, patient safety, transparency, responsibility, accountability, privacy.

INTRODUCTION

Artificial Intelligence (AI) has an increasing realm of use given ever enhancing AI algorithms used in healthcare sector, e.g., for diagnosis of diseases, drug development, personalized medicine, and patient care monitoring (Choudhry & Asan 2020; Guan et al., 2019; Liu et al., 2019; Sahli et al., 2019; Ekins et al., 2019; Banerjee et al., 2019; Bahl et al., 2018). AI-driven healthcare thus promises innovative services and better solutions in the field of patient safety. Enhancing patient safety depends on several factors (e.g., regulation, technology, care quality, physical environment, human factors) that are inter-connected (Choudhury & Asan, 2020). This multi-agent systems perspective also draws the interdependencies between human and non-human factors involved in a healthcare system.

Along the way of transformation driven by AI, 'trust' between AI users (e.g., clinicians) and the AI software arises as an issue that needs to be examined from the perspective of patient safety. This is compelling given the emerging ethical and legal concerns such as opacity (black-box problem) or bias and discrimination for using historical data sets for clinical purposes (Maddox, Rumsfeld, & Payne, 2019; Hashimoto et al., 2018; AlHogail, 2018). To these, accompanying issues of privacy and absent informed consent need to be added, given the increasing tension with these.

In this context, there remains the question(s) of how to define 'trust' and enhance trust relationships in healthcare against the ethical and legal concerns. While how to interpret the role of AI in healthcare needs to be answered in this context, distinction between trust, trustworthiness and reliance on AI also deserves a touch upon. Overall, this study aims to explore trust and elaborate its role within the AI-driven healthcare in view of the ethical and legal concerns, from the perspective of patient safety.

Overall, it is established desirable outcomes for patient safety can be gained from the reliance on AI, which should not be isolated from but needs to be integrated with the other components of a healthcare system. Far from a stand-alone agent, AI cannot be held liable for the flawed decisions in healthcare. Also, AI does not have the capacity to be trusted according to the most prevalent definitions of 'trust' because it does not possess emotive states or cannot be held responsible for their actions. A positive experience of AI reliance needs to be considered indicative of 'trustworthiness' rather than 'trust', for the former focuses on the traits, whereas the latter relationship.

This very fact implies further consequences related to the patient safety. First and foremost, 'trust', built on a relationship, requires all the environmental, psychological and technological conditions being responsive to patient safety in the healthcare context. Second, this approach requires AI being considered as a (sub)component within the overall healthcare system and fits well to the multi-agent systems perspective. Third, for establishment of trust within a healthcare system, multiple actors' responsibilities need be clarified as to how to use AI and how far to rely on it. Furthermore, accountability, transparency, privacy and fairness need to be secured for using AI, given the ethical and legal concerns and their threat to the trust.

DISCUSSION

Role of AI in healthcare

AI is an umbrella term denoting ability of a computer system to perform tasks commonly associated with human mind such as reasoning, generalizing, problem solving or learning. According to the AI Act of the European Union (EU), AI system means "software that is developed with one or more of the techniques and approaches listed in Annex I [machine learning approaches, logic and knowledge based approaches, and Statistical approaches] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with (European Commission, 2021, Art 1(3)).

AI has potential to assist clinicians in making better diagnoses (Bahl et al., 2018; Guan et al., 2019; Liu et al., 2019) and has contributed to the fields of drug development (Chen et al., 2018; Sahli et al., 2019; Ekins et al., 2019), personalized medicine, and patient care monitoring (Jiang et al. (2017); Banerjee et al. (2019); Ciervo et al. (2019); Ronquillo et al. (2018)]. AI has also been embedded in electronic health record (EHR) systems to identify, assess, and mitigate threats to patient safety (Dalal et al., 2019). The integration of AI into the healthcare system is not only changing dynamics such as the role of healthcare providers but is also creating new potential to improve patient safety outcomes (Siau & Wang, 2018).

However, many of these studies are centered around AI development and performance and there is not much scholarly work reviewing the role and impact of AI used in patient safety in view of the trust relationships and accompanying ethical/legal concerns.

