Clemson University

TigerPrints

All Dissertations

Dissertations

8-2024

Integrating Artificial Intelligence and Augmented Reality For Enhanced Task Performance In the Construction Industry

Aasish Chandrika Bhanu

aasishc@clemson.edu

Follow this and additional works at: https://open.clemson.edu/all_dissertations

Part of the Architectural Engineering Commons, Civil Engineering Commons, Cognition and Perception Commons, Construction Engineering Commons, Construction Engineering and Management Commons, Experimental Analysis of Behavior Commons, Hydraulic Engineering Commons, Other Civil and Environmental Engineering Commons, and the Structural Engineering Commons

Recommended Citation

Chandrika Bhanu, Aasish, "Integrating Artificial Intelligence and Augmented Reality For Enhanced Task Performance In the Construction Industry" (2024). *All Dissertations*. 3673.
https://open.clemson.edu/all_dissertations/3673

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

INTEGRATING ARTIFICIAL INTELLIGENCE AND AUGMENTED REALITY
FOR ENHANCED TASK PERFORMANCE
IN THE CONSTRUCTION INDUSTRY

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Civil Engineering

---

by
Aasish Chandrika Bhanu
August 2024

Accepted by:
Dr. Kapil Chalil Madathil, Committee Chair
Dr. Kalyan Piratla
Dr. Vivek Sharma
Dr. Patrick Rosopa

ABSTRACT

The Architecture, Engineering, and Construction (AEC) sector, characterized by its complex tasks, often faces challenges in adopting cutting-edge technologies, a trend that can significantly hinder productivity improvements compared to other industries. This dissertation explores the emerging integration of Artificial Intelligence (AI) and Augmented Reality (AR) within the AEC domain, investigating how effectively this can be implemented to potentially revolutionize industry practices. AI could be used in the AEC domain for advanced analytical and decision-making processes, while AR could be used for training, visualization, and remote collaboration.

In scenarios where construction site workers require expert guidance, these technologies could be helpful. AR, for instance, enables remote experts to offer real-time assistance, while AI can provide data-driven recommendations through AR interfaces, analyzing on-site challenges to suggest practical solutions. This integration of AI and AR harnesses the strengths of both technologies, potentially transforming the AEC landscape. However, this technological integration is not without its challenges, particularly concerning trust in AI, an important factor for the successful implementation of AI across domains.

The initial study of this dissertation investigates the effect of task complexity and AI recommender system reliability on trust, performance, and workload. Utilizing a Partially Observable Markov Decision Process (POMDP) model, the evolution of trust during interactions with recommender AI was modeled. The experimental study identified that the task complexity did not significantly impact trust in AI. Instead, the reliability of the AI agent emerged as a crucial determinant, with higher reliability correlating with increased trust. The trust trajectory predicted by the POMDP model aligned closely with experimental findings under most conditions, offering valuable insights for AI system designers for AR modules while completing construction tasks.

The second study investigated the effect of transparency and explainability on trust in the recommender AI. Specifically, it investigated the effects of varying levels of explanation and transparency on trust, performance, and workload while completing a construction task with the help of AR technology. The study identified that the overall change in trust after introducing a combination of transparency and explainability increased regardless of the reliability of the AI but with the cost of the time taken to complete the task. This combination of transparency and explainability was provided along with all the recommendations throughout the task. This led to the idea of the final study, which is to investigate the effect of providing the combination of transparency and explainability in a non-continuous way, especially in AR modules, to avoid visual clutter of information.

The third study investigated the effect of providing a combination of AI transparency and explainability along with the recommendation in an adaptive manner to maintain appropriate trust, compared to providing them continuously. The study identified that there was no significant reduction in trust level when transparency and explainability were provided when AI's confidence was low and when provided during the first, middle and last steps of the task compared to providing them continuously with an advantage of time taken to complete the task while providing them intermittently.

This research attempts to offer a detailed exploration of the factors influencing trust in AI, particularly when interfaced with AR technology while completing a construction task. The findings are expected to guide the development of more effective, trust-enhancing AI tools, paving the way for their broader acceptance and implementation in complex construction environments.

DEDICATION

This dissertation is dedicated to my beloved parents, K. Bhanu and K. Chandrika, my wife Dhanya Vijayakumar and my children Aaradhya Aasish, Aadhvika Aasish and Aayush Aasish for their unconditional love and support.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Kapil Chalil Madathil. Your unwavering support, understanding and mentorship have been instrumental throughout my doctoral journey. Your invaluable advice and comments have been significant for both my research and future career plans. I am also specifically thankful for your empathetic support during the difficult times in my personal life. You have been a source of inspiration and I hope to embody the same level of dedication and hard work you have shown me when I move on to the next phase in my career. I am also profoundly thankful to my committee members, Dr. Kalyan Piratla, Dr. Vivek Sharma, and Dr. Patrick Rosopa for their insights and support throughout my research. Your guidance has been crucial in shaping my dissertation and I am deeply grateful for your flexibility in scheduling meetings and the continuous help.

I am also extremely grateful to Dr. Jeffery Bertrand for productive and development of all the simulations used for my research studies. I would also like to thank Ms. Barbara Ramirez for grammatically correcting Chapters 2 and 3, and Generative AI (ChatGPT) for providing grammatical correction for the remaining chapters. Additionally, I would like to thank Dr. Hamed Rahimian for collaborating and helping to develop the trust behavioral mathematical model. I am also thankful to Mr. Alastair Boulding for providing me with expert advice on pipe installation concepts and helping me develop the tasks to conduct the simulation-based research studies.

I would like to extend my gratitude to the research group of the Human Systems Integration Laboratory including Amal Ponathil, Sruthy Agnisarman, Shraddhaa Narasimha, Amro Khasawneh, Hunter Rogers, Sarvesh Sawant, Snowil Lopez, Harnish Sharma, Bhargav Upadhyay, Gayatri Anoop, Arvind Kolangarakath, Mitali Bhosekar, and Camden Brady for all the constructive discussions and help at different stages of my research. I will cherish the memories of the wonderful times that we spent together inside, and outside the lab.

Additionally, I would like to thank Soumya Pathy for assisting me with the mathematical model and answering all my related questions, helping me to learn and understand the complexities involved. I would also like to thank all the friends and well-wishers within Clemson University and other parts of the world for their moral support.

One person to whom I shared all my joys and challenges during my time at Clemson and beyond is my wife, Dhanya Vijayakumar. Without your support, this work would not have been possible. You took on many additional responsibilities at home and with our children to support me. Your sacrifice and support have been my pillar of strength. I also want to thank my children, who generously allowed me to work on my dissertation, even at the cost of some of our playtime together. I promise to make up for all that time.

I would like to thank my entire family for the encouragement and the emotional support provided. Specifically, I want to acknowledge my father, K. Bhanu, who worked very hard his entire life to provide for us and help us reach where we are now. I have tried and will continue to try my best to follow your life lessons and advice about life. I also want to express my deepest gratitude to my mother, K. Chandrika whose prayers and unwavering support have been a constant source of strength. One of her greatest wishes has been for me to attain a doctorate degree, and I am overjoyed to fulfill that dream. Your faith in me has been a guiding light throughout this journey. I also want to thank my brother, Aathish Bhanu, and my sister-in-law, Mincy Raj, for providing encouragement and support throughout. I also want to extend my heartfelt gratitude to my in-laws, R. Vijayakumar and Nalini Geethamma, as well as Arjun Vijayakumar and Neethu V. Kini, for their unwavering encouragement and support. Your presence and assistance during times of need have been invaluable. Finally, I would like to thank God Almighty for all his blessings and all my teachers, who have taught both academic and life lessons throughout my educational journey. Without your guidance and support, this achievement would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)..Page

List of Figures (Continued)..Page

CHAPTER ONE

INTRODUCTION

The Architecture, Engineering, and Construction (AEC) domain is a complex and challenging industry involving many complex and critical tasks ranging from structural design to scheduling to system installation, and errors within it can lead to significant delays and impacts. The AEC industry, traditionally seen as slow to adopt the latest technologies (Takim et al., 2013), has begun to incorporate some of the newer technologies to enhance efficiency and safety. For example, AEC projects are associated with a large amount of data such as multiple plans, elevations, and estimates, that are challenging to manage. With the introduction of building information modeling (BIM), this information could be managed digitally and can be visualized in 3D format when required and is one of the most promising changes that the AEC industry has adopted recently (Alizadehsalehi et al., 2020). There is a decline in productivity because of the delays in the adoption of the latest technologies, compared to the other sectors (Akanmu et al., 2021).

Some of these challenges could potentially be addressed to an extent by incorporating technologies such as artificial intelligence (AI), drones, virtual reality (VR) and augmented reality (AR). For example, one of the factors that contributed to the decline in productivity is inefficient on-site execution of construction activities, such as incomplete design and lack of clear scope (Cisterna et al., 2018). For instance, there could be cognitively demanding situations where a worker needs to either install a pipe layout for a heat pump or replace an existing one, all without having access to a design plan. This task is challenging, especially if the worker is not an expert. The challenges include making errors such as placing the pipe fittings in the wrong order or not using a necessary component in the layout, such as a Y-strainer or a regulating valve. Such problems in similar scenarios could potentially be addressed if the expert is onsite. But most of the time, experts must travel from a different location, which is both time-consuming and cost prohibitive. To mitigate this issue, remote support could be offered to the local operator using audio or video calls; however, this approach is not always effective due to the limitations of these technologies in adequately conveying the problem and its solution. To circumvent this issue, a remote expert could provide remote assistance using AR technology, which is more effective than traditional phone assistance for remote collaboration (Havard et al., 2015). Even further, by using AI technology, a well-trained AI agent could potentially provide recommendations to complete the task, reducing the need to have a remote expert helping the local worker.

AI can be defined as the field of engineering and science that focuses on the application and theory of developing systems that display traits like human intelligence (Tecuci, 2012). AI has been used in different industries to assist with the decision-making process, such as healthcare, for detecting cancer by analyzing radiology images (Capobianco, 2022), manufacturing for the detection of quality defects in the final product (Arinez et al., 2020), and aviation for air traffic management (Degas et al., 2022). AI could also be advantageous in the AEC industry for assisting processes such as construction planning and scheduling, risk management, quality control, predictive maintenance, and construction safety (Rafsanjani & Nabizadeh, 2023).

Another technology that AEC adopted recently is AR, which is used in the design, construction, and operation stages of construction (Chen & Xue, 2022). AR technology overlaps the computer-generated virtual information on the user’s view of the real world (Azuma, 1997). AR could potentially benefit users by reducing the cognitive load by displaying the required information while completing a physical task. AR technology can be achieved using handheld devices (HHD), such as phones and tablets, or using head-mounted displays (HMD), or spatial augmented reality (SAR), which is attained using projectors. AEC could benefit from AR in terms of implementation of BIM to support decision-making tasks

(Golparvar-Fard et al., 2011), communication between stakeholders (Noghabaei et al., 2020), real-time visualization for structural health monitoring (Sadhu et al., 2023), collaboration for remote assistance (Chantziaras et al., 2021), and much more.

AEC is a domain where productivity and reducing errors are of utmost importance since one of the main objectives of high-quality projects is to deliver them on time and within budget. By incorporating AI and AR, the benefits of both technologies could be harnessed and could be used to improve error detection, enhanced visualization, safety improvements, and to assist with decision-making. AI-assisted AR in industrial applications has huge potential, such as to increase productivity and minimize errors (Devagiri et al., 2022). For instance, AI can analyze the environment and provide recommendations on the task at hand based on the data that the system is trained in. While completing a task, the user could use an AR technology such as an HMD, HHD, or SAR, on which the AI can provide visual and textual recommendations for assistance. The AI could use the camera associated with the AR technology to learn about the environment and provide the necessary recommendations. This could reduce the use of manpower by avoiding the need for an expert to travel to the site or provide remote assistance to the worker on-site, contributing to solving the issue of skilled manpower shortage crisis that the industry has been going through recently (Kim et al., 2020).

