A Roadmap Toward Building Trust in AI Environment: Research Agenda

Ik-Whan G. Kwon
ORCID: 0000-0002-8705-5172
The Center for Supply Chain Excellence,
Saint Louis University, Saint Louis, MO, USA

Sung-Ho Kim
Department of Healthcare Management,
Cheongju University, Republic of Korea

ABSTRACT

This paper explores the critical interplay between trust in AI technology and trust in its deployment, focusing on the gaps in interpersonal trust often overlooked in AI adoption. While AI enhances operational efficiency and decision-making, it struggles to replicate the emotional and psychological foundations of human trust. Drawing on interdisciplinary literature, the study highlights three trust dimensions—technological, organizational, and user-centric—and their misalignment, which fuels skepticism about AI reliability and ethical behavior. Despite AI's potential to transform industries, trust deficits persist due to concerns about transparency, accountability, and the emotional resonance of AI. The findings underscore that trust in AI must evolve as a complementary construct, bridging human and machine interactions without displacing interpersonal trust. The study proposes a framework emphasizing cultural and organizational adaptation, regulatory oversight, and user engagement to foster a balanced, sustainable AI.

Keywords: Trust, Interpersonal relationship, AI Technology, AI Deployment, AI Environments, oxytocin.

INTRODUCTION

Advent of AI and its deployment in industries has opened a host of research agenda. The most of research has been concentrated in two areas; reliability of AI technologies and its applications. The former mainly addresses on reliability and the level of trust of AI technologies whereas the later discusses the AI foundation; interpersonal trust between and among users of this technology. While discussion on reliability of technical side of AI has been active and widely debated, there has been relatively lack of discussion on interpersonal trust among and between users of this technology. Simple question is: can we build interpersonal relationship among and between users of AI technology? More succinctly, will and can AI replace interpersonal trust building process?

The purpose of this paper is to explore a road map where users of this technology (consumers) search for means to build interpersonal relationship with AI hosts. It is a challenging task to build a trust between the users and hosts (organization) as hosts were or to be replaced by AI technology leaving a big research gap in the trust building process.

LITERATURE REVIEW

Some argue that AI eventually replaces a traditional interpersonal relationship building process especially with institutions [1, 2]. Under this framework, relationship is NOT between "you and I", rather "institution vs. institution" or "I and institution". This line of research stream assumes that AI can build platform which require emotional aspect of human behavior; feeling, anxiety, emotion among and between decision makers during the negotiation/conversation phase. However, a critical review of this type of argument has failed to convince us how AI creates an environment in which trust can be formed, developed and transmitted to those who engage serious serial dialogues among interpersonal decision makers. Human behavior is unpredictable dictated and guided by instinct based on personality, culture and ever-changing surroundings.

Furthermore, trust is an emotional and psychological issues. The process of building interpersonal trust among us requires an extensive interaction to form a sustainable relationship. Trust which is based on weak relationship especially under the task-based relationship fractures easily as soon as the task ends [3]. Human is known to having a strong tendency and desire to trust someone because trusting someone makes us feel "good" and secure. Trust usually brings economic benefit as well by reducing transaction cost [4]. Experiments in neuroscience on human behavior reveal that having a sense of higher purpose in a trusting environment stimulates chemical agent in our body called "oxytocin". Oxytocin produces a sense of "happiness" of being trusted by others [5]. In short, trust "*someone*" (not "*something*") releases oxytocin that furthers desire to trust others and be trusted. It is interaction between and among "human" and not "machine" that produce a feeling and sense of comfort that makes us to pursue unknown journey toward rewards. Trust is generally regarded as a psychological mechanism for reducing uncertainty and increasing the likelihood of a successful (e.g., safe, pleasant, satisfactory) interaction with entities in the environment. When we trust someone, we expend less cognitive, physiological, and economic resources dealing with others. As of today, in spite of many untested claims, we are not yet in a position to accept that AI behaves just like "human", capable of building relationship that leads us to trust someone in our life. The New York Times recently reflects our thought on AI and trust as "artificial intelligence can fabricate entire videos and photos, and those can be hard enough to spot. But now A.I. can also make minute but important alterations in genuine still and moving images, further blurring the lines between real and fake [6].

Interpersonal trust also plays an important role in shaping our work ethics and productivity. For example, it is reported that compared with people at low-trust companies, employees at high-trust organizations have a 74% less stress and 29% more satisfaction with their lives, 106% more energy at work and 50% higher productivity and 13% fewer sick days and 40% less burnout. Compared with people at low-trust companies, people at high-trust companies report: 70% less stress, 106% more energy at work, 50% higher productivity, 13% fewer sick days, 76% more engagement, 29% more satisfaction with their lives, 40% less burnout [7]. Another similar research reveals that the collaboration index based on interpersonal trust accounted for 74% in the variation of performance index [8].