Bigger picture from the perspective of patient safety

Patient safety is a healthcare discipline that emerged with the evolving complexity in healthcare systems and the resulting rise of patient harm in healthcare facilities (WHO, 2019). Sustainable patient safety requires establishment of safety culture based on the values and beliefs systemically followed to ensure the physical, social, and emotional well-being of the patients within an organization (Choudhury and Asan, 2020). Patient safety is prone to many human and technological factors including lack of nursing care (Tuffrey et al., 2013), misdiagnosis, false alarms, late treatment, poor communication, excessive clinical workload, and misinterpretation of health information (Danysz et al., 2019). Out of all these factors deterring patient safety, the application of AI has been extensively focused on addressing errors caused due to (a) misdiagnosis, (b) false alarms, and (c) late treatment (Choudhury & Asan, 2020).

Conventionally, each subsystem in a healthcare system has its own route to achieve the fundamental goal of maintaining quality of care or patient safety. From a holistic systems perspective, the complexity and behavior of a decision support system in the field of healthcare can be described by 3 characteristics: (i) interdependencies, (ii) interactions and (iii) inter-relationships within the system (Choudhury, Asan & Mansouri, 2020).

As all components of a healthcare system including AI are inter-connected, subsystems within a healthcare organization interact with each other enabling functionality of the whole system. This approach, also reminiscent of the "configuration" concept of the SEIPS 2.0 model (Holden et al., 2013), highlights the dynamic, hierarchical, and interactive properties of socio-technical systems and thus can assist in understanding the role of AI in patient safety outcomes. Representing one of the models in the human factors and ergonomics (HFE) discipline, SEIPS 2.0 acknowledges that AI technology is used in a complex, dynamic socio-technical work system which is influenced not only by the people but also by the internal (tasks/ technologies/ organization/ physical) and external environments (Holden et al., 2013). SEIPS 2.0 also has an explicit focus on patient safety.

Crucially, all subsystems (patients, providers, payers, and policy makers) use AI or other technology rendering their respective decisions. AI thus represents a component of the healthcare systems from a holistic multi-agent systems viewpoint. In this context, desirable outcomes for patient safety can be gained from the reliance on AI - which should not be isolated from but needs to be integrated with the other components. This is especially compelling for the patient safety, not least for the functionality of the whole system but also given the need to prevent and reduce risks, errors and harm that would occur to patients during provision of health care.

Ethical and legal concerns

Using AI in healthcare services would bring out some potential challenges based on ethical and legal concerns. While the ethics is an interdisciplinary field of study governing the accumulation and interplay of moral principles (Siau & Wang, 2020), law is a discipline influenced by ethics if not concurring in all terms. The former denotes less normative and non-binding principles whereas the latter most often means binding requirements over the entities who assume responsibility under law.

Ethical concerns arise in relation to using AI, as acknowledged by many (Roseman & Zhang, 2021; Gerke, Minssen, & Cohen, 2020; Mirbabaie et al., 2021; Safdar et al., 2020; Rigby et al., 2019; Schonberger et al., 2019). To address such concerns, ethical guidelines and recommendations are growing across the globe, mostly incorporating voluntary commitments and/or principles. Below are analyzed the most common ethical concerns encountered in the field of healthcare, while the analysis of legal concerns is largely left out for their comprehensive ambit.

Informed consent

AI health apps and chatbots are increasingly used, ranging from diet guidance to health assessments to the help to improve medication adherence and analysis of data collected by wearable sensors (Gerke, Minssen & Cohen, 2020). For data processing, there needs to be an 'informed consent' given by the data subjects (individuals) as required by the data protection laws. According to the Article 4(11) of the EU's General Data Protection Regulation (GDPR), 'consent' means *"any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her"*. In the case of AI-driven healthcare, it is unknown whether or to what extent patients -sometimes in the form of research participants - are specifically informed about the aims of the data controllers for processing of their data. The creation of large cohorts of deeply phenotyped participants raises doubts about the huge amounts of information put in the hands of the governments or the healthcare organizations that stipulate agreements with the former to collect and analyse data from millions of citizens (Blasimme & Vayena, 2020).

