



## OPEN ACCESS

## EDITED BY

Patrick C. Shih,  
Indiana University Bloomington, United States

## REVIEWED BY

Don Donghee Shin  
Zayed University, United Arab Emirates

## \*CORRESPONDENCE

Avishek Choudhury,  
avishek.choudhury@mail.wvu.edu

## †ORCID

Avishek Choudhury  
orcid.org/0000-0002-5342-0709  
Safa Elkefi  
orcid.org/0000-0002-4293-0404

## SPECIALTY SECTION

This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

RECEIVED 10 June 2022

ACCEPTED 08 August 2022

PUBLISHED 23 August 2022

## CITATION

Choudhury A and Elkefi S (2022) Acceptance, initial trust formation, and human biases in artificial intelligence: Focus on clinicians. *Front. Digit. Health* 4:966174. doi: 10.3389/fdgth.2022.966174

## COPYRIGHT

© 2022 Choudhury and Elkefi. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Acceptance, initial trust formation, and human biases in artificial intelligence: Focus on clinicians

Avishek Choudhury<sup>1\*†</sup>  and Safa Elkefi<sup>2†</sup> 

<sup>1</sup>Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States, <sup>2</sup>School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

## KEYWORDS

trust & distrust, artificial intelligence, healthcare, patient safety, technology acceptance, communication, human biases

## Introduction

Healthcare systems have benefited from technology use and development. Different technologies have augmented the healthcare system at different capacities and consequently have advocated for patient safety. *Artificial Intelligence (AI) is one such emerging technology that can extract latent information from a plethora of data to formulate actionable knowledge based on pre-defined or adaptive algorithms.* Unlike other healthcare technologies, AI has not been wholeheartedly welcomed by the healthcare community and has often gained an overall negative perception. We acknowledge that healthcare AI, like any other technology, comes with certain drawbacks, but its potential to augment healthcare systems and care quality is a significant factor encouraging its integration into the clinical workflow.

Research around healthcare AI can be categorized into two groups, one that promises multiple benefits of the AI (1, 2) and the other that criticizes the technology (3). Although both the body of work is crucial and noteworthy, not much evidence exists that can confirm AIs' negative impact on patient safety or health outcome. Studies have demonstrated AI's potential in multiple healthcare domains, including the dentistry (4), pediatric (5), and geriatric and adult populations (6). Given the evidence and potential of AI, it is now the time to emphasize improving AI acceptance and trust among clinicians and healthcare organizations. Currently, several studies, not limited to healthcare, are directed to resolve the problem of distrust in AI, where the trend is to achieve a calibrated trust in the technology (7, 8). Trust in AI can be defined as the tendency of the user to rely on AI recommendation and willingly be vulnerable to the technology. Thus far, most of the studies addressing AI trust have assumed trust as a construct that varies overtime and have explored the role of fairness (9), interpretability, privacy (10, 11), and algorithmic awareness (12) in shaping users' trust in AI. Other factors such as their interaction with the technology, AI accuracy, data quality, and biases, have been considered to significantly impact user trust in AI. However, it is important to understand that these factors are only experienced by the user once they use the technology. For example, to realize the accuracy, transparency, or usability of a given AI system, the user has to first use the

technology. What is missing in the literature is to capture the factors that would initially motivate the user to use the AI system for the very first time. That is, to explore the potential cognitive human factors responsible for shaping clinicians' initial trust in AI. Since clinicians typically resist and hold a negative perception of the technology, it is essential to understand how initial trust in AI is formed.

We define initial trust as a binary (trust = 1 or distrust = 0) latent construct that can be confirmed if the user, without any obligation or influence, acts upon/accepts AI recommendations without having any prior experience with the technology. In other words, initial trust in AI is a measure of the extent to which a clinician is willing to take clinical action or decision based on the information generated by the AI technology, provided free will of the clinician and no prior experience with the technology. Initial trust in AI is a significant precursor to its use and gradual acceptance as a useful healthcare technology. In this article, we portray initial trust in AI as a function of multiple factors such as information relevance, patient risk, and the mental model or perception of the clinician. We also discuss how AI can improve healthcare communication.