AI integrated with AR can be used for simulating the construction process, visualizing building designs and training construction workers (Rafsanjani & Nabizadeh, 2023). Bowman et al. (2022) presented a proof-of-concept prototype of a similar system where intelligent AR technology supports construction workers at the site for steel assembly tasks by detecting, understanding and reasoning about the context of use and presenting only necessary information on AR display while providing step by step instructions. AI can also be used to identify potential safety hazards by monitoring workers' onsite movements and behavior

(Rafsanjani & Nabizadeh, 2023) and provide real-time recommendations to the workers on the AR platform to increase safety at construction sites.

As the demand for sustainable and smart infrastructure grows, the AEC industry's reliance on new technologies will only intensify, highlighting the need for continuous innovation and adaptation. Further research on integrating emerging technologies is necessary within the AEC industry to harness their full potential to enhance efficiency and safety.

Problem Statement

The fusion of AR and AI presents a novel paradigm in human-computer interaction, offering a more immersive and intelligent user experience. While AR provides a connection between the digital and physical worlds, AI enhances this connection with adaptive learning, predictive analytics, and real-time decision-making. However, this interaction also amplifies the challenges associated with each technology. To effectively design an integrated system that combines AR and AI, it is crucial to address individual challenges such as enhancing human interaction, training data for AI, and tackling industry-specific problems like, collaboration issues, and knowledge transfer in the AEC industry (Rebolledo et al., 2023; Wang et al., 2020). One example would be a study by Livingston et al. (2005) mentioning that the text layout and legibility on the AR platform significantly impact the users' performance, potentially hindering their ability to complete tasks efficiently. These aspects should be studied to better understand how performance and workload are affected when integrating these technologies.

In the case of AI, issues caused by system failures have emerged with the advancement and adoption of AI systems and applications. These errors have led to trust issues when using AI to predict and safely use the system (Um et al., 2022). Solberg et al., (2022) defined trust in AI as “the willingness of a person to be vulnerable to the actions of an AI decision aid, based on the expectation that it will perform a decision-making task important to the trustor.” The adoption of AI in AEC is delayed by the lack of trust in the technology (Emaminejad et al., 2021). Regardless of the application domain, while implementing technology for assistance, the user should have appropriate trust in the system for effective use of the technology (Bach et al., 2022). Trust affects the decision-making capability of humans (Park et al., 2008), and it directly affects their willingness to follow or reject the suggestions given by the AI (Freedy et al., 2007). For the implementation of the AI to be effective, the users should have the appropriate amount of reliance on AI rather than under- or over-reliance. The reason for under- or over-reliance on AI is inappropriate trust in the system (Lee & See, 2004). This inappropriate reliance will lead to misuse or disuse of the technology (Parasuraman & Riley, 1997). Even though AI can perform equally or even better than humans at many tasks, it is important to design the AI appropriately so that it is helpful to the users as intended. Poor design of the AI system could detrimentally impact the decision-making process (Lakkaraju & Bastani, 2020) and affect the user's trust in AI (Beauxis-Aussalet et al., 2021). For these reasons, to consider factors that affect trust while using AI technology is important since the design could affect the performance of the task.

Understanding the challenges associated with integrating AI and AR technologies is essential since finding solutions for these challenges and improving factors such as workload and appropriate trust in AI could lead to a better user experience and task performance. In prior studies, it has been demonstrated that providing transparency and explainability of AI recommendations can enhance user trust (Panganiban et al., 2020; N. Wang et al., 2016). Transparency of an AI system provides real-time status of what the system is currently doing (Andrada et al., 2023), whereas explainability provides an explanation or reason for the actions or recommendations provided by the AI (Arrieta et al., 2020). The explainability aspect aims to make the actions of the AI agent clear and understandable for the users, thereby improving trust in the system.

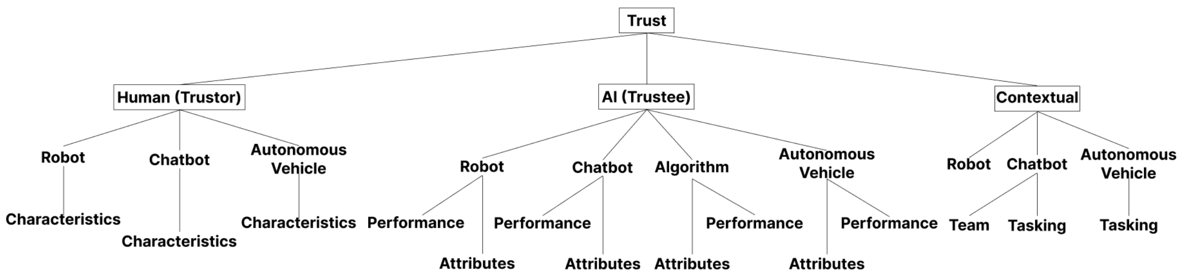
Accordingly, the overarching goal of this dissertation is to characterize human performance in AEC procedural tasks while integrating AI and AR technologies. This research aims to formulate design principles that optimize the utilization of these technologies, with a focus on enhancing explainability and transparency to maintain trust.

Trust Framework

Trust in AI is affected by different factors when human interacts with it. There are different frameworks associated with trust while interacting with automation. Recently, Kaplan et al. (2023) developed a model for trust in AI after conducting a meta-analysis of studies that investigated trust in AI. This trust model offers a detailed understanding of the factors influencing trust in AI. As shown in Figure 1, these factors were categorized into human-related, AI-related, and contextual antecedents of trust in AI.

Figure 1:

Trust in AI Model (Kaplan et al., 2023)



Human-related factors play a crucial role in determining trust in AI. The analysis showed that human-related variables, such as abilities and characteristics, significantly predicted trust levels. For instance, an operator's competency, understanding of the AI system, and expertise were positive predictors of trust. Personality traits and gender were also significant factors, such that lonely operators were less likely to trust AI, whereas innovative individuals were more likely to trust AI and males tended to trust AI more than females.

The AI-related factors, such as performance and attributes of AI systems themselves, are also important in shaping trust. AI's performance, especially its reliability, emerged as a significant predictor. The more reliable and higher performing an AI was, the higher the trust it gained. Additionally, AI's personality, anthropomorphic features, behavior, reputation and transparency positively influenced trust levels. For example, AI systems that were more human-like or focused on teamwork were trusted more, while those that appeared misleading were trusted less.

The context in which AI operates also affects trust. Variables, such as communication patterns and the length of the relationship between humans and AI, were significant predictors. Communication significantly predicts trust, where shared verbal speech is preferred over text communication, and trust is further enhanced when there's similarity between the speech patterns of humans and AI systems. Similarly, as the duration of the operator's interaction with the AI extended, their trust in the system correspondingly increased. Risk emerged as a crucial factor, with riskier situations diminishing trust. This is particularly relevant in high-stakes environments, such as aviation or self-driving cars, where humans and AI share potential risks.

Implementation of AI can be achieved in many ways depending on the use of it. One way of implementing AI technology is using it as a recommender system that assists users in completing tasks by analyzing project-specific data and offering tailored suggestions that enhance decision-making and operational efficiency. A recommender system is a tool designed to assist users in filtering through a vast array of options to find those most relevant to their needs (Ravi et al., 2022). For example, in complex construction activities, AI-enabled recommender systems can guide workers through installation processes, providing real-time, context-aware recommendations that aim to reduce errors and improve task performance. The effectiveness of such systems depends on the quality of their training data and their reliability in providing correct suggestions. As the AEC industry continues to evolve, integrating AI recommender systems could lead to improved decision-making by augmenting human expertise with AI-driven insights. In this dissertation, AI is implemented as a recommender system that assists workers in completing a construction task.

Trust is conceptualized through three distinct but interconnected layers: dispositional, situational, and learned trust (Hoff & Bashir, 2015). Dispositional trust is an individual's inherent tendency to trust, influenced by stable factors such as culture, age, gender, and personality traits. Situational trust varies with context, shaped by external factors like system complexity, task difficulty, and environmental risks, as well as internal factors including the operator's self-confidence, expertise, and mood. Learned trust evolves from an individual's accumulated experiences with specific automated systems, adjusting as they gain familiarity and observe system performance over time. In this dissertation, the concept of trust in AI predominantly pertains to situational trust, as the experimental studies presented herein are methodically designed to provide specific contexts with varying task characteristics and AI attributes. This approach aligns with the situational trust framework, wherein trust in AI is influenced by external factors, such as the complexity of the task, the reliability of AI, as well as internal factors related to the operator’s interactions within the given environment.

This dissertation investigates multiple aspects of trust in AI to help design better human interaction while using AR glasses integrated with AI for completing a procedural task. The meta-analysis by Kaplan et al. (2023) indicates the importance of contextual antecedents in shaping trust, pertinent to the dissertation's exploration of AI within AR environments. By examining how the complexity of the task, reliability, transparency, explainability, and adaptive explanations of AI agents affect user trust in a procedural task context, the research directly addresses situational trust as influenced by AI attributes and human-AI interaction dynamics. AEC is a domain with multiple complex tasks, and to effectively implement the recommender system in the AEC domain, the complexity of the task should be studied in terms of trust in AI. The tasks associated with the AEC domain come with a lot of uncertainties, and if not trained with all these uncertainties, there could be instances where AI makes errors. This might lead to time and recourse loss and even disasters. So, it is important to learn how humans interact with such systems, understand how users' trust in the AI agent varies, and how to maintain appropriate trust.

Research Objectives

The primary objective of this dissertation is to investigate the factors that should be considered while designing recommender systems for completing procedural tasks with the help of AR technology. More specifically, this dissertation explores the following research topics:

1. Understanding how AR technology is used in the AEC domain for remote collaboration.
    a. To understand the facilitators and barriers to using AR for remote collaboration in the AEC domain.
    b. To understand the different technologies and devices used for remote collaboration using AR in the AEC domain.
2. Understanding the effect of the reliability of the recommender AI and complexity of the task on trust in AI.
3. Investigating the effectiveness of transparency and explainability of recommender AI on trust in AI.
4. Evaluate the effectiveness of adaptive explanations in maintaining appropriate trust in AI.

Research Questions

The research questions are listed below:

While completing a procedural task with the help of a recommender AI and AR technology,

1. How and where is AR technology used in the AEC industry for remote collaboration?
2. How does trust vary with the varying reliability of recommender AI agents and the complexity of the task while completing a procedural task with the help of AR technology?
3. How is trust in AI affected by the presence of explainability and transparency while completing a procedural task with the help of AR technology?
4. When are providing explanations on the AR module beneficial to maintaining trust in AI while completing a procedural task?

To investigate these research questions, as an initial step, a systematic review was conducted on how AR technology is used in the AEC industry and its associated challenges. AEC has started to take advantage of the benefits of AR by adopting it for remote collaboration (Ammari et al., 2019). Using AR could overcome the limitations of the current technologies used for remote collaboration to an extent. It could also help real-time communication and spatial mapping (Chantziaras et al., 2021). This review identified a gap in the literature that specifically investigates the use of AR for remote collaboration in the AEC industry. To address this gap, a literature review was conducted to learn about the devices used, domains of application, facilitators, and barriers associated with using AR for remote collaboration in the AEC industry.

