Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians

Onur Asan, PhD; Alparslan Emrah Bayrak, PhD; Avishek Choudhury, MSc

School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Terrace
Hoboken, NJ, 07030
United States
Phone: 1 2012163666
Email: oasan@stevens.edu

Abstract

Artificial intelligence (AI) can transform health care practices with its increasing ability to translate the uncertainty and complexity in data into actionable—though imperfect—clinical decisions or suggestions. In the evolving relationship between humans and AI, trust is the one mechanism that shapes clinicians' use and adoption of AI. Trust is a psychological mechanism to deal with the uncertainty between what is known and unknown. Several research studies have highlighted the need for improving AI-based systems and enhancing their capabilities to help clinicians. However, assessing the magnitude and impact of human trust on AI technology demands substantial attention. Will a clinician trust an AI-based system? What are the factors that influence human trust in AI? Can trust in AI be optimized to improve decision-making processes? In this paper, we focus on clinicians as the primary users of AI systems in health care and present factors shaping trust between clinicians and AI. We highlight critical challenges related to trust that should be considered during the development of any AI system for clinical use.

Introduction

Artificial intelligence (AI), which has been introduced as a technology to improve decision-making involving uncertainty and complexity in systems, has the potential to transform health care practices [1,2]. The role of humans in the practical applications of AI is often overlooked. The development of automated systems to augment human decision-making dates to the 1950s with the Fitts list, which identifies the complementary capabilities of humans and automated systems [3]. The Fitts list includes 11 statements asserting that humans are better at detection, perception, judgment, induction, improvisation, and long-term memory, while automated systems are better at power/speed, computation, replication, simultaneous operation, and short-term memory [4]. Several studies have shown that automated systems may or may not improve human decision-making, depending on whether or how human factors are accounted for in their design [5].

As AI is rapidly developing, unlike other technologies, there is an absence of a clear definition of the process, functioning, and role of AI [6]. Trust is a crucial factor influencing interactions between human beings, including their interactions with AI. Understanding the trust dynamics between AI and humans is crucial, particularly in the field of healthcare, where life is at risk. In this paper, we discuss the impact of trust on the dynamic interactions between AI and clinicians, highlight the key factors that influence trust relationships and identify key challenges and future research directions in the health care domain. While the users of AI systems can be diverse, including patients and insurance providers, the focus of this paper is limited to the domain experts in healthcare, ie, clinicians. We acknowledge that trust relationships could significantly differ for patients and insurance providers.

Definitions

What Is AI?

The term AI has been used in many ways in computer science, engineering, and healthcare. Broadly, it can be defined as a computer program that can make intelligent decisions [7]. This definition includes computer programs that operate with predefined rules and data-driven models. This paper distinguishes these two by referring to the former as *automation*, which can be used to make well-defined and repetitive decisions. While automated systems have been used to augment or replace human operation in health care [8], the generalizability of the process and existence of intelligence in automation can be questionable. The focus of this research is the latter, which is a process to make health care decisions using a mathematical model built on prior or real-time data. Existing literature in the machine learning field provides relatively successful methods to train such mathematical models that learn useful knowledge from data [9-11]. In this paper, we refer to *a computer process that algorithmically makes optimal decisions based on multiple criteria using one or more machine learning-based models* as AI. While trust has been studied in the context of automation [12], the deterministic (ie, consistently providing the same output for a particular input) and relatively predictable nature of automation is an important distinction from our definition of AI, which has implications for trust research.

What Is Trust?

Interpersonal trust is a human belief (or referred to as an attitude in some sources [12]) that is broadly defined based on three main dimensions, namely, benevolence, integrity, and ability [13]. This attitude may be intrinsically formed based on the user's own experience with the system of interest or may stem from an extrinsic source such as the reputation of the system in the user's social circle [14]. Studies highlight some differences between interpersonal trust and trust in technical systems because technical systems may lack intentionality, which is relevant to honesty and benevolence [12]. A user's perception of an AI system's ability remains a significant dimension for trust in AI systems, which depends on the quality of the input data, the mathematical problem representation, and the algorithms used in the decision-making. The level of trust in AI has a significant impact on how much users rely on AI [12], and hence the efficacy of health care decisions. However, the level of trust in AI may not necessarily have a positive correlation with clinical or patient outcomes.