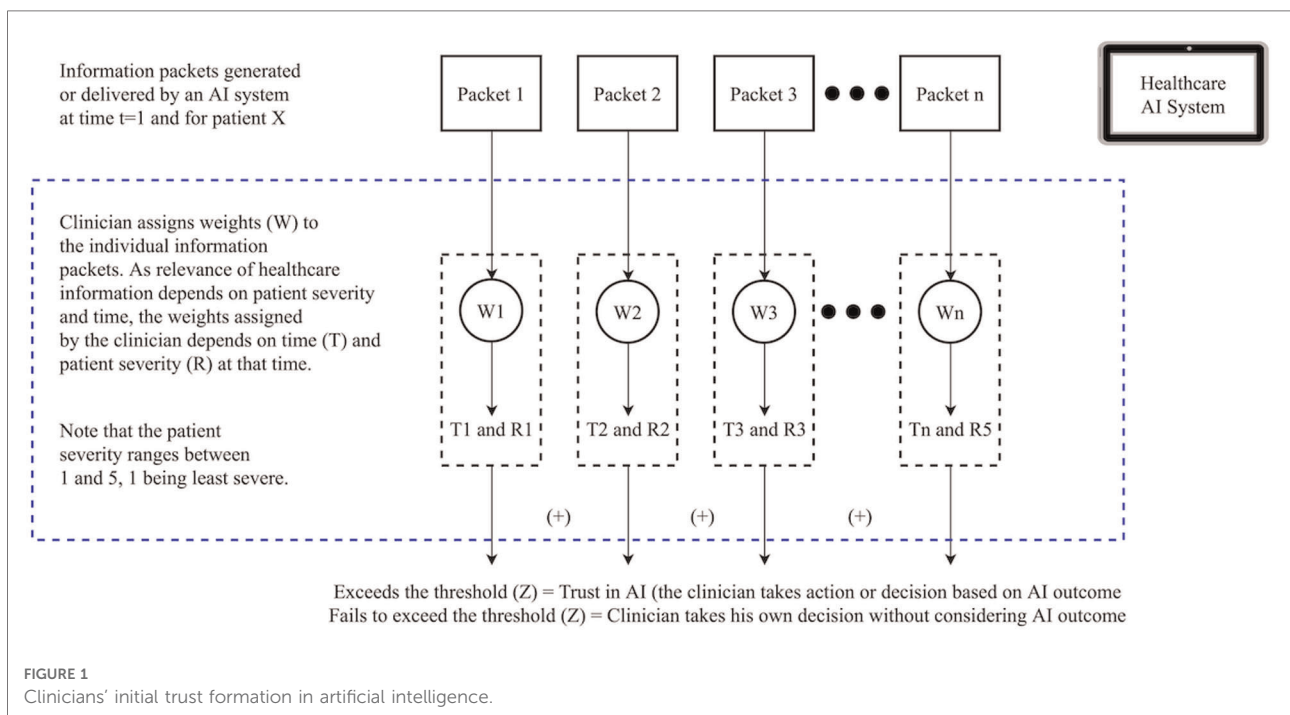
### Initial trust formation in healthcare artificial intelligence: Clinicians' perspective

It is important to understand that an AI system that is implemented into a clinical workflow is extensively tested and approved by the concerned authorities ensuring its accuracy

and overall performance. What's missing is understanding why clinicians hesitate to adopt AI and the factors determining their initial trust formation in the technology. In general, trust in AI fluctuates with prior experience and repeated interactions (iterative) with the technology, but the initial trust (i.e., trust in an unfamiliar entity without prior experience) is formed through assumptions, perceptions, and quick inferences about the trustee from a limited information (13). Understanding the initial trust formation is essential to initiate AI use and acceptance in the healthcare domain.

Literature has identified several important factors that shape users' trust in AI, including AI biases, data quality, reliability, and explainability (14). But these factors primarily depend on the user perceptions and situations, cannot be measured with a numeric standard value. Thus, it is important to incorporate human factors consideration. Whether any clinician (first-time AI user) will place their initial trust in the technology depends on whether the clinician thinks that the information (recommendations) generated by the AI will have clinical relevance and will be applicable to a given context. **Figure 1** illustrates a hypothetical and a general scenario where an AI system embedded in the electronic health record produces a set of recommendations for a clinician.

The recommendation generated by the AI comprises multiple packets of information ( $I = \sum_1^n (\text{Packet})$ ). After receiving this onscreen information (I), the clinician assigns weight (W) to each piece of information. The weights (W) assigned to each information depend on time (T) and patient severity (R) ( $W_n = f(Tn, Rn)$ ). Certain information can be crucial at a given point in time and for a certain patient type.



Note that patient severity (R) is a measure of their health status (deviation from the healthy baseline. For patient acuity, R can range from 1 to 5).