It is also reported that the complexities of changing business environments and the impact of the related decision-making by individuals show that collaborations are often characterized by a lack of trust and due to little information sharing. To address the lack of information sharing

and trust, some people look into artificial intelligence (AI) to transform the traditional forms of information exchange and trust-building process between engaging business partners. Despite the claim that AI capabilities can support supply chain collaborations [9, 10, 11], most firms so far do not engage in AI applications in and for their supply chains, indicating a lack of trust in AI technology.

The speed of AI technology advancement leaves many users behind unconvinced that information carried by AI is reliable and factual [12]. As [13] explains, while AI can assist in decision-making processes, key human variables, such as perception, feelings and trust cannot be programmed into AI systems. [14] even suggested that more interaction between human and AI is needed making AI platform as a mere complementary tool for human intelligent support system rather than a complete replacement of human knowledge.

By using AI technology into the collaboration discourse, however, the two dimensions of interpersonal and interorganizational trust will need to be extended to include another trust dimension: trust in AI technology. Specifically, for example, in supply chain management, whereas in a traditional collaboration, supply chain members build up trust to each other to provide the right information, the inclusion of AI results in shift to more extended trust approach, i.e. trust does not only involve interpersonal and interorganizational information partners, but it also requires trust in the (AI) technology to provide the right data and interpret the data in a meaningful way [15, 16]. Nevertheless, some writers even argue that AI eventually replaces interpersonal trust process altogether. For example, [1] argue that an increase in AI capabilities leads to a gradual reduction of interpersonal trust, i.e. trust in AI gradually replaces trust in interpersonal relationships. In this context, the interorganizational trust factor plays a more significant role, as trust in the AI technology itself depends on the transparency and explain ability of AI systems on a firm level [17]. Such a drastic statement by [1] may be plausible only if AI technology and its associated processes on a firm level are both transparent and explainable, of which many debates are still raging. Their five-stage development of AI toward trust assumes that AI infrastructure is built on a sound process. If the process has flaw, the entire argument becomes vulnerable [2].

However, as of now at least, literature has been relatively silent as to whether AI is capable of triggering releasing such emotional display (e.g., oxytocin) that promotes interpersonal relationships based on trust. In order to investigate interpersonal relationship using AI, some researchers attempt to establish a link between factors affecting trust in human relations and trust in AI technology. [18], for example, hypothesize that if one observes that the concepts of trust in humans and that in AI are highly overlapping, we might be able to transfer some of the knowledge from the concept of interpersonal trust to the concept of trust in AI. Their finding indicates that there is practically no relationship between these two constructs. Their finding is further supported by [19] who found that trust in humans and trust in AI share only small amounts of variance from about 4 to 11% variance depending on the organizational culture.

Another stream of research on trust in AI came from the organizational viewpoint. According to this approach, the degree of trust imbedded in the organizational culture. The degree of trust depends on the quality of the data used for training AI models [20]. But such training may be rooted in murky and ill-understood organizational routines/culture [17]. In addition, we have been warned that the systems based on AI may be developed by inexperienced teams in

organizations who unwittingly may introduce errors and biases [21] which tends to minimize/weaken the degree of trust using AI in a daily decision-making environment.

The above literature review convinces us that there is a sizeable gap between trust in AI technology and trust between and among users of this technology. A recent survey on the degree of trust in using AI reveals discouraging results as shown in the following [Table 1].

Table 1: The level of trust/reliability of using AI

Information from artificial intelligence, chatbots or search results, is reliable and factual	Percent of Returns (%)
Extremely confident	1
Very confident	4
Somewhat confident	30
Not very confident	38
Not at all confident	26
Total	**100**

(Source: The Associated Press-NORC Center for Public Affairs Research and USA Facts, 2024)

Less than 40% of the population seems to convince themselves that AI is reliable and factual. More than half of the population is not convinced that information contained in AI is reliable and trustworthy. The above finding appears to support those by [18, 19].

We always remind ourselves when making multifaced decisions using AI that AI agents are subject to human relationships, acceptance, ignorance and trust [22]. In addition, the degree and depth of using AI in business environment depends also on the level of organizational culture [23]. If engaging players (person or institutions) operate under different organizational culture, the speed and content of conversations via AI may be hampered or distorted to the degree that original intents may be misconstrued leading to an unexpected consequence. Trust is generally regarded as a psychological mechanism for reducing uncertainty and increasing the likelihood of a successful interaction with entities in the environment. When we trust someone, we expend less cognitive, physiological, and economic resources dealing with this entity. Research reveals that we may be not ready to take AI as a major decision-making tool considering persistent criticisms of the AI industry for insufficient trust-building measures related to ethical behavior when developing and implementing AI [24].