The integration of AI with AR technologies could potentially enable the delivery of expertise comparable to that of traditional remote assistance utilizing the AR platform. Trust in AI is a major aspect that needs to be investigated while implementing it, regardless of the domain. A POMDP model was developed to understand trust development while AI assists humans in making decisions when completing a procedural task. An experimental study was conducted to understand the effect of task complexity and reliability of the AI agent on the trust in AI and to validate the model. The main effect of complexity on trust was not significant, whereas reliability had a significant effect on trust. The second study investigated the effect of transparency and explainability of AI on trust and found that trust in AI increases when a combination of transparency and explanations are provided continuously, along with recommendations with the cost of time taken to complete the task. The final study investigated when it is worth providing a combination of transparency and explanations to maintain appropriate trust in AI and found that providing them at the first, middle, and last steps maintains appropriate trust compared to providing them continuously. Additionally, if the combination of transparency and explanation is provided only when the AI's confidence is low or at the first, middle and the last step, instead of continuously, not only the trust in AI is maintained but also the time taken to complete the task is decreased. This dissertation attempts to offer a novel contribution to the literature, guiding designers in developing AI systems that can be seamlessly integrated with AR technologies to effectively complete procedural tasks.

Dissertation Organization

This dissertation is organized as follows: Chapter 2 details a systematic review of the literature on the application of AR technologies for remote collaboration in the AEC industry. Chapter 3 evaluates the effect of varying reliability of the AI agent and the complexity of the task on trust in AI. Chapter 4 evaluates the effect of transparency and explainability on trust in AI. Chapter 5 investigates different strategies to understand when is providing explanations more beneficial in maintaining appropriate trust.

CHAPTER TWO

HUMAN FACTORS CONSIDERATIONS ON THE USE OF AUGMENTED REALITY TECHNOLOGY FOR REMOTE COLLABORATIVE WORK IN THE ARCHITECTURE, ENGINEERING, AND CONSTRUCTION INDUSTRY: A SYSTEMATIC REVIEW

Introduction

The complexities associated with the AEC industry and its increased demand for access to information, especially for evaluation, communication, and collaboration purposes (Rankohi & Waugh, 2013), necessitate adopting and integrating appropriate technologies to enhance human performance and safety. Because of the complex nature of the AEC industry, the teams need to collaborate for information exchange at different stages of a project, as project performance has been found to increase with effective collaboration among stakeholders during the planning, design, and construction stages of a project (Kapogiannis, 2013). For example, computer-supported collaborative work (CSCW), which is becoming prevalent, uses computers as a medium for human collaboration, with such technologies as BIM improving collaboration among various project stakeholders in the construction industry (Barlish & Sullivan, 2012; Demian & Walters, 2014). Using such collaborative technologies contributes to improved scheduling, communication, relationships among partners, project and information management, as well as transparency of project information, among others (Nikas et al., 2007).

One of this industry's most recently adopted technologies is AR defined as a system that combines real and virtual objects in real-time in three dimensions. Using AR, virtual objects are augmented into the real world, bringing the real and virtual worlds together (Bronack, 2011; Klopfer & Squire, 2008). AR is part of the Reality-Virtuality continuum, consisting of the real environment, augmented reality, augmented virtuality, and the virtual environment (Milgram et al., 1995).

AR can be achieved using a variety of device types, including HMD like Microsoft HoloLens, HHD like smartphones or tablets, and SAR, which uses a projector and a camera.

AR sees increasing application across a wide range of areas, from technical industries like education, manufacturing, maintenance, AEC, healthcare, military, aerospace, and engineering to more general industries like tourism, shopping, advertising, and entertainment (Ammari et al., 2019; Azuma, 1997; Zenati-Henda et al., 2014). Specific to the AEC industry, AR technology can be used for multiple purposes, including visualization, communication, information modeling, evaluation, progress monitoring, training, and safety inspection, whereas, in industrial construction, it can be used for different stages of a project from strategizing through execution to inspection (Rankohi & Waugh, 2013; Shin & Dunston, 2008). Other key benefits of using AR technologies include improving communication among the parties involved, increasing project understanding, accelerating decision-making, improving scheduling and budget management, providing real-time visualization, enhancing collaboration, improving safety, and enhancing the implementation of BIM (Danker & Jones, 2014).

AR technology is used for collaborative work in various industries for applications such as education and training, product design and development, and aircraft maintenance (Martín-Gutiérrez et al., 2015; Shen et al., 2010; Utzig et al., 2019; Zhong et al., 2002). Studies have found that AR can improve performance time and mental effort in a collaborative design task (Alem & Li, 2011) and facilitate communication and discussion of engineering processes (Dong et al., 2013). In addition, using AR for collaboration provides virtuality, referring to interaction with computer-generated content, augmentation, referring to adding these computer-generated content to the user's view in the real world, cooperation, referring to interaction between multiple people, independence, referring to controlling the augmented content individually, and individuality, referring to having personalized experience with the augmented content based on the user's preference (Schmalstieg et al., 2002).

There are two types of collaboration using AR technology: face-to-face and remote (Lukosch et al., 2015). In face-to-face collaboration, two or more users in the same room can see the same content on a tabletop using AR technology, whereas, in remote collaboration, users at different locations collaborate using AR technology to complete work or engage in discussions. This review will concentrate on the use of AR technology for remote collaboration.

The construction industry could benefit from the increased use of AR technology and its capability to support remote collaboration between multiple people from multiple locations. This review attempts to understand how and where this technology is used in the AEC industry and learn the barriers associated with it. The specific objectives of the study were to 1) understand how and where AR technology is used in the AEC industry for remote collaboration, 2) identify the different technologies and devices used in this process, 3) understand the facilitators and barriers of using AR for remote collaboration, and 4) identify gaps in the literature and propose future research.

Method

To understand the application of AR technology in remote collaborative work in the AEC industry, a systematic literature review was conducted. Preferred reporting items for systematic reviews and a meta-analysis (PRISMA) format were followed to report the findings from this literature review (Page et al., 2021). The inclusion and exclusion criteria were selected based on the scope of the review, which was using AR for remote collaboration in AEC. Keywords were identified from previous research and were searched in selected databases and journals. The first author then reviewed these articles to remove duplicates and completed the title and abstract screening, followed by the full-text screening based on the inclusion and exclusion criteria.

Eligibility criteria

The inclusion criteria for this review were that the study should investigate AR technology being used for remote collaboration in the AEC industry, and it should have been published in peer-reviewed publications or conference proceedings in English. Review articles, posters, and patented technologies were excluded from this study.

Search strategy

Articles published in English were searched in the following databases/journals: *Institute of Electrical and Electronics Engineers (IEEE) Explore, The American Society of Civil Engineers (ASCE) Library, Automation in Construction, Applied Ergonomics, Computer-Aided Civil and Infrastructure Engineering, Advanced Engineering Informatics,* and *Visualization in Engineering* in July 2020 using combinations of the following keywords using boolean operators (AND/OR); "augmented reality," "mixed reality," "remote," "distant," "distributed," "assistance," "communication," "collaboration," "collaborative work," "computer-supported collaborative work," "construction," "engineering," architecture," and "AEC."

Selection Process

This search resulted in a total of 695 articles; after checking for duplicates, 228 articles were removed, resulting in a total of 467 articles. The titles and abstracts were screened based on the inclusion and exclusion criteria, resulting in 48 articles being selected for the next round of screening. A full-text review by the researcher based on the inclusion and exclusion criteria resulted in a total of 12 papers for further review. During the full-text review, the studies that explored the use of AR for remote collaboration for tasks that could be applied to the AEC, such as general tasks and maintenance tasks, were not excluded. Figure 2 shows this literature selection process, and Table 1 shows the distribution of papers selected from corresponding journals.

Figure 2:

PRISMA Flow Diagram Detailing the Selection Process

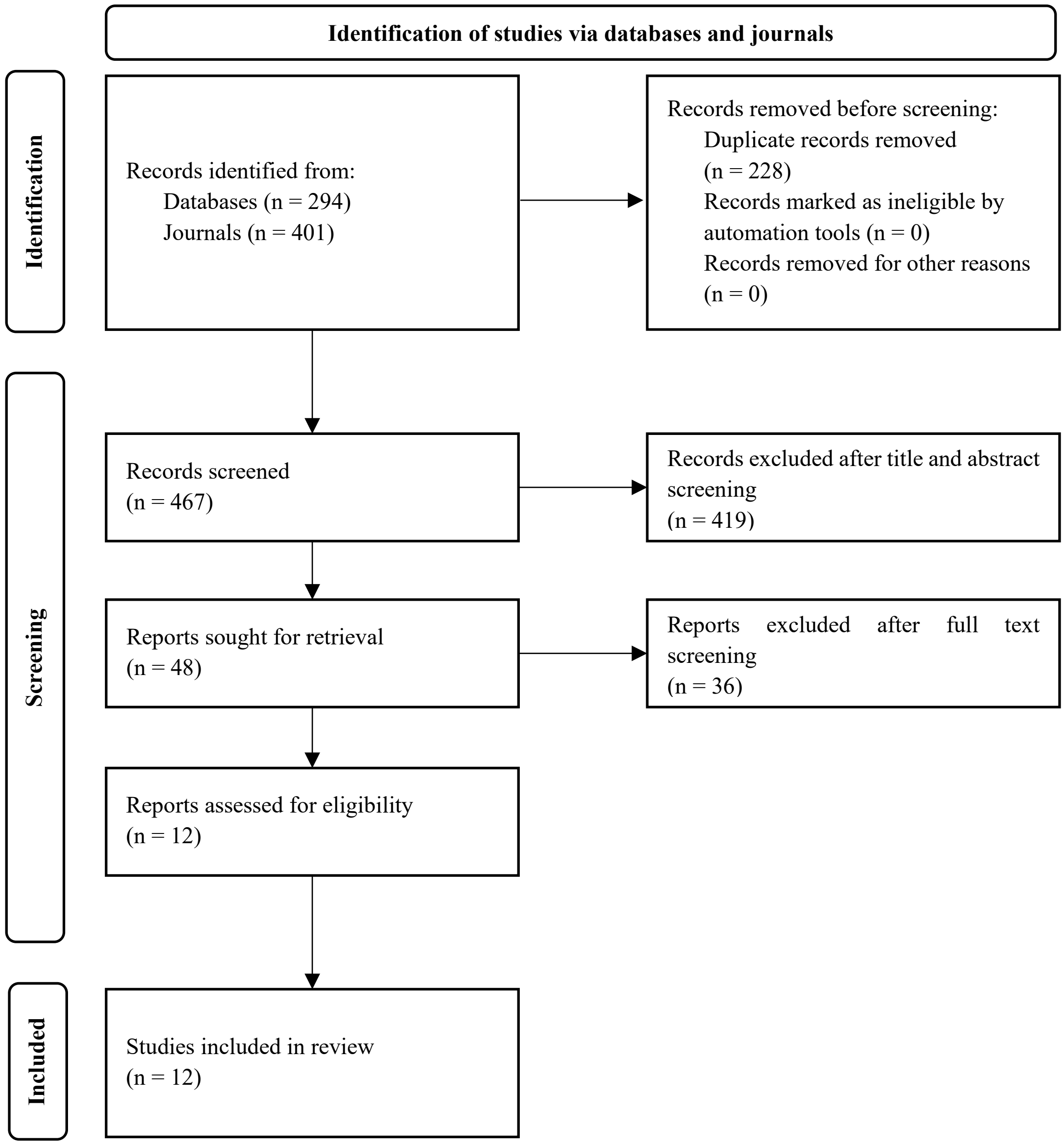


Table 1:

Distribution of Articles from the Database/Journal Search

Database/Journal	**Total number of articles in the initial search**	**The final number of articles selected**
IEEE Explore	166	8
ASCE Library	128	1
Automation in Construction	275	1
Applied Ergonomics	17	0
Computer-Aided Civil and Infrastructure Engineering	5	0
Advanced Engineering Informatics	92	1
Visualization In Engineering	12	1

Data extraction and synthesis

The 12 articles selected for data extraction and thoroughly reviewed are reported in Table 2. Based on the objectives of the study, various themes were identified in the 12 papers and are reported in the results section. The main data items that we investigated in these articles are the area of application of AR technology, the type of hardware used, and the facilitators and barriers to using AR for remote collaboration. To understand the impact of the technology on remote collaboration, the dependent variables primarily used were performance and workload. The majority of studies measured performance using either the time taken to complete the task, the number of errors made, or both. Mental workload is generally measured using NASA TLX and is measured to quantify the cognitive cost to complete a task for predicting performance (Cain, 2007).

