Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8 July, 2021: 6417 – 6424

Research Article

The Role Of Trustworthiness In The Adoption Of Ai-Based Systems

Dr. Ameeta Fernando, Dr. B. Aiswarya, Dr. A. Siluvairaja,

Loyola Institute of Business Administration

Abstract

The objective of technology is to enhance the functioning of the various human task sets in a more efficient and timely manner. This is the basic purpose of any computer/software application that is in existence. The idea is to minimize errors, reduce time, effort and even risk, and improve quality of life. Artificial intelligence, in particular, aims at replacing human effort with machine intelligence in tasks that are either repetitive or risky or both. Despite all this, however we look at it, AI is only a tool at best and requires human interaction in some aspect or the other.

There are several studies that focus on attributes that bring together human nature and technology. For one, technology today has been identified as a leading cause of stress. A typical example is the use of social media platforms, the purpose of which was to bring together people irrespective of their age, ethnicity and geography. Ideally, this should have been a good thing. But given the very fabric of human nature and the attributes of technology, this has become more complicated than it should be. Experts have opined that technologies can dominate the lives of people, by pressurizing them in terms of time and effort, which in turn could have physical and psychological repercussions that arise from such stress (Keith Hampton, 2015).

I. Introduction

Another aspect is the trustworthiness of technology itself. In order for AI systems to realize their full potential, they need to earn the trust of human users. In order to understand the concerns that may pose a threat in the use of technology components like AI in the future, it is imperative that we analyze the aspects of human behavior and those of technology components in tandem. That is the reason why this kind of study, which is termed as Cyberpsychology, has been gaining momentum in the last few years, though it had been a subject of interest even before the birth of the new millennium. This paper aims to unravel the importance of perceived trustworthiness of AI systems among users in the adoption of the same. It also aims to study the factors that contribute to the trustworthiness of AI systems (or the lack thereof) in the Indian context.

In order to identify the factors that cause concern and mistrust with respect to AI systems, Pegasystems conducted a research study with 6,000 users of AI systems from Australia, North America, the United Kingdom, Germany, France, and Japan about their views on empathy with regard to AI systems (Pega Systems, 2019). The results of the study are as follows:

- While deciding about bank loan approvals, 68% of the respondents trust a human more than AI.
- Another interesting finding was that 69% of the respondents would rather tell the truth to a human rather than an AI.

However, 40% of the respondents agree that AI has the ability to enhance customer service and relations. Thus, while consumers may be open to the use of AI systems, the deterrent that organizations will have to worry about while pondering the implementation of more AI systems in the future, would be the way and means of increasing the faith of consumers in these AI technologies. To do that, we need to figure out possible reasons for the lack of trust in AI systems. In their report on 'Principled Artificial Intelligence', the authors identify 8 key themes as a substantial aspect of their findings - Privacy, Human control of technology, Accountability, Professional responsibility, Safety and Security, Fairness and non-discrimination, Transparency and Explainability, and Promotion of human values. One of the key themes is the promotion of human values (Fjeld, 2020). In their paper on 'Trust and Artificial Intelligence', the authors claim that trust, as a fundamental human trait, serves as a mechanism for reducing complexity (Brian Stanton, 2021). While the authors discuss the various factors that lead to trusting or distrusting, they state that while trustor factors include 'general willingness to rely on other people', the factors that contribute, from the trustees' end, include ability, benevolence and integrity, or at least the trustor's perception of these factors. While ability is the possession of some context- or domain-specific skills and benevolence refers to the sense of goodwill that the trustee has towards the trustor, integrity refers to the maintenance of a set of values or principles to which the trustor adheres. In the context of AI systems, however, integrity and adherence to values are attributes that are left to the implementers of the system in question. And these may also be the reasons why users are not completely trusting of AI systems.