CREATING TRUST IN AI ENVIRONMENTS

Studies by [18, 19] appear to suggest that we need a new research path exploring an avenue to build trust between users of AI and institution hosting AI technologies. Furthermore, it is a challenging task to navigate trust building process under AI platform since the outcomes depend more or less upon culture of the institutions [19]. Nevertheless, AI will play a significant role in our life as worldwide investment in artificial intelligence forecast to reach $632 billion in 2028, according to a new IDC Spending Guide. In spite of uncertainty of trust in AI, it has been proved that AI contribution to our life, either personal or business, is beyond debate. It is not whether we should depend our life on AI platform, rather it is how and where AI should be deployed to improve our decision-making process reliable based on trust in technology and by users [15, 16]. AI technology itself does not provide the degree of trust. Rather it is the providers of this technology to convince us that they (providers of AI platform) are trustworthy.

In other words, it is the responsibility of the providers of this technology (interpersonal as well as institutional) to assure the recipients of information through AI is safe and trustworthy.

It has been a challenging task to convince users that AI is safe and reliable as shown in [Table 1]. Trust in AI requires two independently developed constructs (AI technology and users of this technology) align perfectly each other to generate comfortable feeling without any anxiety in using AI technology. In reality, as of today, these two constructs position themselves far from each other as shown in [Figure 1].
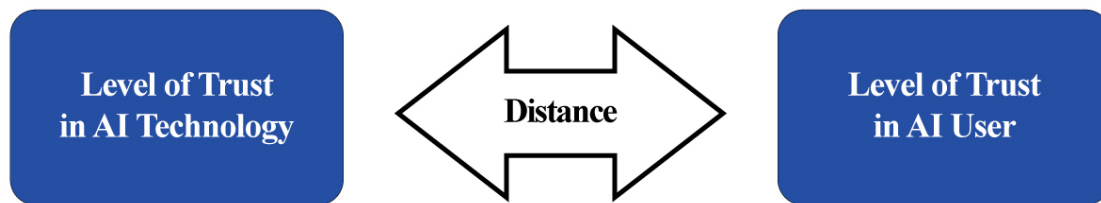

Figure 1: Distance between Level of Trust in AI Technology and Level of Trust by Users

Anxiety and fear are upon users of this technology as the distance between these two constructs is getting wider. Almost every day in IT magazine and some daily newspapers report new and advanced platform in AI technology. The more and faster technology advances, further it leaves users of AI behind. Unless concerted effort is launched to bridge these two constructs, the distance between these two constructs will be wider leaving many consumers into uncertain territories. An increase in anxiety of using AI eventually and to some degree expectedly invites government's intervention by way of legislating this industry. For example, *since 2019, 17 states have enacted 29 bills focused on regulating the design, development and use of artificial intelligence. Some of the highlights include* that the design, development and use of AI is informed by collaborative dialogue with stakeholders from a variety of disciplines. The essence of the legislations among others are to protect individuals from the unintended, yet foreseeable, impacts or uses of an unsafe or ineffective AI system and ensure that those developing and deploying AI systems are complying with the rules and standards governing AI systems. Industries which develop and market AI technologies are being held accountable if they do not meet regulations [25]. In 2023, European Union also issued guidelines on using AI technology. The EU wants to regulate artificial intelligence (AI) to ensure better conditions for the development and use of this innovative technology. The European Parliament priority is to make sure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly. It goes further insisting that AI systems should be overseen by people, rather than by automation, to prevent harmful outcomes. Parliament also wants to establish a technology-neutral, uniform definition for AI that could be applied to future AI systems [26].

Two-dimensional display of AI in [Figure 1] can be extended to accommodate users (consumers) of AI in their daily life as shown in [Figure 2]. Each area in [Figure 2] displays domain of trust in AI platform. As explained in the previous section, trust in AI requires three players; trust in technology (area A), trust in AI host (area B) and trust by AI users (area C). The intersection of area A, B and C creates Zone D where all three players trust AI and feel comfortable to use AI technology.
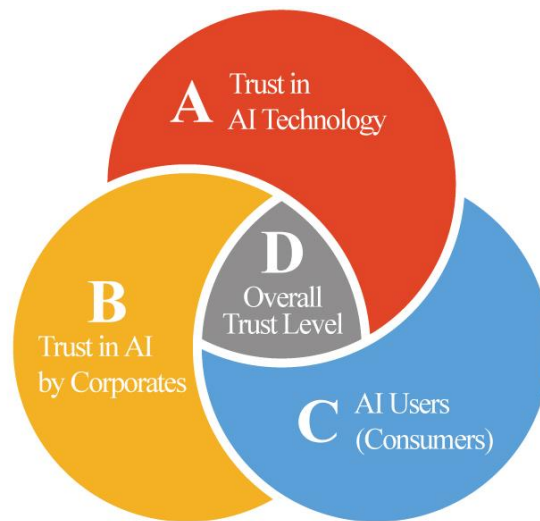
Figure 2: Triangle Intersection