AI in Health Care

AI has shown significant potential in the area of mining medical records, designing treatment plans [15], robotics mediated surgeries [16], medical management and supporting hospital operations, clinical data interpretation [17], clinical trial participation [18,19], image-based diagnosis [20], preliminary diagnosis [21], virtual nursing [22], and connected health care devices [23]. In addition to these applications, significant investments in AI research [24], as well as recent efforts on regulating the use of AI in the medical domain [24], suggest that AI could become an essential technology to assist decision-making in the medical domain in the foreseeable future.

AI research in health care poses unique challenges compared to other technical domains. Physical system models mathematically describe the underlying technical behavior in engineering applications. However, the lack of such quantitative models in many health care applications such as medical diagnostics (eg, the precise relationships between diseases and their causes) creates a significant challenge. The responses from clinicians for the same clinical cases vary significantly. Therefore, it would be a challenge to train AI-based tools on the subjective responses that carry over individual biases from clinicians without any knowledge of the ground truth. Also, AI research must account for the distinct medical problem characteristics across different health care applications. It may not be possible to generalize a process to train a mathematical model for an AI tailored to the needs of cancer applications to cardiovascular applications, for instance. Further, vulnerable groups such as minorities and patients with disabilities may not be sufficiently represented in the data, and their needs may not be adequately accounted for if these groups are not carefully considered during the design of the AI system. A customized AI process might be necessary for each application depending on the type and amount of data available, the target patient population, the amount of variability and useful information in the data, and the nature of the health care decisions to be made.

Further, health care is a highly regulated space where developing an adaptive AI technology to meet regulatory requirements is an additional challenge. 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" [25]. 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 [26]. 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 [25,27].

Unlike the standard SaMD model, an adaptive algorithm changes its behavior using a definitive learning process without requiring any manual input. An adaptive algorithm might generate different outputs each time a given set of inputs is received due to learning and updating. A credible validation and verification plan that ensures safe and reliable operation under adaptive behaviors must be a part of the AI design.

AI in health care has two potential advantages to human performance. First, AI can learn from big data (such as incommunicable silos of unstructured information stored in an electronic health record) more efficiently than clinicians. A successful AI system can efficiently extract relevant information from offline or real-time data to assist in improving organizational performance and help clinicians in making
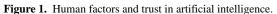
informed decisions in real time. Second, AI systems can perform predefined tasks with higher precision. AI can be in a continuous active state without compromising its performance—it does not suffer from burnout as humans do. This feature of AI technology has the potential to revolutionize complicated surgeries. The Da Vinci robotic surgical system can mimic a surgeon's hand movements with greater precision [28]. Further, existing applications of AI in various domains such as AlphaStar (an AI bot that outperforms an expert player in a video game) and LYNA (an AI capable of detecting breast cancer using images from lymph node biopsies) report successful outcomes comparable to human decision-making [29,30].
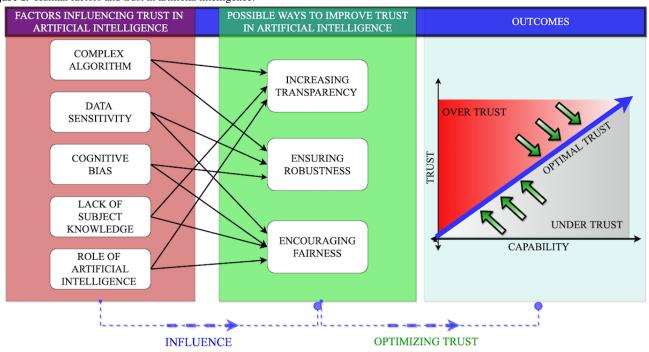
There are limitations of AI that might restrict its application in life-critical areas such as healthcare. AI methods require data inputs to be in a structured form, which limits the type of information that can be provided for medical decisions. Even the deep learning methods, which can find a suitable mathematical representation from a given dataset automatically, are trained for a given input type (eg, medical image data) and, once developed, they cannot accept another input type (eg, statements from patients). Also, AI methods generally lack "common sense," making them unable to identify simple mistakes in data or decisions that would otherwise be obvious to a human being [31]. Therefore, there is significant potential and need for improvement by combining the intuitive and

analytical thinking of medical experts [32] with the computational power of AI in a proper human-AI collaboration architecture.