Privacy

Privacy needs to be considered under both law and ethics, including but not limited to the issue of informed consent. AI and ML rely on large datasets, with various types of data that can include information about disease risks, lifestyle, mental health, family situation, sexual orientation and other sensitive data (Rosemann & Zhang, 2021). As implied above, in the absence of informed consent, there is a risk that such data are shared with third parties on an unjustifiable ground with the original aim of data collection and processing. A similar risk can also take place when the processing is done more than necessary although based on an informed consent.

As large amounts of training materials for ML applications are gathered from multiple and diverse sources (e.g., medical and insurance records, pharmaceutical data, genetic data, and social media), it becomes easier to trace that data to patient referents thereby breaching privacy intentionally or unintentionally (Ford, Price & Nicholson, 2016). Privacy, being a stand-alone right under human rights laws (e.g., European Convention on Human Rights), could be subject-matter of an infringement itself. Notwithstanding, there might be further consequences such as psychological and reputational harms arising out of privacy breaches on part of the patients (Esmaeilzadeh, 2020).

From a broader point of view, protection of patients' privacy and personal data (e.g., against mis or unauthorized use) is of paramount importance for avoidance of not only any human rights violation but also potential further consequences (e.g., anxiety).

Bias and discrimination

ML bears some risks of bias and discrimination, for the historical and/or unrepresentative data being used or the programmers' inherent preferences, the latter being coined with 'cognitive bias' problem. In case an algorithm is trained on data that are biased or reflective of unjust structural inequalities of gender, race or other sensitive attributes, it may 'learn' to discriminate using those attributes (or proxies for them). ML systems identify proxies for personal characteristics based on the patterns generated out of the datasets and their interpretation. As the co-relations rather than actual causations govern this process, there is a risk for the individuals (patients) to be treated unfairly.

The so-called unfair outcomes might reflect on unrepresentative data. For example, imagine an AI-based clinical decision support (CDS) software that helps clinicians to find the best treatment for patients with skin cancer. Let's say the algorithm be predominantly trained on Caucasian patients, and in that case the AI software will likely give unfair outcomes based on the unrepresentative training data which were underinclusive and not representative of other subpopulations such as African American (Gerke, Minssen & Cohen, 2020). Not only underrepresentation of certain ethnicities but also social inequalities may arise when underserved populations are not well factored into the AI software. For instance, if poor or less educated people have performed worse after certain health interventions (due to poor access to care, working schedules, etc.), an algorithm can determine that people with these characteristics will always perform worse and recommend that they are not offered the intervention in the first place (Blasimme & Vayena, 2020).

Opacity (black-box problem)

Black-box medicine promises substantial benefits to the healthcare systems (e.g., regarding diagnostics, personalized treatment, image analysis) (Ford, Price & Nicholson, 2016) although not visible to their programmers and not explainable for the hidden layers. In fact, an adaptive ML algorithm changes its behavior using a definitive learning process without requiring any manual input and might generate different outputs each time a given set of inputs is received due to learning and updating (Asan, Bayrak & Choudhury, 2020).

The US Food and Drug Administration (FDA) categorizes Software into three classes: (a) Software as a Medical Device (SaMD), (b) software in a medical device, and (c) software used in the manufacture or maintenance of a medical device. FDA defines SaMD as "… AI/ML-based Software, when intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions, are medical devices under the FD&C Act and called Software as a Medical Device" (FDA, 2019). SaMD ranges from smartphone applications to view radiologic images for diagnostic purposes to Computer-Aided Detection software to post-processing of images to detect breast cancer (FDA, 2017). FDA has approved several AI-based SaMDs with “locked” algorithms that generate the same result each time for the same input; these algorithms are adaptable but require a manual process for the updates (FDA, 2019). Despite the lack of understanding of AI algorithms, recently, several algorithms have earned regulatory approval for clinical use, and the barricade for entry of novel advanced algorithms has been low (Asan, Bayrak & Choudhury, 2020).