Consumers/users of AI (Part C) rarely interact with AI technology developer (Part A) in their daily activities in the market. Instead, consumers interact with entities/corporates which host AI platform in product design, manufacture and distribution of their products. However, under the AI platform, there are no "persons" to engage person-to-person conversations (interpersonal conversation). Only avenue available for consumers to build interpersonal trust with corporate is through trust in products they market. Brand names of products represent symbol of trust under the AI platform.

Creating the institutional reputation, in turn, rests on the agents (people) in the institution who promote the reputation of the products in the market. From this scenario, it follows that trust of brand names in organization is equivalent to trusting *people* in the organization. It can be argued, therefore, when we purchase products, we in fact trust people in the organization who designs, produces and markets products for the consumers. People in the organizations are well versed in AI technology enough to invest their resources in this technology. In this sense, AI is just a tool like any other technology with exceptional ability of forecasting, planning and managing daily operations from macro to micro in details. The framework of creating trust environment in an AI environment is very similar to a typical trust forming process in interpersonal trust forming process.

Trust forming process with organization requires information on quality of their products in the market. One measure of quality of product among many could be numbers of recalls of their products from the market in a given year. For example, Risk and Insurance reports that in 2023 alone, there are 975 medical device recalls, 916 automotive recalls, 517 pharmaceutical recalls, 506 food and drinks recall [27]. In many instances, social medias and major news networks report on a daily basis information on product recalls. Consumers became well informed and make their choice accordingly. Market disciplines by consumers is an indicative of degree of trust toward the organizations under AI platform. Interpersonal trust building framework under AI platform is replaced by two agents; consumers and organization. Using [Figure 2], the intersection between Area B (providers) and Area C (AI users) can be expanded as information in social media on product safety/recalls becomes more active and visible. Extending this

scenario, it is possible that consumers dictate trust building process with corporates under AI platform.

SUMMARY AND CONCLUSION

This paper started with a question; can we build interpersonal relationship among and between users of AI technology? It has been argued that AI replaces interpersonal trust building process between consumers and those who produce, market and distribute their products. When consumers attempt to contact institutions, AI navigates the conversational traffics to "machine" leaving no human interaction. Consumers are left no choice but being forced to make conversation with "machine". Trusting persons in an organization under AI is being replaced by trusting their products. Product stands for a symbol of corporate's reputation and a part of organizational culture. The above scenario implicitly assumes that consumers indeed trust people in the corporate organization through their products. It is, therefore, no surprise to see so many changes in people in C-Suite based on their products and services especially those in the highly recognizable corporates and public institutions. For example, through October, 2024, more than 1,800 CEOs from major corporations have announced their departures, according to data global outplacement firm Challenger, Gray & Christmas.

References

[1]   Hengstler, M., Enkel, E., and Duelli, S. (2016). Applied artificial intelligence and trust-The case of autonomous vehicles and medical assistance devices. Technological Forecasting and Social Change, 105; 105-120.

[2]   Weisz, E., Herold, D.M., Ostern, N.K., Payne, R. and Kummer, S. (2024). Artificial intelligence (AI) for supply chain collaboration: implications on information sharing and trust. Online Information Review, 49(1); 164-181. Available at: http://www.emeraldinsight.com/doi/10.1108/OIR-02-2024-0083

[3]   Kwon, Ik-Whan and Kim, Sung-Ho (2024). Relationship between economics of trust and transaction cost: A brief exposition. International Journal of Business & Management Studies, 5(2). Available at: DOI: 10.56734/ijbms.v5n2a5.

[4]   Kwon, Ik-Whan, Hamilton, John and Hong, Seock-Jin (2012). Trust and Transaction Cost in Supply Chain Cost Optimization: An Exploratory Study in Inter-Organizational Information Systems and Business Management: Theories for Researchers. Kishor Vaidya edited IGI Global Publication: 70-82. Available at: DOI: 10.4018/978-1-60960-768-5.ch005.