Trust in Human-AI Collaboration

Advances in the capabilities of AI will expand the role of this technology from the automation of repetitive and well-defined tasks to guiding decision-making under uncertainty that is currently done exclusively by medical professionals. As health care providers rely more on AI, a proper trust relationship, also referred to as *calibrated* trust [33], becomes a requirement for effective decisions. The deterministic and relatively predictable nature of a typical rule-based software system is an essential factor contributing to the development of user trust. The resulting behavior of a deterministic system can entirely be determined by its initial state and inputs. However, the nondeterministic nature of AI, where an algorithm can exhibit different behaviors in different iterations for the same input, might introduce unique dimensions to the concept of trust. Figure 1 presents our overview of some important factors influencing trust in AI for health care, possible ways to improve trust relationships, and their impact on trust. Note that the purpose of the figure is not to provide an exhaustive list but rather to highlight important issues relevant to trust in AI for health care applications.

Figure 1. Human factors and trust in artificial intelligence.

Currently, a lack of trust in the AI systems is a significant drawback in the adoption of this technology in healthcare. Trust in AI can be influenced by several human factors such as user education, past experiences, user biases, and perception towards automation, as well as properties of the AI system, including controllability, transparency, and complexity of the model, associated risks, and many others. Among these factors, reliability, which refers to whether the AI technology can perform a task predictably and consistently [34], might be

particularly concerning in health care due to the changes in the reliability of AI in the presence of new data [35]. The reliability of an AI technology is conditioned on the user and input data. Considering that an AI system might be trained with insufficient and subjective data from multiple sources, AI could generate biased or overfitted outcomes of which the clinical user might not be aware. These concerns hinder the performance of this technology [36], thus deterring the user's trust and acceptance of AI systems.

It is also important to note that *maximizing the user's trust* does not necessarily yield the best decisions from a human-AI collaboration. When trust is at maximum, the user accepts or believes all the recommendations and outcomes generated by the AI system. While in some applications, AI can outperform human decision-making by incorporating data from multiple sources [29,30], the limitations above suggest that unnecessarily high trust in AI may have catastrophic consequences, especially in life-critical applications. Therefore, our study supports the concept of *optimal trust* in which both humans and AI each have some level of skepticism regarding the other's decisions since both are capable of making mistakes. The amount of skepticism necessary for the most accurate clinical decisions depends on the capability of the human user and the AI system. The development of AI must incorporate mechanisms that will establish and maintain a properly balanced, optimal level of trust from and to the user that matches the capability of the AI system [12].

Developing a healthy trust relationship in human-AI collaboration is a challenge due to the uncertain behavior of a knowledge-based evolving AI system. We posit that the following factors should be incorporated into the development of AI to achieve an optimal level of trust: fairness, transparency, and robustness (Figure 1). It is often assumed that algorithmic decision-making might lead to fairer and more robust outcomes than human judgment [37]. However, algorithms intrinsically discriminate and assign a weight to some factors over others. Moreover, the properties of machine learning algorithms bear the risk to reflect and aggravate underlying data bias, which might unfairly affect members of protected groups based on sensitive categories such as gender, race, and sexual orientation. For instance, a study [38] that implemented machine learning to detect skin cancer used less than 5% data from a dark-skinned patient population to train the model, potentially leading to bias against dark-skinned patients [39]. According to the UK International Commissioner's Office, AI fairness depends on the effects of the data processing on users (care providers and patients), their expectations as to how their data will be used (clinical and personal data) and the level of transparency provided [40,41] (sharing with other care providers and insurance). Fairness in an AI process is concurrent with bias minimization. Bias, a mismatch between the distribution of training data and a preferred fair distribution, can yield unfair outcomes (prediction/classification). In order to establish fairness in AI systems, biases originating from the subjective responses of clinicians should be identified and curtailed during validation and verification.

