NSF/ASME Design Essay Competition 2022

Undergraduate Submission

St. Louis, Missouri

Commented [BM1]: I put this here to indicate it's an undergrad submission. I am not sure if it is suppose to go here. If the template doesn't have that, then remove it.

Trust Optimization In Human-AI Interaction

Optimizing worker trust in AI systems to increase industry metrics

Advisor: Dr. Beshoy Morkos

The University of Georgia

Cody Carroll

Fatemeh Mozaffar

Logan Smith

Trust Optimization In Human-AI Interaction

Optimizing worker trust in AI systems to increase industry metrics

Abstract

With the rise of practices such as of Industry 4.0 and mass personalization in the manufacturing and design setting, AI systems are becoming more prevalent in day-to-day operations. However, due to the black-box nature of AI agents, fear and lack of trust has restricted the benefits gained by these systems. The manufacturing and design industry is at risk of limiting its potential in profits and productivity due to a lack of understanding in AI interactions. As a consultant for a high-tech manufacturing enterprise, we propose the optimization of trust in human-AI interactions to increase effectiveness in a manifold of industrial facets. While there are many challenges to navigate when exploring human-AI interactions, with the implementation of educational programs and an increase in exposure towards AI for students and employees, industries can optimize the trust in human-AI interactions and ensure an increase in effectiveness for time, productivity and profits.

1. Introduction

AI systems able to learn from data and generate predictions are increasingly leveraged in a myriad of products and services deployed in domains such as healthcare diagnostics, news broadcasting and jobs as well as financial markets[1]. Artificial intelligence (AI) can be described as the ability of a machine to perform cognitive functions that we associate with the human mind and tasks that require human intelligence such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity [2, 3].

AI technologies are associated with a plethora of benefits that range from greater efficiency, faster and more accurate results, reduced error rate at the process level, to more effective and improved strategic outcomes at the organization level. AI has been developed to perform all manner of tasks, but less attention has been on research for AI capable of social or interpersonal interactions with human collaborators [4]. Understanding how users account for AI advice is an important aspect of human AI interaction that has been under-explored, although Human-AI collaboration is critical in the adoption and operation of algorithmic systems [5, 6]. The field of Human-AI teaming seeks to create the right balance of human and AI capabilities that are suited to the context, mission, and operational dynamics within which the system performs. Human-AI teaming focuses on effective integration of humans and intelligent machines by understanding the relative strengths and weaknesses of each. Their combination can utilize each other's strengths while compensating for weaknesses[2]. Therefore, we decided to focus on Human-AI interactions and the impact of user trust on industry.

People use the expression trust in a manifold of different ways to describe a variety of human affairs. In certain cases, they can also express a trusting attitude towards non-human agents. Trust

is thus a construct enriching a wide variety of relationships people establish, nurture, and interrupt in everyday life. E-trust is an interesting approach to trust in digital environments and in the presence of artificial agents[1]. While this mitigates issues related to safety, one of the biggest barriers to adoption remains the black-box nature of current AI systems that limits people's trust in the AI's suggestions[4, 7].

The role of AI in our society and the development and presence of AI devices in consumers' everyday life is regarded with both fascination and fear. AI devices may increase the consumer's quality of life by taking over certain activities they do not like to do[8]. The most dominant theme related to the relational approach in the AI literature relates to trust in the AI system. Trust in AI may depend on the embodiment of the system and the level of machine intelligence, mutual concern, a shared sense of vulnerability, and faith in the competence of the system[9].

Psychological tools and approaches are often used to modify and measure human-AI interactions. For example, when designing the interface for AI agents, it is important to understand the user's goals and intentions, how easily they can manage issues, the workplace social interactions of said user, and their physical ability [10]. Program inputs and outputs, or other methods by which a person may interact with an AI could benefit from a thorough understanding and implementation of psychological tools.

We at "CDS[1] ", in consultation with a high-tech, global manufacturing enterprise, have been tasked with identifying a key characteristic for companies to adopt to position themselves to be a high-tech global design and manufacturing enterprise in the year 2040. We recognize that with the increasing prevalence of computer systems and AI agents in an industrial setting, understanding the relationships between human workers and AI agents is vital to the health of an enterprise. We hypothesize that understanding and increasing the trust between human and AI will increase industry effectiveness going into the year 2040.

2. Predicted Improvements

There are many factors that can contribute to a company's success in design and manufacturing. A company's success will, however, depend mostly on time management and productivity. Through the study of the effects of artificial intelligence on these two fields as well as the challenges facing industry and its results, we can increase the trust in artificial intelligence.