Approval of black-box algorithms would however risk long-standing medical standards (standard of care) being eroded unless pre-marketing processes for SaMD devices are carefully designed and monitored where necessary. In this regard, FDA’s recently published ‘Patient-Centered Approach’ can be deemed as a significant step forward, given the Agency’s emphasis on manufacturers’ ensuring transparency to users (e.g., clinicians) about the functioning of SaMD devices to ensure that the users understand the benefits, risks, and limitations of these devices (FDA, 2021). Notwithstanding, how this approach will be translated to practice remains to be seen and transparency would require ‘systemic oversight’ to cope with such challenges (Blasimme & Vayena, 2020).

Related to opacity, another novel problem concerning human self-determination arguably arises as over-reliance on AI potentially contrasts human dignity and autonomy. Some AI ethicists raise this problem referring to the tension between the user’ autonomy (including right to privacy) and usage for common good. (Balasescu, 2021). Overall, it is vital to guide the implementation of AI by defining both ethical principles and legal obligations towards patients, including clear norms on issues of responsibility, liability and accountability, as they can enable fairness and transparency from an overall understanding.

Not delving into legal concerns, here we simply refer to the most remarkable legal challenges in healthcare. Such challenges, as widely acknowledged, include (1) safety and effectiveness, (2) liability, (3) data protection and privacy, and (4) security (Gerke, Minssen and Cohen, 2021; Ross & Metnick, 2019). As a fully-fledged examination of legal concerns would require a comprehensive analysis going beyond the remit of this paper, we just note that there are some areas of intersections between law and ethics such as privacy, and some novel problems such as quality and (un)representativeness of training data require appropriately designed common response(s).

Trust in AI-driven healthcare

According to Rousseau et al (1998, 395) ‘trust’ is “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another”. All definitions of trust assume the presence of some form of positive expectation regarding the intentions and behavior of the object of trust (Rousseau et al., 1998). The three most commonly cited elements of trust are ‘perceived competence’, ‘perceived benevolence’ and ‘perceived integrity’ (also sometimes called honesty) (Meijer and Grimmelikhuijsen, 2020, 54). Having wide-ranging dimensions (e.g., from political legitimisation to theories regarding social norms and personal relations), ‘trust’ represents one of the central thrusts for the society and its development (Misztal, 1992) and has been described as the lubricant of social interactions (Arrow, 1974). Although being an interpersonal concept and woven with perceptions, ‘trust’ is also echoed with AI and there is a growing body of literature on how to enhance trust in AI (Glikson & Woolley, 2020).

Likewise, it is expected by the policy makers AI systems be designed as trustworthy, ethical and human-centric (European Commission, 2020). Although being conflated, ‘trustworthiness’ marks a different concept comparing to the ‘trust’ which is more situational and relational. The former does not necessarily lead to the latter. This is perhaps best illustrated by the situation where several trustworthy alternatives exist: In that case, the trusting actor (trustor) may choose to engage in a relationship based on different grounds than trustworthiness alone (e.g., an application’s purchase price, or intuitive user interface) (Gille, Jobin & Ienca, 2020). Trust is situation-specific, and can be described with the following parameters: “A trustor A that trusts (judges the trustworthiness of) a trustee B with regard to some behavior X in context Y at time t” (Sharan & Romano, 2020, 2). Nevertheless, the placement of trust in someone often requires a belief about their trustworthiness, and this notion seems to portray the EU’s Ethics Guidelines for Trustworthy AI, where it is prescribed trust “concerns not only the technology’s inherent properties, but also the qualities of the socio-technical systems involving AI applications” (IHLEG, 2019, 5).

This approach can be criticized for the overemphasized role for the AI itself since the definitional elements of ‘trust’ cannot be found in an AI system, except for the competence. For its very nature (i.e., lack of motivation and responsibility), AI cannot be ‘benevolent’ or ‘honest’. AI cannot be something that has the capacity to be trusted according to the most prevalent definitions of trust because it does not possess emotive states or cannot be held responsible for their actions (Ryan, 2020). One’s freedom would enable their assuming responsibility, both ethically and legally; however, this is not relevant to the AI. Under the current legal regimes, it is a common rule that clinicians are responsible for the decisions concerning diagnosis or treatment of patients, and liable for the ensuing harms (Gerke, Minssen & Cohen, 2020).