[5]   Carter, Sue (2014). Oxytocin pathways and the evolution of human behavior. Annual Review of Psychology, 65; 17-39. Available at: DOI: 10.1146/annurev-psych-010213-115110.

[6]   Thompson, Stuart A. (2024). Which Parts of These Images Are A.I.-Generated? The New York Times, December 27, Available at: https://www.nytimes.com/interactive/2024/12/27/technology/artificial-intelligence-generative-fill-photoshop-openai.html.

[7]   Zak, Paul J. (2017). The Neuroscience of Trust; Management behaviors that foster employee engagement. Harvard Business Review, January-February; 84-90.

[8]   Simatupang, T. and Sridharan, R. (2004). Benchmarking Supply Chain Collaboration: An Empirical Study. Benchmarking: An International Journal, 484-503.

[9]   Baryannis, G., Validi, S., Dani, S. and Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. International Journal of Production Research, 57(7); 2179-2202. DOI: 10.1080/00207543.2018.1530476.

[10] Mikl, J., Herold, D.M., Cwiklicki, M. and Kummer, S. (2021). The impact of digital logistics start-ups on incumbent firms: a business model perspective. The International Journal of Logistics Management, 32(4); 1461-1480. Available at: DOI: 10.1108/ijlm-04-2020-0155.

[11] Pournader, M., Ghaderi, H., Hassanzadegan, A. and Fahimnia, B. (2021). Artificial intelligence applications in supply chain management. International Journal of Production Economics, 241, Available at: DOI: 10.1016/j.ijpe.2021.108250.

[12] Collins, Banjamin (2024). Contextualizing Ethics and Equity of Healthcare Artificial Intelligence Through a Lifecycle. The Conference on Bridging disparities in health care using AI, Saint Louis University, December 6.

[13] Subasi, Abdulhamit (2020). Use of artificial intelligence in Alzheimer's disease detection. In Artificial intelligence in precision health, London: Academic Press, 257-278. Available at: DOI: 10.1016/B978-0-12-817133-2.00011-2.

[14] Lainjo, Bongs (2024). Integrating artificial intelligence into healthcare systems: opportunities and challenges. Academia of Medicine, Available at: DOI.org/10.20935/AcadMed7382:1-13.

[15] Chen, R.R., Chen, K. and Ou, C.X. (2023). Facilitating interorganizational trust in strategic alliances by leveraging blockchain-based systems: case studies of two eastern banks. International Journal of Information Management, 68; 102521. Available at: DOI: 10.1016/j.ijinfomgt.2022.102521.

[16] Shin, Donghee (2021). The effects of explain ability and causality on perception, trust, and acceptance: implications for explainable AI. International Journal of Human-Computer Studies, 146; 102551. Available at: DOI: 10.1016/j.ijhcs.2020.102551.

[17] Makariusa, E.E., Mukherjeea, D., Foxa, J.D., FoxbRising, A.K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial. Journal of Business Research, 120; 262-273.

[18] Lukyanenko, R., Maass, W. and Storey, V. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. Electronic Markets, 32; 1993-2020.

[19] Montag, C., Becker, B. and Li, B. (2024). On trust in humans and trust in artificial intelligence: A study with samples from Singapore and Germany extending recent research. Computers in Human Behavior: Artificial Humans, 2(2); 1-7.

[20] Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P. and Aroyo, L.M. (2021). Everyone wants to do the model work, not the data work: Data Cascades in High-Stakes AI. ACM Digital Library, Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 1-15.

[21] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6); 1-35.

[22] Henrique, B.M. and Santo Jr., E. (2016). Trust in artificial intelligence: Literature review and main path analysis. Computers in Human Behavior: Artificial Humans, 2(1); 2-13.

[23] Noor Al-Ma'aitah (2024). Impact of organizational culture on healthcare supply chain resilience in Jordan: Moderating role of technology integration. Problems and Perspectives in Management, 22(4); 68-82. Available at: DOI:10.21511/ppm.22(4).2024.06.

[24] Vardi, M.Y. (2022). ACM, ethics, and corporate behavior. Communications of the ACM, 65(3); 5-5. Available at: DOI: 10.1145/3516423.

[25]   The Council of State Governments (2023). State Leaders Convene, Collaborate at the 2023 CSG National Conference in Raleigh. State Leaders Convene, Collaborate at the 2023 CSG National Conference in Raleigh, December 6.

[26]   European Parliament (2023). EU AI Act: first regulation on artificial intelligence. August 6.

[27]   Risk and Insurance (2024). Product Recalls Reached Five-Year High in 2023, Sedgwick Report Says. March 14. Available at: https://riskandinsurance.com/product-recalls-reached-five-year-high-in-2023-sedgwick-report-says/.