To ensure clinicians' initial trust in AI (i.e., Initial Trust = 1), the summation of the product of weights and information packet ( $W_n I_n$ ) should exceed a given threshold (Z), i.e., ( $W_n I_n > Z$  where  $W_n = \sum_1^n (W)$  and  $I = \sum_1^n (\text{Packet})$ ). The threshold (Z) depends on several human factors, including the clinician's experience (Ex), expertise (Ep), situation awareness (SA), and cognitive workload (WL). In other words,  $Z = f(\text{Ex}, \text{Ep}, \text{SA}, \text{WL})$ . Fundamentally, users' (clinicians') initial trust depends on the information generated by a given AI technology and not the technology or its functioning per se.

## Human biases prevent acceptance of artificial intelligence

AI biases have been extensively discussed in the literature, where studies have acknowledged how skewed data can induce bias in AI. However, the inherent and latent biases in human users are not well recognized in the context of healthcare AI. Human (clinicians') biases are a strong determinant of initial trust in AI and can hinder its acceptance in healthcare.

Clinicians often resist AI integration into their workflow. Their skepticism toward AI builds upon a few presumptions, where lack of initial trust in technology plays a significant role. The Dunning-Kruger effect explains why clinicians refrain from trusting AI systems. According to the Dunning-Kruger effect, people with expertise in a specific field (specialist doctor) often overestimate their own competence in that domain and perceive their own opinion or judgment over anything else (15), including AI. It is common for clinical experts (senior physicians) to express confirmation bias and ignore AI's recommendation whenever it contradicts their presumption/judgment. Thus far, several initiatives have been taken by national and global authorities to regulate, standardize, and improve AI. However, critical factors such as cognitive biases and user perceptions require further exploration.

## Healthcare communication and artificial intelligence

Communication is an essential process in the healthcare industry. Communication between clinicians and AI can foster patient safety and quality of care. However, under excessive workload and chaos, clinicians may miss some of the critical information projected by the AI system or fail to comprehend its actual meaning. To benefit from AI systems and develop the initial trust, ensuring a perception of shared

understanding between AI and clinician, is essential. Once initial trust is established, AI integration into the clinical workflow can play an essential role in improving the clinical communication (16). For example, a speech recognition device powered by AI can detect distinct vocal changes associated with specific health conditions (17). Vocal analysis and speech recognition can help doctors make informed decision-making and encourage personalized care.

The typical notion about AI is that the technology will replace clinicians and reduce interpersonal touch. However, in reality, AI can increase doctor-patient interaction time. Currently, clinician-patient interaction is hindered by excessive clinical workload. Clinicians, during patient visits, spend a significant amount of time taking notes and simultaneously updating the electronic health records system. In this scenario, patients typically feel they are not being listened to. AI technologies (e.g., Voice AI) can eliminate excessive clinical workload by automating taking clinical notes (18). AI can also enter each item into the appropriate field or section of the EHR (18). AI can also play an essential role in coordinating between doctors and patients, and improve information exchange between the stakeholders by facilitating their work and controlling the risks related to clinical tasks (19). AI can also be used to simulate possible scenarios that can help teach nurses how to hone their communication skills before going on to clinical placements using unlimited training attempts in a safe and secure environment (20).

## Discussion and conclusion

In this article, we reemphasize the importance of initial trust formation in AI, clinicians' perception, and how they assign relevance to AI outputs for a given patient. Additionally, we also discuss the role and importance of considering cognitive biases to encourage AI acceptance and how AI can augment healthcare communication. Much of the research around healthcare AI has been designed to improve AI transparency, an important component. However, algorithmic transparency will not suffice or ensure AI acceptance. Not only does making the fundamental function of AI systems transparent may cause proprietary issues, but it can also consume clinicians' working memory because understanding AI algorithms are difficult. Having the complex AI functioning along with clinical procedural information stored in long-term memory while attending to critical patients can be overwhelming and not feasible for many clinicians.

When automated diagnostic systems are used in real-life clinics, it is most likely as assistant or recommendation systems, where the AI system provides information to clinicians or patients as a second opinion. However, if the suggestions made by AI are entirely data-driven without accounting for user's opinion, users could get skewed toward

or against the suggestion of the AI system leading to automation surprises or biases. Therefore, the next push for researchers should be to move AI research out from solely model development into socio-technical systems research and effectively use human factors and psychological principles emphasizing its ecological validity. It is important to understand that a clinical environment is not inherently equivalent to predictable mechanical systems and needs a systematic approach where users' perception is critical. In fact, more realistically, the acceptance of AI into the inpatient clinical workflow entirely depends on how doctors and nurses perceive this technology – their intent to use AI and their initial trust in the technology.