2.1 Effectiveness

In Manufacturing and Design, one of the largest factors that affect a company's success is how well time is managed. Large amounts of research are performed in industry to minimize system process times and time to market for products. The implementation of AI into industry has been proven to save time in numerous different applications. AI systems have already been shown to reduce time to market as well as create smarter automated testing [11], when building customized software packages. One bottleneck factor when minimizing process times that involve AI agents is the interaction with human workers or users. Many factors contribute to this, such as system proximity to the user, ease of operation, and the trust between the human and AI. Human beings are inclined to distrust things they do not understand, and this can be translated in industry as questioning or second-guessing actions taken or decisions made by AI agents. While blind trust

[1] Collaborative Design Solutions

can introduce error to a system, causing an inability to recognize incorrect decisions made by the agent, a lack of trust in AI agents is equally as inefficient, both of which are depicted by Case B and C in **Figure 1**. The optimization of trust, where human users are able to recognize when an agent is wrong while not second-guessing correct decisions, will assist in minimizing time wasted. However, it should not be suggested that machine errors during the human interaction are a waste of time as is in Case A are inefficient since they inevitably yield finer tuning of the AI and increase the agent's ability to make correct decisions in the future.

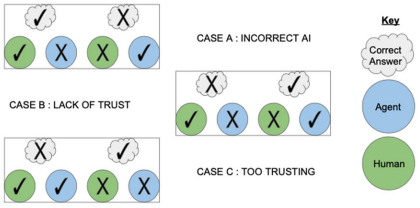

***Figure 1**: Visual Representation of Human-AI Iterations*

This aspect of the human AI interaction is conceivably the largest hurdle to the implementation of these systems to many of the following fronts. The possibility of joint false positives and false negatives as the human AI interaction unfolds. Inevitably both man and machine will introduce errors into the process; however, it is reasonable to assume that the time saved in both contexts where man and machine agree, and where man realizes some issue with the logic of the machine would significantly enhance both the manufacturing and design aspects of the process. There are several applications where agents currently interact with the industrial environment, but the rationale for finely calibrating user trust is highlighted in some contexts to be explored.

Furthermore, AI's ability to sense connections in ways that humans might not be able to lend credence to its usage in sensing potential issues within the scope of one's supply chain. After the impact of COVID-19, companies around the globe have been reassessing the level of depth to which they source parts and components of their design; however, the interconnectedness of the 21st century makes this task understandably difficult for humans alone. Having AI tools to diagnose where issues might arise carries understandable benefits, but the hurdle of ensuring human AI interaction maintains the balance previously mentioned remains an issue.

AI is considered an engine for productivity and progress; on the other hand, AI entails societal upheaval, since AI services are perceived as inherently risky due to the lack of human interaction; thus, a preference for personal contact with a service employee seems to prevail[3, 9]. When AI is successfully assimilated into an organization, employees are more likely to experience better psychological outcomes and higher performance. Research suggests that effective socialization can lead to greater productivity as well as higher commitment and lower turnover from employees. Relatedly, a greater understanding of AI can influence emotional states and career satisfaction. In particular, AI is expected to improve productivity by automating routine tasks and allowing employees to focus on work that adds more value to the organization[9]. Additionally, Employees are supported in the decision-making process, strategic decisions are facilitated, and human-AI

collaboration frees humans from repetitive tasks, helping them embrace strong economic potential[2]. Research has demonstrated that human-AI interaction is able to shape processes more effectively and enhances the individual performance of employees and when humans and AI work together, optimal and improved outcomes can result[3, 9].While AI had a 7.5% error rate when operating on its own and employees had a 3.5% error rate operating on their own, combining the work of people and AI machines dropped the error rate to 0.5% . Collaboration between humans and AI has also shown to increase the IQ of business teams. Shirado and Christakis (2017) find that AI technology helped human players enhance their performance in an online game, shortening problem solving time by 55.6%. Research has shown that AI socialization increases productivity, which leads to a greater level of trust between users and AI[3, 9, 12].

2.2 Challenges faced by Industry

One of the main challenges presented by Human-AI interaction is the human's fear of losing control or privacy. In an effort to decrease workplace error and increase productivity, Industries are giving AI agents a more active role in daily operations. This implementation must be done carefully, making sure workers and technicians do not feel threatened by the agent. An example of this concept is the Amazon Ultrasonic Bracelet patented in 2018, shown in **Figure 2**. The proposal of this patent stated "Ultrasonic tracking of a worker's hands may be used to monitor performance of assigned tasks" [13]. The system was described to be used as a sorting aid for workers to replace conventional scanners, allowing workers' hands to be free. However, this tool received strong public backlash, with many citing it as a break of privacy, "akin to the idea of wearing an ankle bracelet during house arrest" [14].