Table 2:

Summary of Articles

Article	Domain	Technology/equipment used	Task	Variables	Objectives	Findings
Lee et al., (2020)	Search and assembly	Microsoft Hololens mounted with Ricoh Theta S 360 camera for local workers and HTC Vive for the remote expert	Locate tans (flat polygons) and assemble them in collaboration	IV: 1) View sharing technique: 2D video, 360 video, and 3D model augmented with 2D video 2) Task roles: independent task, divided task DV: Search time, assembly time, workload, social presence	To compare the three view-sharing techniques for the remote expert that supports the local workers with AR glasses	The task completion time was significantly less for the 3D model condition than that of the other two conditions for the independent task scenario, whereas the 2D condition had more completion time compared to the other two conditions in the divided task. The workload was less for 3D condition, whereas usability, social preference, and preference were more for the 3D condition for both the scenarios
Zenati-Henda et al., (2014)	Industrial maintenance	Vizur wrap 920 AR HMD and MSI Windpad 110 W Tablet for the local worker and interactive table connected to a desktop for remote expert	1) Replace the printer cartridge 2) Check and add engine oil in a car	IV: The mode of collaboration: HMD vs. Tablet DV: Ease of use, usability, usefulness, task completion time, evaluate the usefulness and the usability of the proposed system	To observe how the users used the new system and to study its usability and usefulness	The users were satisfied with the system, especially with the HMD, since it was handsfree, but HMD provided less flexibility for application control
Wang & Dunston, (2009)	Design error detection support in the construction industry	AR based face-to-face system, and AR based virtual space system using ARvision HMD for both parties- ARvision	Collaboratively find the errors of a pipe layout	IV: Mode of collaboration: AR vs. paper (face to face) and AR vs. Navis Work Roamer DV: Performance time	To test the capabilities of the AR system by comparing the performance time	AR based face-to-face system reduced performance time by 64%, and AR-based virtual space system reduced the performance time by 67% compared to the prevalent methods
Aschenbrenner et al., (2018)	Industrial Robot repair	For the local user, Epson Moverio BT200 with an android control device for AR HMD, Panasonic PT-VZ575N projector with PointGrey Black-fly camera for the projector condition, and ASUS MEMO ME302C Tablet for the tablet condition and a laptop for the remote expert	Exchange of a controller in a switch cabinet of an industrial robot	IV: Mode of collaborations: tablet-based AR (video, screenshot, and tracking), optical see-through AR, phone, and projections-based AR DV: Task duration, workload, and situational awareness	To measure the effectiveness of the different modes of collaboration for a collaborative industrial repair task	Significantly less task duration was found for the projector condition compared to the phone condition

Article	Domain	Technology/equipment used	Task	Variables	Objectives	Findings
Wang et al., (2019)	Assembly training in the manufacturing industry	Projector and Camera for SAR at the local site and HTC Vive and Leap Motion for VR at the expert side	Assemble a vise	IV: Mode of interaction at the remote expert side- Controller-based and Gesture-based DV: Performance time and System Usability Scale (SUS)	To compare the performance of gesture-based and controller-based conditions and to identify user's preference	No significant differences between the performance times for the two conditions were found. Participants preferred the gesture-based condition compared to the controller-based one.
Obermair et al., (2020)	Industrial maintenance support	Smartphone for the local worker and laptop for the remote expert- models not mentioned	Change a part of an industrial PC	IV: Mode of support DV: Average completion time and error percentage	To determine the benefits of AR remote support in maintenance by comparing the completion time and error percentage with support using paper instructions.	For the AR remote support, the number of errors was 75% less compared to the paper-based support. The total duration of the tasks was similar for both conditions but significantly different for the type identification task.
Olorunfemi et al., (2018)	Risk communication and hazard identification in the construction industry	Microsoft HoloLens for the local user and a computer tablet for the expert	Communicate a risk scenario between the expert and local worker	IV: Model of communication: AR, phone call, walking up to people and talk, video conferencing, and emails DV: Communication accuracy	Evaluate the accuracy of using AR technology communicating safety risks at construction sites	AR technology had more accuracy of risk communication compared to the other methods
Le Chenechal et al., (2016)	Industrial maintenance	Oculus Rift with two stereo cameras for both the parties	Physical bimanual selection task by simultaneously touching two targets	IV: 1) Mode of communication: AR system and desktop screen-based system 2) Complexity of the task DV: Task completion time	Evaluate the effectiveness and user preference of the AR system	The AR system provides a faster completion time compared to the desktop screen based system on the complex task. The cognitive load on the AR system is less compared to the traditional approach. Participants preferred the AR system except for the visual comfort.
Ammari et al., (2019)	Facilities management support	Android tablet (Samsung Galaxy Tab S2) with FLIR ONE thermal camera for the local worker and Oculus Rift for the remote expert	Replace a thermostat attached to a wall	IV: 1) Mode of view (Immersive Augmented Virtuality and AR module) 2) collaborated/non-collaborated DV: Task completion time, number of errors, Usability, and	1) Usability testing and time and error analysis 2) accuracy of remote marking and sketching	The number of errors and time to identify marked tasks were less when collaborating compared to the non-collaborative mode. System effectiveness was found to be 90.2%. Participants suggested using AR glasses instead of tablets for the AR module

Article	Domain	Technology/equipment used	Task	Variables	Objectives	Findings
				accuracy of remote marking and sketching	3) System effectiveness	
Fernández del Amo et al., (2020)	Aircraft maintenance support	Microsoft Hololens for the local worker and desktop for remote expert	Visual inspection of an aircraft's fuel hatch	IV: Mode of collaboration: AR and non-AR DV: Task completion time, errors	To evaluate the proposed effectiveness of the proposed communication framework for AR remote communication	The average time to communicate was reduced by 56% when compared to the traditional phone call and email communication. No significant difference in the number of errors for the different conditions.
Wang & Dunston, 2013)	Remote support in the construction industry	ARvision HMD for both local worker and remote expert	Find errors in a 3D CAD model of ductwork	IV: Mode of collaboration: Navisworks Roamer and HMD DV: Attitude towards the effectiveness of AR system, the user experience of AR system, and usability issues, task completion time	Evaluate the effectiveness and user experience of the AR system	Task completion time was much less for the AR system. Level of immersion, sense of location and orientation, and suitability of making decisions and performing tasks on design models was significantly better for the AR system
Sasikumar et al., 2019)	Not mentioned	Magic Leap One for the local user and Vive Pro VR HMD with Leap Motion hand gesture sensor	The local user has to pick up and place the lego block at a location specified by the remote expert.	IV: Natural cues: User-centric and Device-centric DV: Task completion time, social presence	To understand the use of natural communication cues by comparing a controller and annotation-based cues in a collaborative task.	No difference in task completion time or workload was identified between the two AR conditions. Above average copresence was noticed, and the remote users have less physical workload compared to the local users.

Note: IV- Independent variable; DV- Dependent variable

Results

Based on the first three objectives of this study, the results were classified into the four categories of the application of AR for remote collaboration, the technology and hardware used for both remote experts and local users, the facilitators of using AR for remote collaboration, and the barriers associated with using AR for remote collaboration. Of the 12 studies reviewed, one of each was conducted in the United States, New Zealand, South Korea, Austria, Canada, France, or Algeria, and five included researchers who collaborated from multiple countries. These 12 papers reviewed include qualitative, quantitative, and usability studies, with all 12 articles conducting quantitative research as an experimental study or as a pilot study for their technology developed; 3 of them contained qualitative studies, and 5 of them had usability studies of a technology or a system. 11 studies used performance as a metric in investigating the effectiveness of AR for remote collaboration. Table 3 shows the different dependent variables used in the reviewed studies.

Table 3:

Dependent Variables

Authors	Dependent variables						
	Performance		Workload	Usability	Situational awareness	Social presence	Communication accuracy
	Time	Errors					
Lee et al., (2020)	•		•			•	
Zenati-Henda et al., (2014)	•			•			
Wang & Dunston, (2009)	•						
Aschenbrenner et al., (2018)	•		•		•		
Wang et al., (2019)	•			•			
Obermair et al., (2020)	•	•					
Olorunfemi et al., (2018)							•
Le Chenechal et al., (2016)	•						
Ammari et al., (2019)	•	•		•			
Fernández del Amo et al., (2020)	•	•		•			
Wang & Dunston, (2013)	•			•			
Sasikumar et al., (2019)	•		•			•	

Application of AR for remote collaboration

The papers reviewed here provide insights into the application of collaborative AR support for various areas within the AEC domain, ranging from design review to facilities management support, with maintenance being the most frequently identified area (50% of the studies). Other applications involved training (17% of the studies), design review (17% of the studies), assembly (8% of the studies), and risk communication (8% of the studies).

Lee et al. (2020) explored search and assembly tasks where multiple users at the local site were supported by a remote expert using AR technology, the only study identified in this review connecting multiple local users with a single remote expert, and Le Chenechal et al., (2016) explored the concept of virtual arms, specifically Vishnu, at a physical site for use in industrial maintenance, procedure learning, and sports training. Olorunfemi et al. (2018) investigated AR technology for use in the construction industry for risk communication and hazard identification. The experts from a remote location collaborated with a user concerning a hazard happening in real-time on the site, thus avoiding a safety visit that would interrupt the workflow. Similarly, AR was also used to maintain industrial personal computers (PCs) with the expert guiding the customer to fix the issue remotely (Obermair et al., 2020). Wang et al. (2019) applied an AR platform in the manufacturing training industry to virtually train an operator using VR on the expert side and AR as Spatial Augmented Reality on the local worker side. The expert guided the worker by moving the virtual models using hand gestures or VR controllers, with these real-time motions of the models being projected to the local worker using a projector. Wang & Dunston (2009) tested the capabilities of an AR system in their design review by detecting the design errors of a pipe layout using a team of two participants, one local and one remote, who collaborated to identify the errors associated with the pipe layout design. These researchers conducted a second study on the user perception and impacts of using mixed reality for remote design review (Wang & Dunston, 2013). Table 4 shows the list of application areas of AR for remote collaboration identified from the reviewed papers.

Table 4:

Application Areas of AR for Remote Collaboration

Authors	Maintenance	Training	Design Review	Assembly	Risk Communication
Lee et al., (2020)				•	
Zenati-Henda et al., (2014)	•				
Wang & Dunston, (2009)			•		
Aschenbrenner et al., (2018)	•				
Wang et al., (2019)		•			
Obermair et al., (2020)	•				
Olorunfemi et al., (2018)					•
Le Chenechal et al., (2016)	•				
Ammari et al., (2019)	•				
Fernández del Amo et al., (2020)	•				
Wang & Dunston, (2013)			•		
Sasikumar et al., (2019)		•			

Hardware and technology used

The 12 studies reviewed here used various AR technologies and hardware for remote collaboration. All the studies reviewed used some kind of AR device at the local user side, which includes HMD, HHD, and SAR. The HMD used included AR HMD, like Microsoft HoloLens and MagicLeap, and VR HMD, like Oculus Rift, with stereo cameras, while the HHD included smartphones and tablets with a camera and SAR included using a projector along with a camera.