This very fact disproves ‘trust in AI’ as a concept and implies further consequences in relation to the patient safety. Both the EU’s approach and the bigger picture drawn above, signifies trust needs to be considered from a holistic multi-agent systems perspective. Trust, as to be built on a multidimensional relationship, requires all the environmental, psychological and technological conditions being responsive to patient safety. The relationship (between the trustor and trustee) as centered in the ‘trust’ unpacks the AI’s role as a component of the overall healthcare system, not a stand-alone actor to be trusted, held accountable or responsible.

On the other hand, humans (e.g., patients or clinicians) can develop trust to AI as we know today (Gille, Jobin & Ienca, 2020). This reality should not be taken rendering responsibility to AI systems for this is legally and ethically unacceptable. Rather, we should distinguish this factual situation as illustrating a trustworthy relationship within the broader domain of healthcare system in which sub-elements interact with each other resulting in an overall ‘trust’ to the whole healthcare system.

From this perspective, fostering trust in an AI-driven healthcare system should not be limited to the stand-alone AI itself, but should rather cover the entire socio-technical ecosystem in which AI systems are embedded (Aroyo et al., 2021, 432). Mapping an ecosystem is therefore key to better understand trust within the healthcare system, and in this context, trust needs to be taken as a matrix type multidimensional relationship that is fertilized for the whole healthcare system.

Dynamics between trust and ethics in AI-driven healthcare

Within the complex landscape of healthcare systems AI might lie at a critical point for the maintenance of the patient safety giving rise to ethical and legal concerns in potential. Since AI itself should not be taken as a stand-alone trustee and whole ecosystem needs to be analyzed from a holistic perspective, potential ethical and legal concerns require a similar response that needs to be crafted holistically. From this point of view, there should be clarity as to ‘transparency’, ‘responsibility’ and ‘accountability’, as widely acknowledged (UK Department of Health and Social Care; Meijer & Grimmelikhuijsen, 2021; Gille, Jobin & Ienca, 2020). These fundamental thrusts are largely found to build trust, as explained by Meijer & Grimmelikhuijsen (2020, 60):

> “[W]ell-considered choices in terms of how the algorithm is to be used in the organization and a consistent monitoring and evaluation of the desired and undesired outcomes contribute to the perception that [organization] is using these algorithms in a rational manner”.

From a broader point of view, to these three principles (‘transparency’, ‘responsibility’ and ‘accountability’), ‘privacy’ and ‘fairness’ need to be added since AI-driven healthcare often entails privacy and bias/discrimination related concerns, as detailed above. Overall, we uphold the view that the ethical concerns delineated above require adoption and implementation of five main guiding principles: (1) transparency, (2) responsibility, (3) accountability, (4) fairness and (5) privacy. Crucially, an organization’s adhering to an ethical policy for patient safety can be linked and found to contribute strengthening trust.

The structural core of a “work system” comprises of people (e.g., providers, patients, patient family), performing various tasks (e.g., diagnosis, treatment), within a physical environment (e.g., cancer setting, home care), using tools and technologies (e.g., AI enabled technologies, consumer health informatics tools), within an organizational context (e.g., guidelines to integrate AI results into decision-making) (Choudhury & Asan, 2020). Given this, there are many inter-related factors that affect ‘trust’ within the overall landscape of healthcare systems.

Against this background, responsibilities of the actors involved in healthcare provision need to be clarified as to how to use AI and how far to rely on AI. In addition, certain regulatory principles and norms are needed to ensure AI-driven processes be ‘transparent’, along with the clarification of ‘accountable’ decision makers. Furthermore, privacy of the patients, protection of their data as well as representativeness and quality of the datasets for training of AI algorithms need to be secured. All these require a holistic regulatory approach being adopted following a multi-agent systems perspective.