## Author contributions

AC conceptualized the study and participated in manuscript writing, figure development, and revision. SE participated in

manuscript writing, particularly focusing on section 4. All authors contributed to the article and approved the submitted version.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Choudhury A. Toward an ecologically valid conceptual framework for the use of artificial intelligence in clinical settings: need for systems thinking, accountability, decision-making, trust, and patient safety considerations in safeguarding the technology and clinicians. *JMIR Hum Factors*. (2022) 9:e35421. doi: 10.2196/35421
- Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. (2019) 6:94–8. doi: 10.7861/futurehosp.6-2-94
- Panch T, Mattie H, Celi LA. The “inconvenient truth” about AI in healthcare. *npj Digit Med*. (2019) 2:77. doi: 10.1038/s41746-019-0155-4
- Nguyen TT, Larrivé N, Lee A, Bilaniuk O, Durand R. Use of artificial intelligence in dentistry: current clinical trends and research advances. *J Can Dent Assoc*. (2021) 87:1488–2159. <https://jcd.ca/l7>
- Choudhury A, Urena E. Artificial intelligence in NICU and PICU: a need for ecological validity, accountability, and human factors. *Healthcare (Basel)*. (2022) 10(5):952. doi: 10.3390/healthcare10050952
- Bates DW, Levine D, Syrowatka A, Kuznetsova M, Craig KJT, Rui A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. *npj Digit Med*. (2021) 4:54. doi: 10.1038/s41746-021-00423-6
- Lebiere C, Blaha LM, Fallon CK, Jefferson B. Adaptive cognitive mechanisms to maintain calibrated trust and reliance in automation. *Front Robot AI*. (2021) 8: e652776. doi: 10.3389/frobt.2021.652776
- Okamura K, Yamada S. Adaptive trust calibration for human-AI collaboration. *PLoS ONE*. (2020) 15:e0229132. doi: 10.1371/journal.pone.0229132
- Shin D, Lim JS, Ahmad N, Ibrahime M. Understanding user sensemaking in fairness and transparency in algorithms: algorithmic sensemaking in over-the-top platform. *AI & Soc*. (2022b):1–14. doi: 10.1007/s00146-022-01525-9
- Shin D, Zaid B, Biocca F, Rasul A. In platforms we trust? Unlocking the black-box of news algorithms through interpretable AI. *J Broadcast Electron Media*. (2022c) 66:235–56. doi: 10.1080/08838151.2022.2057984
- Shin D. Expanding the role of trust in the experience of algorithmic journalism: user sensemaking of algorithmic heuristics in Korean users. *J Pract*. (2022) 16:1168–91. doi: 10.1080/17512786.2020.1841018
- Shin D, Kee KF, Shin EY. Algorithm awareness: why user awareness is critical for personal privacy in the adoption of algorithmic platforms? *Int J Inf Manage*. (2022a) 65:102494. doi: 10.1016/j.ijinfomgt.2022.102494
- Mcknight DH, Cummings LL, Chervany NL. Initial trust formation in new organizational relationships. *Acad Manage Rev*. (1998) 23:473–90. doi: 10.5465/amr.1998.926622
- Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. *Acad Manage Ann*. (2020) 14:627–60. doi: 10.5465/annals.2018.0057
- Dunning D. The Dunning-Kruger effect: On being ignorant of one's own ignorance. In: J. M. Olson & M. P. Zanna, editors. *Advances in experimental social psychology*. San Diego, CA: Elsevier (2011). p. 247–296.
- Ryan P, Luz S, Albert P, Vogel C, Normand C, Elwyn G. Using artificial intelligence to assess clinicians' communication skills. *BMJ*. (2019) 364:161. doi: 10.1136/bmj.l161
- Leibold LJ, Browning JM, Buss E. Masking release for speech-in-speech recognition due to a target/masker sex mismatch in children with hearing loss. *Ear Hear*. (2020) 41:259. doi: 10.1097/AUD.0000000000000752
- Jadczyk T, Wojakowski W, Tendera M, Henry TD, Egnaczyk G, Shreenivas S. Artificial intelligence can improve patient management at the time of a pandemic: the role of voice technology. *J Med Internet Res*. (2021) 23:e22959. doi: 10.2196/22959
- Xie B, Tao C, Li J, Hilsabeck RC, Aguirre A. Artificial intelligence for caregivers of persons with Alzheimer's disease and related dementias: systematic literature review. *JMIR Med Inform*. (2020) 8(8):e18189. doi: 10.2196/18189
- Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. *J Med Internet Res*. (2019) 21:e14658. doi: 10.2196/14658