The hardware types used for the remote expert included HMDs, desktops/laptops/tablets, and interactive tables. The HMD for the remote expert included the Oculus Rift and the HTC Vive, and the interactive table was connected to a desktop. Of the studies reviewed here, nine used HMD, four HHD, and two SAR for the local user, while seven studies used HMD, four desktops/laptops/tablets, and one an interactive table for the remote expert. Five studies used HMD for both remote and local users, and one did not use any kind of HMD for either user. Table 5 shows the hardware types used by the studies for both the local and remote users.

As Table 5 shows, the majority of the studies used HMD for the local user, with AR HMD being the most frequently used: Microsoft HoloLens was used in 25% of the studies (Fernández del Amo et al., 2020; G. Lee et al., 2020; Olorunfemi et al., 2018), ARvision in 17% of the studies (Wang & Dunston, 2009, 2013); and each MagicLeap, Epson Moverio BT200, and Vizur Wrap 920 in 8% of the studies (Aschenbrenner et al., 2018; Sasikumar et al., 2019; Zenati-Henda et al., 2014). Oculus Rift, a VR HMD, along with two stereo cameras, was used in 8% of the studies on the local user side for the AR condition (Le Chenechal et al., 2016). Similar to the local user, the device type most frequently used for the remote expert was also HMD but most often VR HMD: HTC Vive was used in 17% of the studies (Lee et al., 2020; Wang et al., 2019); Oculus Rift in 17% of the studies (Ammari et al., 2019; Le Chenechal et al., 2016); and each ARvision and Vivie Pro in 8% of the studies (Sasikumar et al., 2019; Wang & Dunston, 2013).

Table 5:

Hardware Types Used by the Local User and Remote User

Authors	Local User			Remote User		
	HMD	HHD	SAR	HMD	Desktop/laptop/tablet	Interactive table
Lee et al., (2020)	•			•		
Zenati-Henda et al., (2014)	•	•				•
Wang & Dunston, (2009)	•			•		
Aschenbrenner et al., (2018)	•	•	•		•	
Wang et al., (2019)			•	•		
Obermair et al., (2020)		•			•	
Olorunfemi et al., (2018)	•				•	
Le Chenechal et al., (2016)	•			•		
Ammari et al., (2019)		•		•		
Fernández del Amo et al., 2020)	•				•	
Wang & Dunston, (2013)	•			•		
Sasikumar et al., (2019)	•			•		

Note: HMD- Head Mounted Device; HHD- Handheld Device; SAR- Spatial Augmented Reality

Microsoft HoloLens was used in a safety hazard communication study to view and interact with the holographic collaborative environment created from the user's field view (Olorunfemi et al., 2018). Hand and finger gestures and the user's gaze were used to control what they saw. These users communicated with the remote team members using Skype through the HoloLens, where both parties could annotate spatially and textually in the virtual world. Wang & Dunston (2009) used ARvision stereoscopic HMD connected to a color video camera to compare the performance time of collaborative error identification using AR systems to that of the current methods. One of the AR systems they used was an AR-based virtual space system in which two participants collaborated on a task from two locations. Lee et al. (2020) compared three view-sharing techniques—2D Video, 360 Video, and 3D Model—for use in situations where multiple local workers are supervised by a remote expert using the HTC Vive VR HMD to see the local workers. The first condition involved using videos directly linked to the local worker's live streaming using the camera of Microsoft HoloLens, while the second condition had live panoramic video captured from the 360-degree camera mounted on the HoloLens. The third condition used a reconstructed 3D workspace that the expert can virtually walk through and give directions to the local user. However, since it is difficult to reconstruct the dynamic scene accurately in real-time, they used a pre-reconstructed static workspace environment and displayed the dynamic scene captured using the camera of Microsoft HoloLens. In all three conditions, the expert has the freedom to switch between the views of the local workers.

The second most used AR device type by local users for collaboration was HHD. As seen in Table 5, four studies used HHD for communication between local users and remote experts; of these four, three used tablets, and one used a smartphone as an AR device. More specifically, Ammari et al., (2019) used Samsung Tab S2 as the HHD, while Zenati-Henda et al., (2014) used an MSI Windpad 110 W Tablet, and Aschenbrenner et al., (2018) an ASUS MEMO ME302C Tablet.

SAR, the third device type, was used in only two of the 12 studies reviewed (Aschenbrenner et al., 2018; P. Wang et al., 2019). SAR uses a projector to augment information from the real world to help the remote expert share it with the local user in conjunction with a camera that transfers the data from the local site to the remote expert. The remote expert uses the projector to point to or annotate the real world on the local side to express his ideas. For the SAR condition, Aschenbrenner et al., (2018) used a Panasonic PT-VZ575N projector along with a PoitnGrey Black-fly camera for their research, while Wang et al., (2019) used SAR for augmenting the real-time motions of the 3D CAD models manipulated by experts at a remote location using VR. They used a camera to display the local situation in real-time to the expert, who then guided the worker using the 3D CAD models in VR. The real-time movements of the CAD models were projected to the local worker using a projector.

The hardware was combined with various software/technologies in these 12 studies. Hand gesture-based AR was used in three studies by the remote expert to communicate with local users (Le Chenechal et al., 2016; Sasikumar et al., 2019; Zenati-Henda et al., 2014) while in the study conducted by Wang et al., (2019), the remote experts used hand gestures to manipulate virtual CAD replicas while wearing VR HMDs during communication with the local user. Five studies indicated that they used the Unity game engine to develop the prototype for the research (Ammari et al., 2019; G. Lee et al., 2020; Obermair et al., 2020; Sasikumar et al., 2019; P. Wang et al., 2019). Obermair et al., (2020) used it to create visualization tasks and AR Foundation for annotation anchoring and environment tracking while audio/video communication with the remote expert was achieved using web real-time communication (WebRCT). They used smartphones as AR devices, and since smartphone cameras are not equipped with depth cameras for tracking, the simultaneous localization and mapping (SLAM) algorithm, which is based on the 2D information of the camera image, was used. Two other studies also mentioned using SLAM for their future research to improve the reconstruction of the live environment (Le Chenechal et al., 2016; G. Lee et al., 2020). For example, Le Chenechal et al., (2016) mentioned that in the future, they would consider replacing optical tracking using stereo cameras with SLAM for image-based tracking during 3D reconstruction. In their study, AR technology was used to augment the virtual arms of the remote expert to assist the local worker in completing a task. An Inverse Kinematics algorithm was used for the elbows and hands of the virtual arm while the shoulders were fixed with respect to the local worker. The system dynamically reconstructed the real environment in 3D based on Kinect™ so that the occlusion of the real objects by the virtual ones could be avoided. This also helped to increase the ease of presence and depth perception using the virtual shadows. An interaction technique based on a virtual hand metaphor was used to manipulate these virtual objects, and the researchers used the GoGo arms navigation technique, which allows users to reach and manipulate objects that are beyond their physical arm's reach to reach the virtual objects in immersive environments. The OptiTrackTMV 120: Trio IR tracking system was used to track the agent's head (6 degrees of freedom) and real objects.

Facilitators of AR for remote collaboration

The impact of using AR for remote collaboration was measured using performance and workload for most of the studies, performance being quantified using task completion time (11 studies) and the number of errors made during the task (three studies) and workload using NASA task load index (NASA-TLX) survey tool (three studies). Out of the 11 studies that measured the time task time, only six studies directly compared some kind of AR versus other modes for collaboration. All six studies found a significant decrease in the task completion time when some kind of AR was used compared to their counterparts. Of the three studies that measured the number of errors, two found fewer errors when AR technology was used for collaboration, and one study did not find any significant difference. Out of this, one study that compared the number of errors while using paper-based instructions with AR-based remote support found that paper-based support led to more errors in completing remote maintenance. Another study that used AR for remote collaboration found that using interactive virtual collaboration (IVC), which includes using enhanced visualization support using arrows and sensory data, found fewer errors than not using IVC while completing a facilities management task. Out of the three studies that measured workload, one study that compared AR with other modes of communication found that the workload of the user decreased when AR technology was used for collaboration. The other two studies did not compare AR with other modes; instead, one compared visual communication cues in AR and found that user-centric cues (sharing view frustum from the local user and controller annotation from the remote user) reduced the workload compared to the device-centric one (sharing eye gaze from the local user and natural hand gesture from the remote user)da. The second study compared view-sharing techniques in AR and found that when 3D model techniques were used, the workload was less compared to the 2D video and 360 video conditions.

One study compared HMD, tablet, and projector for achieving AR, finding that using projector-based AR resulted in a shorter task completion time than the other two (Aschenbrenner et al., 2018). Le Chenechal et al., (2016) found that using a virtual arm for guidance (Vishnu) resulted in a shorter task completion time compared to remote sketch support for a complex task. Further, they also found that this system provided an easier mapping process between the guided instruction and the physical task interaction space than the remote sketch support for both simple and complex tasks. Wang & Dunston, (2009) found that the performance time when using AR systems to identify errors in a pipe layout design is significantly less than the current methods used for collaboration. They compared an AR-based face-to-face system (table-top AR system) to paper drawing and an AR-based virtual space system to NavisWork Roamer for collaborative error identification. For both AR systems, the time taken to identify the design errors was less than that of the paper drawing or the NavisWorks Roamer. The AR-based face-to-face system reduced the performance time for error detection by 64%, whereas the AR-based virtual space system reduced the performance time by 67%.

Zenati-Henda et al., (2014) found that AR collaboration using hand gestures in conjunction with HMD provides additional flexibility as hands are free, but the application control was difficult. Sasikumar et al., (2019) introduced a wearable AR remote collaboration system, the Wearable Remote Fusion, finding that this system not only reduced the workload of the participants but also, by sharing augmented natural cues, gave the participants a strong feeling of co-presence. Olorunfemi et al., (2018) found that using AR technology on construction job sites increased the accuracy of risk communication compared to such traditional communication methods as phone calls, emails, face-to-face, and video conferencing, in addition to finding a strong correlation between the accuracy of risk communication and the first three methods of communication, and a weak relation for the fourth.

Obermair et al., (2020) found that for an industrial maintenance process of an industrial PC, AR-based remote support reduced the number of errors, and the users were able to complete the complex tasks faster than paper-based instructions. The findings of a study conducted by Wang et al., (2019) supported the potential for training workers in the manufacturing industry using an AR remote collaborative platform in conjunction with 3D CAD models. The advantage of SAR/VR technology is that it permits the local workers to see the virtual cues without wearing any AR devices. The study conducted by Lee et al., (2020) involving multiple local users wearing AR headsets supervised by a remote expert wearing a VR HMD evaluated the view-sharing techniques. They found that the 3D model augmented with 2D video was productive and functional and reduced effort as well as was preferred. They further provided guidelines for implementing view-sharing techniques for pragmatic use. Ammari et al., (2019) proposed an innovative BIM based approach to support a large-scale facilities field task. The AR framework for facilities management they developed includes a field AR module and an office immersive augmented virtuality module. Field task efficiency is improved by minimizing the data entry time and the errors using interactive virtual collaboration compared to the non-collaborative mode. In addition, the ability to remotely mark a building element in an AR environment successfully was beneficial for visually interacting and communicating with a remote co-worker.

Barriers associated with AR for remote collaboration

Since AR technology is still a fairly recent development, there are various challenges associated with using it. Zenati-Henda et al., (2014) investigated remote AR collaboration using hand gestures and found that using HMD provides less flexibility for application control. To address this issue, they plan to use a hand gesture recognition technology that allows better interaction between the worker and the expert while using the HMD. According to a later study conducted by Le Chenechal et al., (2016) that used virtual arms to assist agents at a remote location, this approach is helpful only if the agent is close to the expert's virtual location. Once the distance increases, the expert cannot control the virtual targets using the virtual arms. To address this problem, these researchers used color codes from red to green based on the distance between the two users. They also found that wearing the see-through HMD is less comfortable compared to the screen-based setting since the perception of the real world is slightly altered, meaning a learning process is needed to be comfortable.