CONCLUSION

AI promises key innovations and significant added value for the healthcare and patient safety. Notwithstanding, far from a stand-alone agent, AI cannot be held liable for the flawed decisions or responsible for their actions in healthcare. While there is an ongoing academic debate over the parameters (e.g., anthropomorphisation, cognitive elements) for enhancing trust in AI, this paper counts on the idea that we need to revisit ‘trust’ to better structure it in the healthcare system and to have a sustainable patient safety.

Findings of this paper demonstrate that there are various ethical and legal concerns emerging with the unmitigated and over reliance on AI, surrounding informed consent, privacy, bias/discrimination and opacity, which all threaten trust in an AI-driven healthcare system. While this is widely acknowledged, a missing point seems to be that without deeply recognizing trust and accompanying relationships, counter principles and remedies would not promise much. It is thus concluded that without a holistic multi-agent systems perspective and translation of this to regulation, any analysis focused on trust and stand-alone AI would be incomplete.

There are many factors that affect ‘trust’ within the overall ecosystem of healthcare system, incorporating human actors, clinical processes, physical environment, technological tools and means. AI appears to be one of the (sub)component that has an ever-increasing role to play for patient safety. Given this, establishment of trust require all these interdependent actors working in cohesion for rebuilding trust on the part of the patients. Once this happens all the actors including clinicians as well as patients would benefit from the overall trust as they both need to maintain a trust relationship rather than simply rely on AI.

From this point of view, robust, multi-dimensional and longitudinal trust relationship is key for patient safety, which needs to be supported by regulatory principles and norms. For the latter, there needs to be further research as this requires an in-depth analysis of how to respond to ethical and legal concerns via regulation from the given holistic perspective.

REFERENCES

AlHogail A. (2018). Improving IoT Technology Adoption through Improving Consumer Trust. *Technologies, 6*(3) 64. doi:10.3390/technologies6030064

Aroyo, A. M., de Bruyne, J., Dheu, O., Fosch-Villaronga, E., Gudkov, A., Hoch, H., Jones, S., Lutz, C., Sætra, H., Solberg, M., & Tamò-Larrieux, A., Overtrusting robots: Setting a research agenda to mitigate overtrust in automation https://doi.org/10.1515/pjbr-2021-0029

Arrow K. J. (1974). *The limits of organization*. New York: Norton

Asan, O., Bayrak, A. E. & Choudhury, A. (2020) Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research, 22*(6). doi:10.2196/15154

Bahl M., Barzilay R., Yedidia A. B., Locascio N. J., Yu L. & Lehman C.D. (2018). High-Risk Breast Lesions: A Machine Learning Model to Predict Pathologic Upgrade and Reduce Unnecessary Surgical Excision. *Radiology, 286*(3), 810-818. doi:10.1148/radiol.2017170549

Balasescu, A. (2021). Ethics, Health and AI in a Covid-19 World: Why Contextual Dynamics Matter and Culture Beats Algor. In *Ethical Implications of Reshaping Healthcare with Emerging Technologies*.

Banerjee I., Li K., Seneviratne M., Ferrari M., Seto T., Brooks J. D., et al. (2019). Weakly supervised natural language processing for assessing patient-centered outcome following prostate cancer treatment. *J Am Med Inform Assoc, 2*(1), 150-159. doi: 10.1093/jamiaopen/ooy057

Blasimme, A. & Vayena, A. E. (2020) The Ethics of AI in Biomedical Research, Patient Care, and Public Health. In *The Oxford Handbook of Ethics of AI*, 702-718

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., . . . Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med, 144*(10). 742-752. doi:10.7326/0003-4819-144-10-200605160-00125

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discov Today, 23*(6), 1241-1250. doi: 10.1016/j.drudis.2018.01.039

Choudhury, A. & Asan O. (2020). Role of Artificial Intelligence in Patient Safety Outcomes: Systemic Literature Review. *JMIR Medical Informatics, 8*(7), 1-24. http://medinform.jmir.org/2020/7/e18599