Aside from the potential for AI-based systems to improve decision-making, reduce organisational resources, and root out human biases, their unexpected effects have been largely ignored so far. Researchers are divided on whether the perceived trust in AI enabled services is consistent and whether incorrect advice is a hindrance to system use or if these systems are blindly trusted. For instance, while transparency about a recommendation system's accuracy levels increases confidence and reliance when given wrong guidance, it could have a different impact when given correct information. Because the efficacy of transparency techniques in AI is linked to system performance, they should be used with prudence.

Artificial intelligence-based information systems are suffering from opaqueness, limited robustness and reliability. They offer outcomes that are often not fully predictable, unexplainable and further, bring along a certain probability of inaccuracy. Against the backdrop of these challenges, trust plays an important role in understanding the adoption and use of AI-based systems. Lack of trust towards such systems' recommendations can impede successful adoption and deployment.

When exploring relevant user outcomes, most empirical work has black boxed the nature of AI-based systems and overlooked the validity of its output. The distinction between conceptions of trust and related behavioural consequences is now blurred in studies. In the field of artificial intelligence, inaccurate system guidance has been identified as a major issue that has received insufficient attention. If an error in algorithmic guidance causes AI-based systems to be misused or abandoned, the question of how to design for interaction with failure prone information systems (IS) emerges. So far, there hasn't been much research into how specific design characteristics mitigate possible over- or under- reliance on AI-based guidance. It would really help to get a more nuanced understanding of both the antecedents (the underlying technical

system and the accuracy of advice) and the effects of AI-based guidance.

Recent research findings help us better understand whether people trust AI-based advice and if this acceptance is impacted by the system's accuracy. By distinguishing between trust and dependence on AI-based guidance, we give a more sophisticated explanation of perceptual and behavioural results. We intend to contribute to the current research corpus on algorithmic aversion and appreciation by throwing light on these phenomena and experimentally assessing proposed solutions for AI trustworthiness.

ALGORITHMIC AVERSION VERSUS APPRECIATION

Two dominant research streams, namely algorithmic aversion and algorithmic appreciation proposed two conflicting conclusions regarding the reliance on AI-based advice. Algorithm appreciation claims that users would rather rely on advice stated to come from an algorithmic source, as compared to a human. The existing research body has identified numerous reasons for decreased trust in algorithmic advice, including the desire for perfect prediction and human confidence in their own reasoning.

While most studies support the notion of algorithmic aversion, a more nascent research stream observed an exaggerated appreciation of AI based advice. Algorithmic appreciation has been found in time-critical situations and has been explained by humans attributing more objectivity and rationality to algorithmic advisors compared to human judgment. Algorithmic advice can be relied upon more than humans to make decisions, but little is known about when it is over

or under-utilized. Liel and Zalmanson's 2020 paper was one of the few that investigated the effects of erroneous AI on advice reliance in the context of a simple judgment task where algorithmic mistakes were quite apparent to the user (Liel Y., 2020).

According to established trust theories, trust can be described as a human reaction to reduce complexity although an undesirable outcome is possible. Users presume a favourable behaviour of the IS despite the uncertainty of the algorithms providing erroneous recommendations. Algorithmic aversion predicts a general distrust in AI- based systems, but we suggest that perceptions of trust in and reliance on advice will be influenced by the correctness of the algorithmic advice. Literature generally offers mixed results regarding the preference for source of advice, yet also the implications of erroneous algorithms. Considering a decision task defined by an objectively measurable outcome, users should be able to detect incorrect advice.

Performance and accuracy of AI-based IS vary from system to system and are dependent on numerous factors like quantity and quality of data and labelling of data. The nature of machine learning (ML) based systems, by default, introduces outcomes that are not fully predictable nor explainable, and bring along a certain probability of inaccuracy. Improving accuracy rates and model performance in practice can be achieved by building high quality datasets, but the improvement of accuracy rates is a trade-off between resources and performance.