Achieving fairness through awareness or transparency can also improve trust [42,43]. Facilitating secured access to patient data can improve transparency [44] in the data and allow patients to validate their information. In our view, transparency fosters an understanding not only of the working principles of AI algorithms but also the algorithmic biases and biases due to underrepresentation. Depending on data availability and quality, an AI system might perform remarkably well at some tasks while performing poorly in others. Revisiting the 'skin cancer example' cited above [38], if clinicians are informed about the bias in the training data (eg, underrepresentation of minorities),

then it will be easier for them to identify the suitable patient population (eg, Caucasians) on which the AI algorithm can be implemented. Explainable AI is another route to AI transparency that might help clinicians arrive at clinically meaningful explanations about the outcome of AI applications and make informed judgments. However, there are tradeoffs between the explainability and sophistication of an AI algorithm. Explainable AI models such as decision trees intuitively tend to be simple and might not be accurate, especially when dealing with big and complex data. Determining the precise balance between explainability and sophistication is crucial to enhance trust in AI. Finally, robustness is the sensitivity of the decisions made by the AI models to the input dataset. Poor robustness can lead to significant changes in the outcome of an AI model with small perturbations in the input data. Insufficient or erroneous data can impede AI robustness. Proofing AI models against such volatilities can help to build trust in AI systems.

Future Research Directions

The efficacy of human-AI collaboration is not only a function of the accuracy of the underlying mathematical process behind the AI system but also human factors, including trust. A holistic approach recognizing health care as a dynamic socio-technical system in which sub-elements interact with each other is necessary to understand trust relationships in human-AI collaboration. For instance, trust in AI systems might be affected by organizational policies, culture, specific tasks assigned to the health care providers, other similar computational tools used by the providers, providers' interaction with other individuals such as patients and other providers, as well as internal and external environmental factors. This viewpoint is consistent with and complementary to the research roadmaps proposed in the systems engineering literature on AI [45]. Applying human factors methodologies such as the SEIPS model [46] to the health care domain can assist researchers in capturing the entire socio-technical work system. These holistic human factors models provide a useful conceptual framework for researchers to capture contemporary and dynamic issues relevant to trust modeling in healthcare.

Second, recent approvals of AI algorithms reveal the limitations of existing regulatory standards. In early 2018, the FDA authorized the WAVE Clinical Platform, an early-warning system that utilizes vital sign data to identify high-risk patients [35]. FDA standards, designed for traditional rule-based algorithms, do not apply to advanced AI systems whose predictive performance might change when exposed to new data [35]. To measure the impact of AI systems, FDA should benchmark the predictive model and ensure clinically meaningful outcomes. As it does for drug approvals, FDA should rigorously confirm and test surrogate endpoints in potential evaluations of AI systems. Recently, the FDA announced that it is developing a framework for regulating AI systems that self-update on new data and seeking comments on how to regulate self-updating algorithms [47].

AI raises profound concerns regarding medical responsibility. Under current practice, clinicians are only responsible when they deviate from the standard care procedure for a given health

XSL·FO

RenderX

condition (irrespective of patient health outcome) [48]. However, clinicians will be held responsible if they follow the AI recommendation when it is different from the standard care process and negatively affects patient health outcomes [48]. What would this mean for an AI system and trust between AI and users? Clinicians using AI systems are expected to use them for decision support, not as a replacement for trained clinical decision-making. In this sense, clinicians are still responsible for any medical errors that may occur as humans remain the final decision-makers. Then, in what capacity can AI assist clinicians and will clinicians be able to use and assess the reliability of an AI system? The influence of these factors on clinicians' trust in AI applications needs further study.

Considering the limitations of both human cognition and AI approaches, a quantitative measure for the optimal level of trust between clinicians and AI systems to make the most accurate and reliable clinical decisions remains unknown. Linking this optimal level of trust to specific design attributes in an AI system is another unknown. An analysis of that problem should account for the individual human factors specific to the user of the system, including the sizeable aleatory variability associated with it and continuously evolving capabilities of AI methods. The results of such an analysis should inform the regulatory policy decisions.

Finally, trust in AI is expected to have completely different characteristics for patients compared to clinicians. First, patients generally do not have expertise in the medical field, as opposed to clinicians. Further, patients, regardless of whether they are the users of the AI system or not, will be directly impacted by the clinical decisions (or suggestions) that the AI system provides. With the increased patient involvement in a patient-centered health care model, especially educated patients might question clinicians' decisions and want to be informed whether the decisions are based on AI recommendations or not. If AI systems are an essential part of this shared decision-making process between patients and clinicians, trust relationships between patients and AI systems deserve further in-depth research.