Choudhury, A., Asan, O. & Mansouri, M. (2020) Role of Artificial Intelligence, Clinicians & Policymakers in Clinical Decision Making: A Systems Viewpoint, *2019 International Symposium on Systems Engineering (ISSE)*, 1-8 doi:10.1109/ISSE46696.2019.8984573

Ciervo, J., Shen, S.C., Stallcup, K., Thomas, A., Farnum, M.A., Lobanov, V. S. & Agrafiotis, D. K. (2019). A new risk and issue management system to improve productivity, quality, and compliance in clinical trials. *JAMIA Open, 2*(2), 216-221. doi:10.1093/jamiaopen/ooz006

Dalal A. K., Fuller T., Garabedian P., Ergai A., Balint C. & Bates D. W. (2019) Systems engineering and human factors support of a system of novel EHR-integrated tools to prevent harm in the hospital. *J Am Med Inform Assoc, 26*(6) 553-560. doi: 10.1093/jamia/ocz002

Danysz, K., Cicirello, S., Mingle, E., Assuncao, B., Tetarenko N., Mockute, R., . . . Desai, S. (2019). Artificial intelligence and the future of the drug safety professional. *Drug safety*, 42(4) 491–497. doi: 10.1007/s40264-018-0746-z

Ekins S., Puhl A. C., Zorn K. M., Lane T. R., Russo D. P. & Klein J.J., et al. (2019). Exploiting machine learning for end-to-end drug discovery and development. *Nat Mater*, *18*(5), 435-441. doi: 10.1038/s41563-019-0338-z

Esmaeilzadeh, P. (2020) Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. *BMC Medical Informatics and Decision Making, 20*(170), 1-20. https://doi.org/10.1186/s12911-020-01191-1

European Commission (2020) White Paper on Artificial Intelligence - A European approach to excellence and trust, COM (2020) 65 final. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission (2021) Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (AI Act) and amending certain Union legislative acts, COM(2021) 206 final, 2021/0106 (COD)

FDA. (2021). Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

FDA. (2019). Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based Software as a Medical Device (SaMD). https://www.regulations.gov/document?D=FDA-2019-N-1185-0001

FDA. (2017). What are examples of Software as a Medical Device? https://www.fda.gov/medical-devices/software-medical-device-samd/what-are-examples-software-medical-device

Finney Rutten, L. J., Blake, K. D., Greenberg-Worisek, A. J., Allen, S. V., Moser, R. P., & Hesse, B. W. (2019). Online Health Information Seeking Among US Adults: Measuring Progress Toward a Healthy People 2020 Objective. *Public Health Rep, 134*(6), 617-625. doi:10.1177/003335491987407

Ford, R.A., Price, W. & Nicholson, I. (2016). Privacy and accountability in black-box medicine, *Mich Telecomm & Tech. L Rev*, *23*(1), 1-43

Gerke, S., Minssen, T. & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare, 295-396, *Artificial Intelligence in Healthcare*, doi: 10.1016/B978-0-12-818438-7.00012-5

Gille, F., Jobin, A. & Ienca, M. (2020). What we talk about when we talk about trust: Theory of trust for AI in healthcare, *Intelligence-Based Medicine*, *1-2,* 100001

Glikson, E. & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Guan M., Cho S., Petro R., Zhang W., Pasche B. & Topaloglu U. (2019). Natural lanuage processing and recurrent network models for identifying genomic mutation-associated cancer treatment change from patient progress notes. *JAMIA Open*, *2*(1), 139-149. doi: 10.1093/jamiaopen/ooy061

Hashimoto D. A., Rosman G., Rus D. & Meireles O. R. (2018). Artificial Intelligence in Surgery: Promises and Perils. *Ann Surg*, *268*(1), 70-76. doi: 10.1097/SLA.0000000000002693

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, 56(11), 1669-1686. doi:10.1080/00140139.2013.838643

Independent High-Level Expert Group (IHLEG) on Artificial Intelligence (set up by the European Commission). (2019). Ethics Guidelines for Trustworthy Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*, *2*(4), 230-243. doi: 10.1136/svn-2017-000101