An increased understanding of how systems work leads users to assign increased capabilities to the AI based IS, and thus trust it more. Yeomans et al. gather evidence on how explanations on the underlying workings of an algorithm decrease algorithmic aversion (Yeomans, Shah, Mullainathan, & Kleinberg, 2019). Berger et al. (2021) suggest exploring transparency as a further moderator to consider when studying the impact of erroneous algorithm advice (Berger, Adam, Rühr, & Benlian, 2021). A research gap still

exists in measuring how to enhance trust in algorithms. Transparency has been mentioned as one of the key dimensions in establishing trustworthiness in AI-based systems.

FOCUS GROUP DISCUSSION

A study was carried out in this regard, with respondents who are informed about the use of Artificial Intelligence and Machine Learning through Predictive Analytics. The respondents were asked to discuss the topic 'Trustworthiness of AI systems' and document their thoughts and arguments. The exercise was conducted with 50 respondents, all of them made aware of the basic concepts used in Predictive Analytics and Business Intelligence. There was no limit on the points that each respondent could come up with. They were asked to individually identify the perceived causes for trusting or not trusting an AI system. The respondents were given access to information from the internet for gathering facts that could favor their arguments. The points that were documented in favor of trusting AI systems mostly had to do with the convenience, efficiency, handling of enormous data, timeliness of processing, precision in the execution of processes and removal of risk from manual systems. The points that were put forth against the trustworthiness of AI systems were initially categorized under 17 headings and were then scrutinized for relevance and rationality. While there were many reasons cited (mentioned below), some of them were not found to be very relevant to the trustworthiness aspect of AI systems and hence were not considered for the study:

- The loss in jobs and opportunities for humans due to the emergence of AI technologies
- The high costs of implementation, maintenance, upgradation and upscaling
- The rendering of humans lazy, lacking in experiences and over-dependencies on machines
- The robbing of desirable activities such as driving a car (as in the case of unmanned automobiles)

Based on the discussion, four major points emerged as valid and relevant reasons as to why AI systems are not/cannot be trusted by users. They were:

1. AI is not completely accurate/error free

2. AI does not possess Emotional Intelligence

3. AI is not completely private and secure

4. AI in not naturally focused on ethics

1. AI is not completely accurate/error free (Competence)

The responses in this category covered several aspects of automated systems and specifically brought to light the following concerns that users have regarding these systems:

GIGO: The system is as intelligent as the person who designed and developed it, sans the human advantages of emotional intelligence and value systems. Therefore, the lack of accuracy itself could be built in by faulty algorithms and aberrant data.

BIAS: This could also include the various kinds of biases that human thinking suffers from. These points were explained with the examples of Amazon's AI enabled HR Recruitment system and racial biases in machine learning systems.

The presence of false positives and false negatives, the correctness of the process and data sets were all discussed under this cause for lack of trustworthiness. Concerns regarding the magnitude of the repercussions that could result from these errors, the consequences of such errors in real-time, especially

hard real-time systems such as healthcare systems and automobiles (quoting the example of the accidents in unmanned vehicles) seem to be a real concern for prospective consumers of this kind of system.

2. AI does not possess Emotional Intelligence (Human Values)

This was the second most common reason cited by the respondents for not being able to trust AI systems as much as they could trust human counterparts. Emotional intelligence has been discussed in terms of empathy, taking into consideration extraordinary situations while making decisions that could affect the future of individuals, remorse for actions gone wrong, taking responsibility for wrong decisions and their consequences, creativity in arriving at alternate solutions and assessment of these alternatives, intuition that goes beyond logical reasoning at certain times, working based on instincts that have proved correct in previous situations and so on. According to the respondents, these are, by and large, important dimensions of decision making and, therefore, reasons for why humans can be trusted more than machines. While advancements in AI and ML claim that Pepper Robots can determine human emotions and moods with certain biometric data, at least in the Indian context, according to the group, consumers are not yet ready to trust these systems, especially in decisions pertaining to sensitive issues.