Other challenges mentioned in this study include collocation, inverse kinematics, perception of remote collaborators, and gesture coordination. Le Chenechal et al., (2016) proposed using stereo cameras for 3D reconstruction and replacing the optical tracking system with image-based tracking SLAM and/or a 3D model. Several studies also mentioned that challenges such as the lack of experience using AR devices might affect the validity of the results (Le Chenechal et al., 2016; Olorunfemi et al., 2018) and Obermair et al., (2020) found that for simple tasks in the industrial PC repair process, paper-based instruction is more efficient than AR-based remote support. In addition, according to the research conducted by Ammari et al., (2019), mobility support is lacking in the latest AR headsets, and they support only visual or vocal command input functionality, a challenge when using AR module functionality to interact with the AR contents. They also found that augmenting BIM elements while working on critical tasks may cause confusion and could impact the workers' safety.

A pilot study by Lee et al., (2020) found that the participants reported three challenges in the view-sharing techniques that they subsequently modified. First, in the 2D video-sharing technique, the large screen size was an issue as it was difficult to perceive and understand the image at a glance. To address this issue, they disabled the full-screen mode. In the 360-video sharing technique, an issue of image distortion caused by the video projection on the sphere makes it difficult to identify the distant object; by adding zoom-in and zoom-out functions, they could resolve this problem. Third, in the 3D model sharing technique, participants commented that the pre-constructed 3D model did not show the changes completed on the real project site. In addition, participants also complained about the visibility of the billboard, saying that it was only visible when their view direction was close to the avatar and that it was difficult to switch between avatars. This issue made it difficult for one-to-many collaboration scenarios, the primary focus of the research. To address this, the researchers added a function facilitating the ability to switch between the video billboard and the video projector.

Discussion

This section focuses on discussing the data that was analyzed from the 12 studies that were identified. The discussion is divided into four sections: 1) Human aspects, which talks about the human factors aspects of using AR for remote collaboration, 2) Technology interventions, which discusses the technology used in the identified studies, 3) Collaborative work, which discusses the collaborative aspects while using AR for remote collaboration, and 4) A framework for remote collaboration using AR technology which incorporates the human, technology and environmental aspects.

Human Aspects

AR has been found to improve a person's ability to perform their tasks by reducing time and the number of errors. With AR remote support, complex maintenance tasks are addressed efficiently without an expert traveling to the location, which is both cost- and time-intensive. The use of AR for this purpose is more efficient, usable, flexible, and less demanding for the operator, in addition to completing the task with less workload. Further, the use of AR has reduced the average time to communicate between the workers on the job site, which is important since communication is critical in any construction project.

100% of the reviewed studies comparing the task time between AR technology and other modes found that the task completion time was shorter for the AR condition. For example, Aschenbrenner et al. (2018) found that the task completion time was significantly less when SAR technology was used compared to the collaboration over the phone. This shows the effectiveness of SAR in improving performance while collaborating to complete a task. The AR facilitates communication in an effective manner without any confusion since the remote user can point to a particular component effectively compared to a phone call. Also, the video, while using an HMD, provides a subjective camera angle, which helps the remote expert to understand the problem at hand in an improved manner and can provide instructions in a clearer manner.

Three studies measured the number of errors for performance, and 66% of them reported that the errors were fewer when AR technology was used. This is because of the mistake the users made while using the other modes to complete the task. For example, the study conducted by Obermair et al., (2020) found that the users made errors while replacing a heat pipe using the paper-based instructions since half of the users opened the wrong page for instruction, while the other half removed the wrong components. This mistake did not happen when the AR module was used mostly since the expert could see what the user was doing and was guided accordingly. This shows the effectiveness of AR for remote collaboration to reduce the number of errors while completing a task. More research needs to be done to understand the effect of AR technology in reducing the number of errors since one study did not find any significant difference.

Out of the 12 studies reviewed, only one study compared the workload of AR technology with other modes of collaboration and found that the workload was less when AR was used (Aschenbrenner et al., 2018). Specifically, the SAR condition had the least workload compared to the other modes of collaboration compared. This could be because all the other modes of collaboration had to interact with some kind of device, such as an HMD or tablet, whereas in the SAR condition, once the device is set up, the user doesn’t have to interact with it. Moreover, it has the advantage of visual cues compared to the phone condition.

The research included in this review validates that AR-based technology is sustainable in the working environment. This technology is evolving and creating more opportunities to increase the productivity of construction workers; however, there is a potential for improvement in the area of ergonomic factors and arrangement. Although AR has found several applications in the construction industry, few studies focused on evaluating the human factors aspects during the interaction with the AR system, emphasizing the need for more research on how to increase the movement and efficiency of the user while using this technology.

Technology Intervention

The construction industry is experiencing transformative benefits from integrating AR technologies, ranging from safety training to risk mitigation and enhanced productivity. The deployment of AR applications necessitates the integration with sophisticated hardware. So, researchers are increasingly focusing on smart devices such as AR headsets (HoloLens, Meta Quest2, Magic Leap) and smart glasses for their advanced automation applications.

The importance of AR for remote collaboration in the AEC industry is evident from the reviewed studies, which highlight the adaptability and versatility of AR in addressing unique challenges. For instance, the Microsoft HoloLens has been effectively employed for safety hazard communication, enabling users to interact with a holographic collaborative environment seamlessly, as highlighted by Olorunfemi et al., (2018). The use of AR is significant for the growth of the construction industry, and as Hala et al., (2020) suggest, more research is needed for the holistic integration of AR in the AEC industry.

The commonly used remote collaboration technologies primarily offer basic audio and video communication. However, integrating these with AR can create an immersive environment, fostering a sense of co-presence with experts. Hand gestures, while providing a real-time collaboration system through AR, have their limitations, necessitating further research to enhance their controllability and introduce additional features. This is evident in the studies by Le Chenechal et al., (2016) and Sasikumar et al., (2019).

SAR is effective for remote maintenance tasks due to its spatial awareness and shared experience capabilities, as seen in the works of Aschenbrenner et al., (2018) and Wang et al., (2019). SAR's practicality and precision across industries make it invaluable in error prevention and detection. Its ability to overlay AR content directly onto real-world objects without the need for wearable devices offers a hands-free experience, especially beneficial for tasks like assembly. While AR glasses, such as HoloLens, are recommended for hands-free operations and have shown accuracy in construction risk communication, they come with their set of challenges. Mobility and input functionality issues can hinder their adoption in work environments. Data entry tasks, which are more straightforward on traditional desktop setups, become cumbersome on AR devices. Also, cluttered visuals on AR glasses can overshadow real-world information, compromising situational awareness. Properly curating the information displayed on AR modules is essential to maintain clarity and effectiveness. As the industry moves towards a hands-free operational model, innovative methods for projecting AR content onto surroundings will be crucial.

Another important consideration when using AR in construction involves the integration of artificial intelligence (AI) agents to assist the workers as they complete various tasks. AI could be used in construction, for example, for visual inspection (defect detection using BIM) and visual site exploration and to compare as-planned vs. as-built project status (Abioye et al., 2021). While there are several applications of AR supported by AI in manufacturing applications (Sahu et al., 2021), more research is needed to understand how workload, performance, and situational awareness of the workers are affected when using these technologies together.

In conclusion, the adoption and integration of AR in the construction industry are both promising and challenging. While the benefits are manifold, addressing the associated challenges will be significant for effective implementation of AR in the sector.

Collaborative work

In the domain of collaborative work, a multitude of factors come into play. Patel et al., (2012) outlined these into seven principal categories: context, support, tasks, interaction processes, teams, individuals, and overarching considerations. This discussion explores the factors within these categories that are pertinent to the studies that were reviewed.

Contextual Factors. This includes culture and environment, with the former referring to the type of profession and the latter to the physical space where individuals work. The majority of the studies reviewed here were conducted with an expert working remotely with a non-expert to complete a particular task. For example, Ammari et al., (2019) investigated the effectiveness of using the AR module at the local site to receive help from an expert at a remote location to replace a thermostat attached to a wall.

Supportive Elements. This involves the tools required for collaboration, training, and knowledge management. Tools refer to the technologies provided for collaboration; here, the use of AR for remote collaboration has been found to be more effective compared to other mediums. The AR was achieved using devices such as Microsoft HoloLens, (G. Lee et al., 2020), Magic Leap One (Sasikumar et al., 2019), and tablets (Ammari et al., 2019). However, for the collaboration to be effective, providing training on the newer technologies and tools for the parties involved is important. If they are new to the system, there will be complications in the beginning, but they should resolve over time. The parties involved should also have access to the knowledge database of the system they are using.

Task-Related Considerations. This involves the type, structure, and demand of the task. Task type may affect performance since the mental demand required will be different for the various team members. This review identified the following tasks: search and assembly, industrial maintenance, error detection, repair tasks, assembly training, risk communication, and hazard identification, among others, all of which have benefitted from using AR technology for remote collaboration. The structure of the task is also important in collaboration. Some of the tasks identified from this review were simple, while others were complex. For example, Obermair et al., (2020) found that simple tasks did not benefit from the AR technology compared to traditional paper-based support. This might be attributed to the unnecessary use of technology for a straightforward task, which could, in fact, be easily accomplished using conventional methods. This area should need further study focusing on the various tasks in different domains of application. The demand that the task requires from the individuals is also an essential contribution to effective collaboration.

Interaction Processes. This refers to how individuals collaborate to complete a particular task and includes learning, coordination, communication, and decision-making. Learning occurs through training or while completing a task multiple times. It is especially effective through collaboration, especially when an expert helps a novice complete a task; while the worker is learning from the expert, the expert is also learning something with every new scenario. Coordination is important when working as a team to achieve shared goals. To emphasize the importance of this factor, Lee et al., (2020) explored using remote collaboration to connect one expert with two local users. Similarly, communication between parties is another determining factor for effective collaboration, as the lack of proper communication can lead to errors and low team performance. Through proper communication, knowledge is transferred, and shared awareness is maintained (Patel et al., 2012). In AR, there are different methods to communicate than just using audio and video. For example, Sasikumar et al., (2019) investigated the impact of employing user-centric and device-centric cues on enhancing communication using AR, with the ultimate goal of enhancing interaction between two users situated in different locations. This review found that AR is an effective mode of communication when using remote collaboration to complete various tasks in the AEC domain, as effective communication and collaboration result in better decision-making.

Team Dynamics. This includes subfactors such as roles, relationships, and shared awareness, among others. In this review focused on remote collaboration using AR, most of the studies included an agent helping a remote worker complete a task, with the roles of each being clearly defined. The relationship between these members of a team is important for team performance and effective collaboration. Shared awareness of the issue faced by the worker at the location is important for remote assistance because if it is not correct, the solution or help provided by the remote worker will be ineffective and time-consuming. More research in understanding shared situational awareness or team situational awareness will help in designing effective AR systems.

Individual performance. This is the most measured metric that we found from the reviewed papers in this study. The individual performance contributes highly to the effective completion of a task. In a remote assistance task, the individuals at both ends possess different skills to complete the task, especially in a construction setting. The expert guiding the worker should possess the experience required to understand the problem at hand and the skill set to solve it. Similarly, the worker should possess the skill to execute the instructions provided by the expert to complete the task. The majority of the studies in this review focused on the performance of the local user. Future work could investigate the performance of the remote expert and team performance while collaborating over the AR module to complete a task.

Overarching factors. There are other overarching factors that contribute to collaborative work performance, for example, trust, experience, goals, constraints, and incentives, among others. Trust is an important factor since, for successful remote collaboration, the worker must trust the remote expert to follow their directions. However, none of the 12 papers analyzed in this review investigated the trust of either local users or remote experts, which is important because it is a critical element for effective teamwork. With the advancement in technology, AI agents are now assisting workers to complete various tasks, and research focused on trust in AI agents while completing a procedural task would provide valuable information about trust for human teams integrated with such agents.