Liu, Y., Kohlberger T., Norouzi M., Dahl, G. E., Smith, J. L. & Mohtashamian A. et al. (2019). Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection: Insights into the Black Box for Pathologists. *Archives of Pathology & Laboratory Medicine*, *143*(7), 859-868. doi:10.5858/arpa.2018-0147-oa

Maddox T. M., Rumsfeld J. S. & Payne P. R. O. (2019). Questions for Artificial Intelligence in Health Care. *JAMA*, *321*(1), 31-32. doi: 10.1001/jama.2018.18932

Meijer, A. & Grimmelikhuijsen, S. (2020) Responsible and accountable algorithmization: How to generate citizen trust in governmental usage of algorithms. In Schuilenburg, M. and Peeters, R. *The Algorithmic Society: Technology, Power and Knowledge* (Routledge 2020) 52-66

Mirbabaie, M., Hofeditz, L., Frick, N. R. J., Stieglitz, S. (2021). Artificial Intelligence in hospitals: providing a status quo of ethical considerations in academia to guide future research, *AI & Society*, https://doi.org/10.1007/s00146-021-01239-4

Misztal, B. A. (1992). The Notion of Trust in Social Theory. *Policy*, *Organisation and Society*, *5*(1), 6-15. doi:10.1080/10349952.1992.11876774

Rigby, M. J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics*, *21*(2), E121-E124

Ronquillo, J. G., Winterholler, E. J., Cwikla, K., Szymanski, R., Levy, C. & Health, I. T. (2018). Hacking, and cybersecurity: national trends in data breaches of protected health information. *J Am Med Inform Assoc 1*(1), 15-19 doi:10.1093/jamiaopen/ooy019

Rousseau D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. (1998) Not so different after all: a cross-discipline view of trust. *Academy of Management Review 23*(3), 393-404. https://doi.org/10.5465/amr.1998.926617,

Roseman, A. & Zhang X. (2021). Exploring the social, ethical, legal, and responsibility dimensions of artificial intelligence for health - a new column in Intelligent Medicine. https://doi.org/10.1016/j.imed.2021.12.002

Ross, D. & Metnick, C. (2019). Artificial Intelligence and Health Care: Legal and Practical Considerations (AHLA). https://www.americanhealthlaw.org/publications/health-law-hub-current-topics/artificial-intelligence-and-health-law

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics 26*, 2749-2767. https://doi.org/10.1007/s11948-020-00228-y

Safdar, N. M., Banja, J. D., Meltzer, C. C. (2020). Ethical considerations in artificial intelligence, *European Journal of Radiology*, 122, 108768. doi: 10.1016/j.ejrad.2019.108768

Sahli C. F., Matsuno K., Yao J., Perdikaris P. & Kuhl E. (2019). Machine learning in drug development: Characterizing the effect of 30 drugs on the QT interval using Gaussian process regression, sensitivity analysis, and uncertainty quantification. *Comput Meth Appl Mechan Eng*, *348*, 313-333. doi: 10.1016/j.cma.2019.01.033

Schonberger, D. (2019). Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications, *IJLIT*, *27*, 171-203. https://doi.org/10.1093/ijlit/eaz004

Sharan, N. N. & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence, *Heliyon*, *6*, e04572

Siau, K. & Wang, W. (2020). Artificial Intelligence (AI) ethics. *J Database Manag*, *31*, 74-87. https://doi.org/10.4018/jdm.2020040105

Siau, K. & Wang, W. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, *31*(2), 47-53.

Tuffrey, I., Giatras, N., Goulding, L., Abraham, E., Fenwick L., Edwards, C., & Hollins, S. (2013). Identifying the factors affecting the implementation of strategies to promote a safer environment for patients with learning disabilities in NHS hospitals: a mixed-methods study-Health Services and Delivery Research. In *Health Services and Delivery Research. Southampton*. doi:10.3310/hsdr01130

UK Department of Health and Social Care. (2021). A guide to good practice for digital and data-driven health technologies, https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology.

World Health Organization (WHO). (2019). Patient Safety: Key facts. https://www.who.int/news-room/fact-sheets/detail/patient-safety