3. AI is not completely private and secure (Responsibility)

The fact that the information about users is shared across multiple platforms is detrimental to the perceived trustworthiness of AI systems. Take a chatbot, for example. When we furnish our choice while exploring the options for a service or a product that we wish to buy, the fact that user responses are consolidated and sold to marketing agencies for promoting the product or service, which in turn may be used by marketing agencies to promote other brands who are also their customers, leaves the customers with a feeling of being let down by the system. When we browse the internet for a certain service or product, it ends up as an advertisement on our Facebook pages. It does not take an expert in today's age to figure out how this happened. We live in an era of state-of-the-art digital marketing tools and techniques which use our data for profit maximization by several parties, but that may not be taken well by all consumers. Well-informed consumers may slowly wean themselves out of such technologies and go back to traditional methods of purchasing goods and services without the help of AI enabled systems. Another point that came to light was the cybercrime factor. Users are cynical about using AI systems since they feel that "unethical yet software savvy individuals" may misuse data furnished by consumers on AI enabled devices and platforms to misuse their data. The fact that entire bank accounts could be swindled within a few seconds of accessing security credentials also does not help AI systems gain customer trust. The lack of a controlling authority which can draw a line when it comes to data sharing between marketing agencies, the not-so-transparent nature of legal implications pertaining to data privacy/security and the lack of awareness of mitigation procedures all contribute to the lack of trust in AI systems.

4. AI is not naturally ethical (Integrity)

When it comes to ethical decision making and the importance of the human value system in determining our choice from alternatives, there is not much that machines can do by themselves. If a machine is taught to be unethical, it can cause the worst possible harm, as met with in the case of AI technologies being used by terrorists to realize their missions. The fact that machines lack emotions and empathy makes them less trustworthy, especially when it comes to personal and sensitive situations. Though there are new innovations in this area to recreate the human value system in machines, the makers of AI solutions have a long way to go before they can convince users that in every given situation (where

there are infinite possibilities), the machine will render ethical solutions every single time. Another factor that contributes to this belief is that machines are only as ethical as their creators and that if the designers and developers of these systems are unethical, so will the machines be. While we have a choice to stay away from people who do not share our same ethical space, there may be no choice when it comes to machines since people naturally are more likely to believe machines with mathematical and statistical intelligence but may not know where the line is drawn with respect to ethics. Users may get carried away until they reach a point where they realize the ethical/unethical perspective of a certain process. The lack of remorse or guilt after an erroneous process or result is another aspect that works against trusting an AI system. If a machine makes a mistake, it does not "feel" anything and if left as it is, will cause the same damage if used again. Normal humans differ in this aspect, and to rebuild machines incorporating "remorse" is going to be a challenge even in the case of sophisticated machine learning algorithms.

Based on these broad areas that emerged from the discussion, a questionnaire was designed to further probe and study the perceived trustworthiness of AI systems. The objective of the questionnaire was to elicit responses that could help identify factors that affect the perceived trustworthiness of AI systems among users. The target respondents are consumers who have a moderate understanding of the working of AI systems who, however, may not have in-depth knowledge of the technical implementation of these systems. The questionnaire was aimed at aiding this study that focuses specifically on what users perceive as AI's adherence to human values like empathy and honesty, and the role that these factors play in the overall adoption of AI systems. This comes at a time when the world is marching towards the implementation of end-to-end AI solutions in all aspects of information processing. This study will help provide valuable insights to AI solution developers, organizations that implement these solutions, and consumers and would help them get an idea of what the future would look like, given the perceived trustworthiness of AI systems.