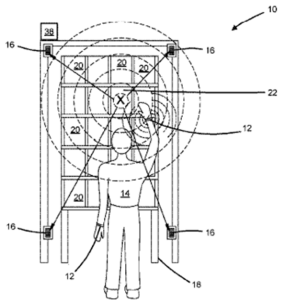

Figure 2 :*Patent US00988176B2 given to Amazon Technologies Inc. Jan. 30, 2018[13]*

There are different negative scenarios in which AI devices will become more intelligent than their creators, being self-aware, capable of scientific creativity, and gaining social skills, which will increase their independence in society and replace humans. Derived from these characteristics, negative or pessimistic scenarios about the development of AI have been forecasted. All these negative scenarios about the development of AI devices raise a certain skepticism in consumers' minds. Hence, it is important to gain the trust and acceptance of AI among consumers and humans in general, and firms should introduce complementary systems (e.g., incentive systems) that foster interactions within the reduced social environment AI might create. Furthermore, they need to focus on the fact that people will react less emotionally to algorithmic decisions, as they attribute less agency and intentionality to algorithms[8, 15].

If the user profile methods were to be implemented, care would need to be taken to ensure that the AI did not adapt to just convince the user. It would need to adapt to learn the correct answers. Overall, this means that the AI would need to present data in a fashion that better suits an individual user, but it should be equivalent in a manner where entirely separate conclusions could not be drawn from the output.

In Table 1, a developed "Model of AI Integration" in the workplace has been shown that combines the level of AI novelty (low/moderate/high) and the scope of AI (content changing/ incremental AI versus context changing/radical AI) dimensions. "Content changing" refers to the changes occurring in a narrower task domain, mainly based on the component knowledge of the associated system. "Context changing" refers to the changes that may occur in broader task domains based on the architectural knowledge of the system. The resultant quadrants entail six types of AI and associated human-AI relationships and human roles.

Table 1 : AI integration model[9].

Level of AI Novelty	Scope of AI: Content Changing AI	Scope of AI: Context Changing AI
High	**Autonomous** (Self driven vehicles) **Employee-AI Relationship**: Independence **Human Role**: Keep in Check IIIa	**Authentic** (Superintelligence) **Employee-AI Relationship**: Singularity **Human Role**: Comprehend IIIb
Medium	**Augmentation** (Robots in Surgery) **Employee-AI Relationship**: Complementary **Human role**: Collaborator IIa	**Alteration** (Deep Learning) **Employee-AI Relationship**: Symbiotic **Human Role**: Co-creator IIb
Low	**Automation** (Assembly line robots, DSS, automated online assistants) **Employee-AI Relationship**: Substitution **Human Role**: Controller Ia	**Amplification** (Predictive AI) **Employee-AI Relationship:** Supplementary **Human Role**: Conductor Ib

The resulting quadrants should be treated as ideal scenarios. As AI systems continue to become more complex, far reaching, and novel, the ideal types described in our typology may move from one quadrant to another. Moreover, some narrow, content changing AI systems may have broader context changing applications in the adopting organizations such as systems developing more 'contextual awareness' and having more self-learning capabilities. Such context changing applications of AI systems in new domains will also depend on employees who use such systems and find new ways to apply them within the organization[9].

3. Advanced Preparations

In order for implementation to be successful, we must plan today as much of the technology and integration will require significant time. We foresee advanced preparations necessary in the following elements: Education, worker training, and the implementation of increased exposure to AI.

3.1 Educational Impact

Engineering and technological education is ever evolving in terms of techniques as well as its content. Topics exist in present day collegiate curriculums that would not have been present in the curriculum of a student from the 1950s such as having a base level familiarity with coding. There will be an inevitable shift of what topics an average engineer will be required to know in order to be competent in the job market. With this we believe that more students should end up with a substantially larger set of course interactions in order to increase the quantity of technical users.

It is the belief of this proposal that the integration of these AI systems will become increasingly easier over the years as more students become more technical users. However, there are ways in which this transition could be reached quicker based on several personal experiences of the company.

Experience is usually the metric by which competency is judged in technical fields. There exists a gap between education and experience that, in recent years, has been bridged in part by the flipped classroom setting. The structure of a class based around granting users greater understanding of the inner-workings of agent decision making would benefit from a project-based structure class that allows the students the flexibility to choose an application of AI, but also gain footing into the experience that engineers might be missing before addressing these systems in other industry environments.

Since experience is also derived from a set of hands-on encounters, linking the two could prove to be beneficial for the students and make the training of these workers much easier. Non-technical users might benefit from the experience of technical users as one might be able to readily explain certain aspects of the agent.