Framework for remote collaboration using AR technology

The information flow in remote collaboration between a local user and a remote expert is a cyclical process, as shown in Figure 3. It begins with the local user who uses an AR module to collect visual and audio data, which he transmits over a server to the remote user. The remote user, who receives the data on a computer, processes it and then gives instructions, which are received by the local user on the AR module. If the local user agrees with the suggestions, those actions are executed and sent to the remote user over the server. This process continues until the problem is fixed or the intentions are satisfied.

Figure 3:

Remote Collaboration Process Flow

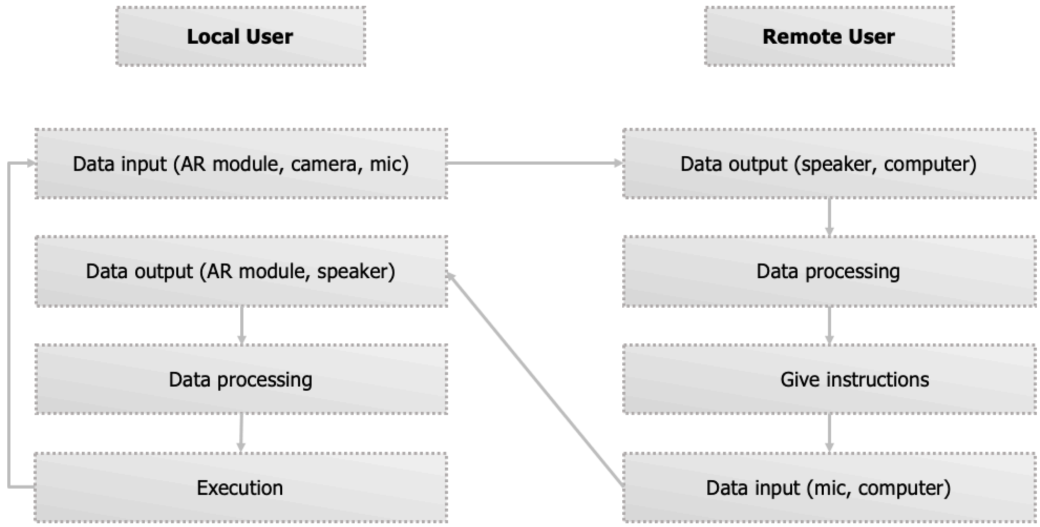
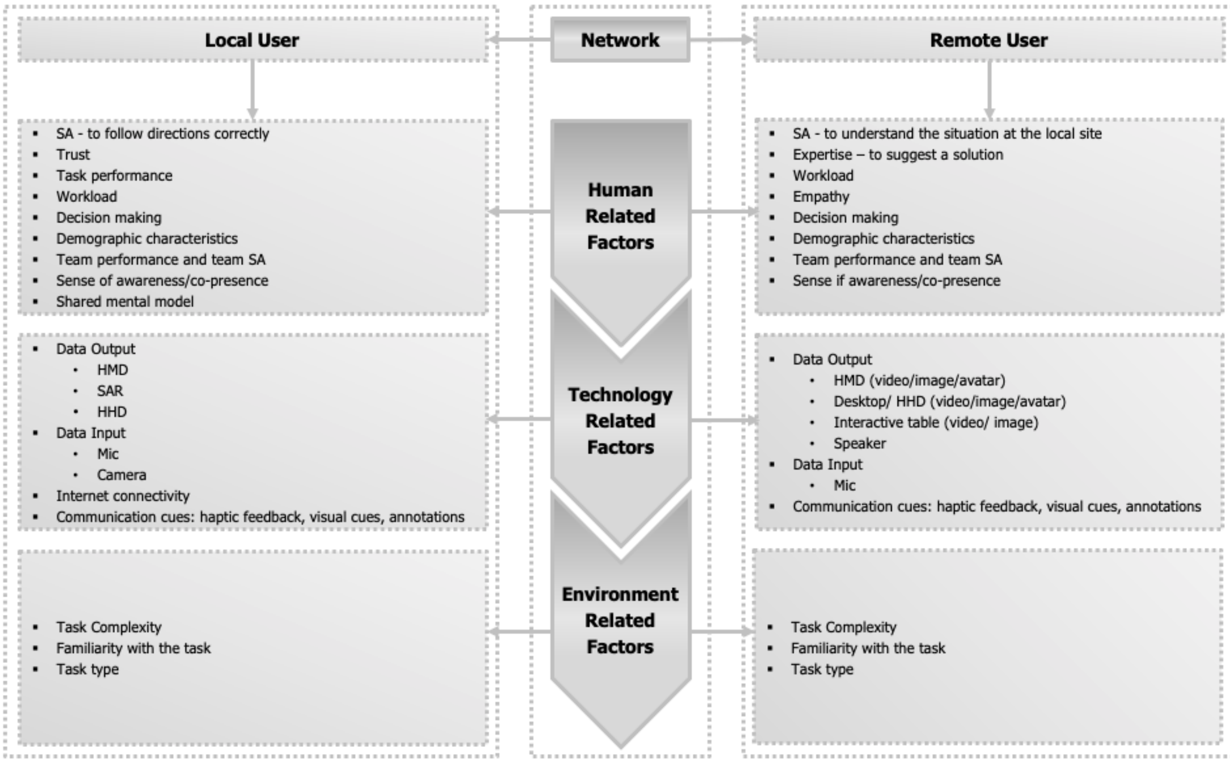


Figure 4 shows the proposed framework for remote collaboration using AR technology, which is impacted by human-, technology-, and environment-related factors. Human-related factors include situation awareness (SA), trust, workload, and task performance, among others. Effective collaboration requires effective team performance, achieved through effective communication, a sense of awareness, a sense of co-presence, team situational awareness, and shared mental models, among other factors.

Figure 4:

Proposed Framework for Remote Collaboration using AR



At the local location, SA is required for the user to understand the problem at hand and execute the directions provided by the remote expert correctly. The local user's trust in the remote expert also affects the remote collaboration as this factor is needed between the collaborators for an effective exchange of information and execution. Task performance and the workload of local users are other factors that need to be considered for effective collaboration: an increase in workload could affect the performance of the user in completing the task as well as his decision-making capabilities.

For the remote user, SA is an important aspect of collaboration, especially in understanding the problem faced by the local user at a different location. The remote user should have enough expertise to give instructions for or to suggest solutions to the problems faced by the local user. In addition, the remote user should have empathy for the user to understand the situation at the local location. This can lead to effective decision-making and, thus, effective collaboration.

Technology-related factors influencing effective remote collaboration include input and output devices, internet connectivity, and such communication aspects as visual cues, haptic feedback, and annotations, among others. Input devices include microphones and cameras, while output devices include HMDs, HHDs, SAR, desktops, and speakers. These factors affect the human-related factors, in turn affecting the effectiveness of the collaboration between the parties involved.

Environmental-related factors that contribute to effective collaboration include task complexity, task type, and the users' familiarity with the task. Other aspects include multi-tasking requirements for the task at hand and the physical environment, such as lighting and noise level.

The systematic review conducted identified several gaps in the literature and avenues for future research in the context of AR-integrated work environments within the AEC industry. These gaps include the need for more robust user training programs to facilitate seamless AR adoption, research focused on enhancing shared situational awareness, understanding the development of trust between the local and remote users, addressing integration challenges into existing workflows, and addressing privacy and security concerns. Also, with the evolving AI technology, the potential of AI as a recommender system to support AR users, potentially reducing their reliance on remote experts, demands thorough investigation. Addressing these gaps can contribute to its successful implementation in the AEC industry, ultimately fostering improved productivity and safety.

Conclusion

This systematic review has analyzed the application of AR for remote collaboration with a focus on the impact of the device or technologies used on the user's performance. AR technology is very versatile, and as a result, it is increasingly being used in the construction industry for a variety of applications, ranging from safety training collaboration to an efficient working environment and many more. This review identified several important considerations concerning the different devices to apply for implementing AR, in addition to finding that its use resulted in a significant reduction in the time needed to complete the task, the number of errors, and the workload. Further, accuracy, risk, and communication among the participants were remarkably improved. However, the use of HMD for the application of AR was not flexible nor comfortable as the perception of the real world was altered, and hand gesture coordination was challenging. Further research is required to address the mobility challenges of the AR operating devices, and more advanced study explicitly exploring the techniques for sharing the data using AR operating devices remotely is important for increasing the efficiency and productivity of the AEC industry.

After investigating studies on collaborative AR support in the AEC domain, it's evident that there is limited research specifically focusing on the remote expert and the trust between the remote user and the local user. In this era of technological advancement, integrating AI into the AR landscape has emerged as a promising avenue for assisting workers by providing recommendations. This integration can potentially reduce the local user's reliance on remote experts when faced with uncertainties during task execution. However, successfully incorporating AI into any technological framework necessitates thorough investigation, considering the complex factors that could influence its effectiveness. Accordingly, the first study reported in Chapter 3 aims to investigate the effect of the reliability of the AI agent and the complexity of the task on the trust the humans have in the recommender system.

CHAPTER THREE

UNDERSTANDING TRUST IN ARTIFICIAL INTELLIGENCE AGENTS WHILE COMPLETING A PROCEDURAL TASK

Introduction

The fourth industrial revolution is increasing the importance of AI. This digitalization era refers to the period in which we increasingly rely on digital technologies to connect the physical and virtual worlds, which includes using AI systems and AR/VR technologies to create a more active and seamless connection between these two technologies (Zhang et al., 2022). Industries are undergoing a significant transformation as they move toward digitalization, such as through the integration of AI in manufacturing for predictive maintenance, the use of AR for improving precision and efficiency in engineering design processes, and the application of VR for training and simulations in various sectors. AI is a rapidly evolving field that can potentially transform various industries and aspects of our daily lives. Healthcare, finance, transportation, manufacturing, and construction are some of the key sectors where AI has significantly impacted industries by improving efficiency, safety, collaboration, and training. Using computer systems, AI simulates many aspects of the human intellect, including learning, reasoning, and self-correction. As a result, machines can analyze massive amounts of data to identify patterns and base decisions based on that data (Gillath et al., 2021).

AR technology is another innovation that has significantly impacted various industries and has many applications. AR technology enhances the user’s perception by overlaying computer-generated information, such as pictures, text, and 3D models, onto the real world. AR application covers the fields of entertainment, maintenance, manufacturing, healthcare, and construction (Hincapie et al., 2011). AR is used to create immersive gaming experiences and interactive media content. On the other hand, this technology provides technicians with real-time information and guidance for maintenance tasks improving the accuracy and efficiency of assembly line operations and quality control. AR is used for project planning and design, safety training, and on-site construction management in the AEC industry. AR technology provides architects, engineers, and construction workers with a visual representation of the building design, allowing them to identify and solve potential problems before construction begins. Overall, AR technology has the potential to revolutionize various industries by providing new perspectives and improving the accuracy and efficiency of complex operations (Hincapie et al., 2011).

The rapid development of digital technologies has had a significant impact on the AEC industry, leading to improved efficiency and productivity. However, adopting these new technologies can be challenging, as the construction industry is complex, and projects often involve multiple stakeholders, each with unique requirements and needs. Incorporating new technologies in the construction process requires a high level of professional knowledge and expertise, and incorporating AI in the construction industry changes how a construction project performs (K. Wang et al., 2022). AI can automate several operations and increase the efficiency of the building process. AR, on the other hand, benefits the AEC industry in the areas of visualization, information retrieval, and interaction. It can help architects and designers to create more immersive and interactive visualization in the early phase of the design. It can help to reduce the risk and improve communication between the project stakeholders. AR is also used in the AEC domain for remote collaboration between two users, for example, an expert from a remote location providing recommendations to a user to complete a task (Bhanu et al., 2022). AR can also be used for information retrieval, allowing workers on construction sites to access important information quickly and easily. For example, with AR glasses, workers can view detailed plans and schematics while working on a building without carrying physical copies of documents. This can improve efficiency and reduce errors, resulting in cost and time savings in the design and construction process.