Conflicts of Interest

None declared

References

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
2. Tekkeşin A. Artificial Intelligence in Healthcare: Past, Present and Future. Anatol J Cardiol 2019 Oct;22(Suppl 2):8-9 [FREE Full text] [doi: 10.14744/AnatolJCardiol.2019.28661] [Medline: 31670713]
3. de Winter JCF, Dodou D. Why the Fitts list has persisted throughout the history of function allocation. Cogn Tech Work 2011 Aug 25;16(1):1-11. [doi: 10.1007/s10111-011-0188-1]
4. Fitts P. Human engineering for an effective air-navigationtraffic-control system. Fitts PM. editor. Oxford, England: National Research Council, Div. of; 1951.
5. Parasuraman R, Riley V. Humans and Automation: Use, Misuse, Disuse, Abuse. Hum Factors 2016 Nov 23;39(2):230-253. [doi: 10.1518/001872097778543886]
6. Siau K, Wang W. Building Trust in Artificial Intelligence, Machine Learning, and Robotics. Cutter Business Technology Journal 2018;31(2):47-53.
7. McCarthy J, Hayes P. Some philosophical problems from the standpoint of artificial intelligence. Readings in nonmonotonic reasoning: Morgan Kaufmann Publishers Inc; 1987:26.
8. Nundy S, Montgomery T, Wachter RM. Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence. JAMA 2019 Aug 13;322(6):497. [doi: 10.1001/jama.2018.20563]
9. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan 7;25(1):24-29. [doi: 10.1038/s41591-018-0316-z]
10. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2017 May 06:1236-1246. [doi: 10.1093/bib/bbx044] [Medline: 28481991]
11. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clin Infect Dis 2018 Jan 06;66(1):149-153 [FREE Full text] [doi: 10.1093/cid/cix731] [Medline: 29020316]
12. Lee JD, See KA. Trust in Automation: Designing for Appropriate Reliance. hfes 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]
13. Mayer RC, Davis JH, Schoorman FD. An Integrative Model of Organizational Trust. The Academy of Management Review 1995 Jul;20(3):709. [doi: 10.2307/258792]
14. Han Yu, Zhiqi Shen, Leung C, Chunyan Miao, Lesser VR. A Survey of Multi-Agent Trust Management Systems. IEEE Access 2013;1:35-50. [doi: 10.1109/access.2013.2259892]
15. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. Correction: World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLoS Med 2015 Dec;12(12):e1001940 [FREE Full text] [doi: 10.1371/journal.pmed.1001940] [Medline: 26701262]
16. Davenport TH, Glaser J. Just-in-time delivery comes to knowledge management. Harv Bus Rev 2002 Jul;80(7):107-11, 126. [Medline: 12140850]

17. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]

18. Fink E, Kokku P, Nikiforou S, Hall L, Goldgof D, Krischer J. Selection of patients for clinical trials: an interactive web-based system. Artificial Intelligence in Medicine 2004;31(3):241-254 [FREE Full text] [doi: 10.1016/j.artmed.2004.01.017]

19. Beck JT, Rammage M, Jackson GP, Preininger AM, Dankwa-Mullan I, Roebuck MC, et al. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center. JCO Clin Cancer Inform 2020 Jan;4:50-59 [FREE Full text] [doi: 10.1200/CCI.19.00079] [Medline: 31977254]

20. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Comput Struct Biotechnol J 2019;17:995-1008 [FREE Full text] [doi: 10.1016/j.csbj.2019.07.001] [Medline: 31388413]

21. Zhao M, Jiang Y. Great expectations and challenges of artificial intelligence in the screening of diabetic retinopathy. Eye 2019 Dec 11;34(3):418-419. [doi: 10.1038/s41433-019-0629-2]

22. Peete R, Majowski K, Lauer L, Jay A. Artificial Intelligence in Healthcare. Artificial Intelligence and Machine Learning for Business for Non-Engineers 2019:96. [doi: 10.1201/9780367821654-8]

23. Durkin K. Artificial Intelligence-driven Smart Healthcare Services, Wearable Medical Devices, and Body Sensor Networks. Am. J. Med. Res 2019;6(2):37. [doi: 10.22381/ajmr6220195]

24. White House American artificial intelligence initiative year one annual report. In: policy Oosat, editor. USA executive office of the president of the United States; 2020:E-29.