3.2 Worker Training

After employment, no matter the level of education obtained prior, worker training is necessary for each industry. For every process or change in manufacturing and design, there is a need for a unique training module or course. These trainings are used to make workers familiar with tools and processes used in each industrial setting. It has been shown that AI agents can be incorporated into training systems via a VR environment to improve the measured performance on a custom given task[16]. This type of training environment doubles as a tool that can be used to increase worker's trust in AI systems before beginning official tasks. As workers undergo training with assistance from AI agents, the increased exposure to these agents will increase the inherent trust workers have in AI systems.

4. Conclusion

It is our belief here at "CDS" that AI has and will continue to be an ever increasingly utilized tool that industry will begin to adopt. Provided the proper training and future educational interactions, many workers of the future will be well suited for the task of iterating with such entities. It is up to the next generation of cutting-edge industries to facilitate the growth of these tools by implementing and overcoming the challenges associated with the benefit of optimizing trust between those users on all sides of the process. With uses in both design and manufacturing, it is easy to see how monetarily an individual company might be able to benefit from streamlining these tools by using the AI to adapt with the user. In short, if an industry seeks a competitive edge in the markets of 2040, steps should be taken now to further optimize the trust in human-AI interaction.

References

[1] A. Ferrario, M. Loi, and E. Viganò, "In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions," *Philosophy & Technology,* vol. 33, no. 3, pp. 523-539, 2019, doi: 10.1007/s13347-019-00378-3.

[2] B. Liu, "In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction," *Journal of Computer-Mediated Communication,* vol. 26, no. 6, pp. 384-402, 2021, doi: 10.1093/jcmc/zmab013.

[3] M. M. F. B. N. R. J. M. F. S. Stieglitz, "The rise of artificial intelligence – understanding the AI identity threat at the workplace," 5 October 2021, doi: 10.1007/s12525-021-00496-x.

[4] R. Bendell, J. Williams, S. M. Fiore, and F. Jentsch, "Supporting Social Interactions In Human-Ai Teams: Profiling Human Teammates From Sparse Data," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* vol. 65, no. 1, pp. 665-669, 2021/09/01 2021, doi: 10.1177/1071181321651354b.

[5] D. A. Hyanghee Park, Kartik Hosanagar,Joonhwan Lee, "Human-AI Interaction in Human Resource Management: Understanding Why Employees Resist Algorithmic Evaluation at Workplaces and How to Mitigate Burdens," 2021, doi: 10.1145/3411764.3445304.

[6] T. G. Kailas Vodrahalli, James Zou, "Do Humans Trust Advice More if it Comes from AI? An Analysis of Human-AI Interactions," 14 Jul 2021.

[7] D. A.-T. Mohammad Naiseh, Nan Jiang & Raian Ali "Explainable recommendation: when design meets trust calibration," 02 August 2021 2021, doi: https://doi.org/10.1007/s11280-021-00916-0.

[8] C. Pelau, D.-C. Dabija, and I. Ene, "What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry," *Computers in Human Behavior,* vol. 122, 2021, doi: 10.1016/j.chb.2021.106855.

[9] D. M. Erin E. Makariusa, Joseph D. Foxa, Alexa K. Foxb, "Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization," 24 July 2020 2020, doi: 10.1016/j.jbusres.2020.07.045.

[10] P. a. R. Abhishek, "The relationship between artificial intelligence and psychological theories," *International Journal of Conceptions on Computing and Information Technology,* vol. 1, no. 1, 2013.

[11] H. Hourani, A. Hammad, and M. Lafi, "The Impact of Artificial Intelligence on Software Testing," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 9-11 April 2019 2019, pp. 565-570, doi: 10.1109/JEEIT.2019.8717439.

[12] H. Shirado and N. A. Christakis, "Locally noisy autonomous agents improve global human coordination in network experiments," *Nature,* vol. 545, no. 7654, pp. 370-374, 2017/05/01 2017, doi: 10.1038/nature22332.

[13] c. J. Evan, "ULTRASONIC BRACELET AND RECIEVER FOR DETECTING POSITION IN 2D PLANE," 2018.

[14] A. Boyle, "Amazon Wins a Pair of Patents for Wireless Wristbands That Track Warehouse Workers," 2018.

[15] M. K. Lee, "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management," *Big Data & Society,* vol. 5, no. 1, p. 2053951718756684, 2018/01/01 2018, doi: 10.1177/2053951718756684.

[16] A. O. BUTEAN, Marco Leon; TOCU, Nicolae Adrian; FLOREA, Adrian, "Improving Training Methods for Industry Workers though AI Assisted Multi-Stage Virtual Reality Simulations," 2019, doi: https://doi.org/10.12753/2066-026X-19-007.