SIGNIFICANT FINDINGS

1. Only 18% of the respondents confidently agree to depend on AI systems for life dependent functions such as healthcare
2. Only a little more than half of the respondents feel that AI systems are not biased
3. More than 61% of the respondents feel that AI systems can never be taught to have human values like empathy or honesty
4. Only about 55% of the respondents agree that AI systems are always accurate
5. When asked if they believe that AI systems can take correct decisions in extraordinary situations, only a little over 20% of the respondents agreed
6. Likewise, only less than 38% of the respondents believe that AI systems can be taught to learn from the consequences of bad decisions
7. A whopping 72% of the respondents feel that they cannot freely give their personal information to AI systems fearing the compromise on confidentiality
8. For a similar question, an astounding 82% of respondents opine that those enterprises that capture user data through AI systems might easily sell their data to others

9. More than 55% of the respondents will trust a human more than an AI system when it comes to revealing confidential information

10. Over 70% of the respondents fear that their data could be misused by an AI system

TO CONCLUDE

We live in an age where there may not be a return from where we have reached in terms of technology. The way forward does not look bright for AI either if consumers are not willing to trust these systems. Human intuition and outlook on aspects pertaining to their own security and well-being, at some point of time, may far outweigh the comforts and conveniences that technology provides. It is the utmost duty of AI technology providers therefore to ensure that these doubts and uncertainties are clarified in the minds of the consumers so that they may favour the use of AI technologies for a prolonged duration. It is far better to slow down, pick the pieces up, put them together and continue the race than to finish in a hurry only to realize that the objectives were not met. Issues pertaining to confidentiality, security, transparency, credibility, integrity and empathy need to be sorted out here and now. These aspects should be built into every AI-based system and should not be incorporated after the effect. If we are looking for a long-term adoption of AI-based systems, the results mentioned in literature as well as from our own study show an immediate and urgent need to address the afore-mentioned concerns at the earliest and, thereby, bringing about a change in the way users view AI-based systems.

Bibliography

1. Ancis, J. R. (2020). The Age of Cyberpsychology: An Overview. American Psychological Association: Technology, Mind and Behavior, 1.
2. Banavar, G. (2016). What It Will Take for Us to Trust AI. Harvard Business Review Digital Articles, 11/29/2016, 2-4. 3p.
3. Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn. Business and Information Systems Engineering, 55– 68.
4. Brian Stanton, T. J. (2021, March). Trust and Artificial Intelligence. Retrieved from National Institute of Standards and Technology:
5. https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8332-draft.pdf
6. Canty, D. (2017, July 25). Cyberpsychology and How AI Affects Humans. Retrieved from DZone: https://dzone.com/articles/advancing-human-cyber-psychology
7. Fjeld, J. N. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Berkman Klein Center for Internet & Society.
8. Harjit Sekhon, C. E. (2014). Trustworthiness and Trust: Influences and Implications. Journal of Marketing Management, Vol. 30, , 3–4, 409–430.
9. Keith Hampton, L. R. (2015, January 15). Social Media and the Cost of Caring. Retrieved from Pew Research Center: https://www.pewresearch.org/internet/2015/01/15/social media-and-stress/#fn-12625-1
10. Kerasidou, A. (2021). Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. Journal of Oral Biology and Craniofacial Research, 612 - 614.

11. Liel Y., a. Z. (2020). What If an AI Told You That 2 + 2 Is 5? Conformity to Algorithmic Recommendations. Association for Information Systems.
12. Nadine Schlicker, M. L. (2021). Towards Warranted Trust: A Model on the Relation Between Actual and Perceived System Trustworthiness. MuC '21: Mensch und
13. Computer 2021 (pp. 325–329). Ingolstadt Germany : Association for Computing MachineryNew YorkNYUnited States.
14. Pega Systems. (2019, July). AI and Empathy: Combining artificial intelligence with human ethics for better engagement. Retrieved from https://www.pega.com/: https://www.pega.com/system/files/resources/2019-11/pega-ai-empathy-study.pdf
15. Rossi, F. (2019). BUILDING TRUST IN ARTIFICIAL INTELLIGENCE. Journal of International Affairs Editorial Board, 127-134.
16. Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making Sense of Recommendations. Journal of Behavioral Decision Making (32:4),, 403–414.