25. FDA. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based Software as a Medical Device (SaMD) URL: https://www.regulations.gov/document?D=FDA-2019-N-1185-0001 [accessed 2020-04-01]

26. FDA. Food and Drug Administration. 2017. What are examples of Software as a Medical Device? : U URL: https://www.fda.gov/medical-devices/software-medical-device-samd/what-are-examples-software-medical-device [accessed 2020-04-01]

27. Networks Inc. K. Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices. CMR 2019 Apr 02:1. [doi: 10.31525/fda2-ucm635083.htm]

28. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019 Jul;8(7):2328-2331 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_440_19] [Medline: 31463251]

29. DeepMind. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II URL: https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii [accessed 2020-05-05]

30. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial Intelligence–Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. Archives of Pathology & Laboratory Medicine 2019 Jul;143(7):859-868. [doi: 10.5858/arpa.2018-0147-oa]

31. Goodfellow I, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. 2015 May 07 Presented at: Proceeding of 3rd International Conference on Learning Representations; May ; San Diego; 2015; San Diego p. 7-9.

32. Evans JSBT. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. Annu. Rev. Psychol 2008 Jan;59(1):255-278. [doi: 10.1146/annurev.psych.59.103006.093629]

33. Hoffman RR, Johnson M, Bradshaw JM, Underbrink A. Trust in Automation. IEEE Intell. Syst 2013 Jan;28(1):84-88. [doi: 10.1109/mis.2013.24]

34. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans. Manage. Inf. Syst 2011 Jun;2(2):1-25. [doi: 10.1145/1985347.1985353]

35. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science 2019 Feb 21;363(6429):810-812. [doi: 10.1126/science.aaw0029]

36. AlHogail A. Improving IoT Technology Adoption through Improving Consumer Trust. Technologies 2018 Jul 07;6(3):64. [doi: 10.3390/technologies6030064]

37. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. International Journal of Law and Information Technology 2019;27(2):a. [doi: 10.1093/ijlit/eaz004]

38. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]

39. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature 2018 Jul;559(7714):324-326. [doi: 10.1038/d41586-018-05707-8] [Medline: 30018439]

40. Office I. ICO. 2019. Guide to data protection URL: https://ico.org.uk/media/for-organisations/guide-to-data-protection-1-1.pdf [accessed 2020-04-04]

41. Butterworth M. The ICO and artificial intelligence: The role of fairness in the GDPR framework. Computer Law & Security Review 2018:257-268 [FREE Full text] [doi: 10.1016/j.clsr.2018.01.004]

42. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through Awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. Cambridge, Massachusetts: Association for Computing Machinery; 2012 Presented at: 3rd Innovations in Theoretical Computer Science Conference; 2012; Cambridge. [doi: 10.1145/2090236.2090255]

43.   Feldman M, Friedler S, Moeller J, Scheidegger C, Venkatasubramanian S. Certifying and Removing Disparate Impact. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Sydney, NSW. Australia: Association for Computing Machinery; 2015 Presented at: 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015; Sydney. [doi: 10.1145/2783258.2783311]

44.   Ferreira A, Lenzini G. Springer. Can Transparency Enhancing Tools Support Patient?s Accessing Electronic Health Records? New Contributions in Information Systems and Technologies: Springer; 2015:1121-1132.

45.   McDermott T, DeLaurentis D, Beling P, Blackburn M, Bone M. AI4SE and SE4AI: A Research Roadmap. INSIGHT 2020 Mar;23(1):8-14. [doi: 10.1002/inst.12278]

46.   Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013 Nov;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]

47.   FDA. Artificial Intelligence and Machine Learning in Software as a Medical Device URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-devicefiles/7614/artificial-intelligence-and-machine-learning-software-medical-device.html [accessed 2020-05-01]

48.   Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. JAMA 2019 Oct 04:1765-1766. [doi: 10.1001/jama.2019.15064] [Medline: 31584609]

Abbreviations

AI: artificial intelligence
FDA: Food and Drug Administration
SaMD: Software as a Medical Device

XSL•FO
RenderX