Though AR has improved construction efficiency, it has been challenging to adopt AR technology in the AEC industry (Devagiri et al., 2022). For example, some workers experience discomfort using HMDs on construction sites, prompting ongoing hardware-based research to address these limitations. Incorporating AI systems into the AR framework has its advantages, and their integration is becoming increasingly prevalent due to the rapid advancement in these technologies. For example, by combining AI algorithms with AR sensors and cameras, it is possible to automatically detect defects, deviations from design specifications, and other issues in real time, allowing for faster and more accurate inspections and quality control. However, issues caused by unexpected operations and AI system failures have occurred with the continual increase in the development and use of AI systems and applications. These errors have led to trust issues when using AI to predict and safely use services (Um et al., 2022). Trust can be extended not only to human beings but also to artificial intelligence systems, such as chatbots or virtual assistants. As more and more AI systems become a part of our daily lives, it is essential to consider how we can trust them to behave in safe, reliable, and ethical ways. This requires designing AI systems with transparency, accountability, and explainability in mind so that users can understand how the system works and why it makes certain decisions. A lack of trust in AI systems can lead to underutilizing the technology.

Measuring and managing trust and foreseeing the potential risk associated with AI is essential for increasing transparency and accountability. However, controlling and assessing the dependability of AI systems and algorithms is highly difficult for a variety of reasons (Gasser & Almeida, 2017). Measuring the reliability of AI can be challenging, especially as the level of machine intelligence increases. One of the reasons for this difficulty is that machine learning algorithms can exhibit different behaviors depending on the data they are trained on, even if the underlying objective function remains the same (Glikson & Woolley, 2020).

However, one of the criteria used to decide whether to believe an agent is their reliability: 'In judging that someone is reliable we look to their past performance; in placing trust in them we commit ourselves to relying on their future performance' (O'Neill, 2002). Lee & See (2004) found that reliability is an important determinant of trust in automation. Their study showed that users are more likely to trust and use an automated system if it consistently performs its tasks accurately and reliably. This finding is echoed in the work of Hoff & Bashir (2015), who found that perceived reliability significantly influences trust in AI systems. AI agents are tools or systems that can enhance human capabilities and help solve complex problems. Although trust may be developed, conferred, or rejected based on reliability and experience, this is not the only or defining feature of trust. While prior performance is a basis for reliability, it is not the only factor when deciding whether to trust someone. The responsibility burden falls on those responsible for creating, implementing, and employing reliable AI (Ryan, 2020).

Trust is a dynamic process that can change over time depending on several factors rather than a static or fixed concept. How much trust a person is willing to place in another person or an AI system can be influenced by a person's disposition, experiences in the past, and personal biases. The trustworthiness of an AI system may also depend on its capacity to carry out the task at hand and provide the promised outcomes. It is important to understand that trust is a complicated and multifaceted idea that various factors, including the complexity of the task, can influence.

The construction industry is a complex domain that includes several procedural tasks. Procedural tasks involve completing a series of actions by interacting with the physical environment to achieve selected objectives (Henderson & Feiner, 2011). These tasks range from simple to highly complex, often requiring specialized expertise to assist the workforce in their execution, depending upon the nature of the task and the experience of the individual worker. Advancements in digital technologies, such as BIM, AI, and AR, have significantly enhanced the efficiency with which these procedural tasks are performed. For example, AR has been shown to help complete an assembly task significantly faster and with fewer errors (Henderson & Feiner, 2011).

As mentioned earlier, AR technology is used for remote collaboration to assist in completing a task for a user at a site by an expert from a remote location without needing to travel. Integrating a well-trained recommender AI agent into the AR technology could eliminate the need for a human expert to assist in completing a task. The recommender AI could only be beneficial if it can provide recommendations at or above the level of the expert it replaces. There is a gap in the literature that investigates the effects of the reliability of the recommender AI and the complexity of the task while completing a procedural task while using the AR technology embedded with the recommender AI. This study tries to fill this gap and aims to explore how the reliability of a recommender AI agent and the complexity of the task impacted overall trust, trust development, user performance, and workload while completing a procedural task with the help of AR and AI technologies. Specifically, the study focused on a procedural construction task and evaluated how trust levels changed when the recommender AI agent's reliability varied across different conditions. The results from this research have a significant potential to inform the design and implementation of recommender AI systems on an AR technology for completing a procedural task.

Research Questions:

1. How does trust vary for the varying reliability of recommender AI and the complexity of tasks while completing a procedural task with the help of augmented reality technology?
2. How is performance affected when the reliability of the recommender AI and the complexity of tasks varies while completing a procedural task with the help of augmented reality technology?
3. Does the workload increase with a decrease in the reliability of the recommender AI?

Hypotheses:

1. The trust in AI increases with an increase in the reliability of the recommender AI.
2. The task performance could be higher while interacting with a high reliable recommender AI.
3. The workload increases with a decrease in the reliability of the recommender AI.

Partially observable Markov Decision Process (POMDP) Model for Trust in AI

In this section, we modeled human trust as a Markov model. A Markov model is a mathematical model that has been used to represent the state of a system that changes over time (Grassmann, 1983). In the Markov model, the probability of transitioning from one state to another does not depend on previous states but solely on the current state. The state represents the condition that the system can be in at a given point in time. Markov models can be used for predicting the dynamics of human trust in AI agents as the trust changes over time during the interaction with the AI agent. POMDP is one type of Markov model used for decision-making problems where the state of the system is partially observable. Human trust during the interaction with AI systems is uncertain and dynamic, which is influenced by different factors such as reliability of the AI, the context, the user's previous experiences etc. POMDP model could handle situations where the state of the system, which in this case is the trust users have in AI, is not fully observable and evolves probabilistically over time (Chen et al., 2018). POMDP models have been previously used in the literature to model trust during human-robot interaction (Chen et al., 2018; Tilloo et al., 2022; Williams et al., 2023). In this study, we model the human trust during the interaction of the user with a recommender system. In POMDP, the system does not know what the human trust level is at any time step, and the state level is changed based on the feedback it gets.

In this experiment, the recommender AI gave recommendations for the next component. The humans can either follow the agent's recommendations or, if they believe the recommendations to be incorrect, they have two options: one, to proceed with the pipe fitting that they think is correct based on their knowledge and experience, and two, to refer to the map of the pipe layout to determine the correct pipe fitting. The AI agent will get feedback on the trust that the humans have in them based on the actions that the humans take. If the humans follow the recommendations given by the agent, the AI thinks that the humans trust the agent. In contrast, if the humans don't follow the recommendations provided by the agent, the AI understands that the trust in them has decreased. The trust level of humans is updated based on the correctness of the recommendations given by the agent.

A mathematical framework is developed to model the decision-making based on POMDP (Kaelbling et al., 1998). In this context, the POMDP is defined by the tuple (S, A, Ω, T, R, γ, b_1). S is the set of states and represents all possible values of human's trust in the AI agent, and the human's trust as perceived by the AI agent. A is the set of agent's actions, which is recommending the correct component or a wrong one. Ω is the set of observations the agent can make about the environment and the human's actions and help the agent infer the current state of the environment and the human's trust level. T is the transition function and defines the probability of transitioning from one state to another given a particular action and is influenced by the agent's actions and the human's reactions. R is the reward function and specifies the reward for each action taken at a given state. γ is the discount factor considered for the future rewards, and b_1 is the shape parameters of the initial trust probability distribution. In a finite-horizon POMDP, the goal is to optimize the expected reward. In order to optimize the model, we first need to describe the value function and action function, crucial tools for analyzing a POMDP. The value function is the optimal expected cumulative reward starting from state s_k at time k and is represented as $V_k(s_k)$.

$$V_k(s_k) = \max_{a_k^r \in A}\{ R(s_k, a_k^r) + \gamma \sum_{s' \in S} T(s_{k+1} = s' | s_k, a_k^r) V_{k+1}(s_{k+1})\} \tag{1}$$

The action function is the action taken to achieve $V_k(s_k)$ starting from state s_k at time k and is represented as $A_k'(s_k)$.

$$A_k'(s_k) = \operatorname*{argmax}_{a_k^r \in A}\{R(s_k, a_k^r) + \gamma \sum_{s' \in S} T(s_{k+1} = s' | s_k, a_k^r) V_{k+1}(s_{k+1})\} \tag{2}$$

Action Space and State Space

The action space is $A = \{0, 1\}$, which is a set of all possible actions of the agent. The values of a_k^r could be 1 and 0 that implies that the agent's recommendations are correct and incorrect respectively. Let their respective probabilities be d_k and $1 - d_k$. The state space is $S = [0,1]$ which is the set of all possible trust values perceived by the agent. The state space consists of trust at any point in time. The state of human trust perceived by the AI agent is represented as:

$$s_k = \frac{\alpha_k}{\alpha_k + \beta_k} \tag{3}$$

Trust Dynamics Model

An important component of this model is the trust dynamics. Trust, before each interaction, is assumed to be a Beta distribution and can be represented as:

$$t_k \sim Beta(\alpha_k, \beta_k) \tag{4}$$

The shape parameters α_k and β_k are updated based on the agent's performance and its observations of the human's actions. The belief at time k, which characterizes the trust t_k, is represented as $b_k = (\alpha_k, \beta_k)$. The belief is updated based on the actions of the humans and the AI agent's recommendation. This can be represented as p_k where its value could be 1 if the human accepts the recommendation or 0 if the human doesn't follow the recommendation the agent provided. Let w^f and w^s be the experience gains due to the human's performance. The trust dynamics could be formulated as:

$$(\alpha_{k+1}, \beta_{k+1}) = \begin{cases} (\alpha_k + w^s, \beta_k), & if\ p_k = 1 \\ (\alpha_k, \beta_k + w^f), & if\ p_k = 0 \end{cases} \tag{5}$$

We denote the state of the system as s_k at any step k. The human's trust, t_k^h is represented as the state space s_k^h, which is updated based on the agent's performance in recommending the correct recommendations, and the human's belief as $b_k^h = (\alpha_k^h, \beta_k^h)$. We assume that agent perceives the same amount of trust humans have in them, i.e., $s_1^h = s_1$. The update rule for the human is the same as (5). The state remains the same if the human receives no feedback about the agent's performance.

State Transitions

The state transitions happen based on the robot's observation of the human's action and the agent's performance.

Trust-behavior Model

We define a trust behavior model to explain this transition probability that relates the trust to human action. We define a Bernoulli random variable, z_k, whose value is 1 when the user follows the recommendation by the AI agent, and 0 when they don't. We assume the probability of $z_k = 1$ as $\psi(sk)$. Thus, the trust-behavior model can be shown as follows:

$$\pi(z_k \mid s_k) = \begin{cases} \psi(s_k), & if z_k = 1 \\ 1 - \psi(s_k), & if z_k = 0 \end{cases}$$

This implies that when the trust is high, the user is more likely to follow the agent's recommendations whereas when the trust is low, the user is less likely to follow the recommendations. When the trust is low, the user has two option to complete the task, either by clicking the map button to see the layout or by using their own knowledge to make a decision.

Transition Probabilities

The transition function depends on the interaction between the agent and represents the change of state of the agent. The state of the agent changes when the trust increases or decreases based on the feedback. Based on the values of a_k^r, the transition function can be written as:

$$\mathbb{T}(s_{k+1}|s_k) = \begin{cases} \begin{cases} \psi(s_k), & if\ s_{\{k+1)} > s\_k \\ 1 - \psi(s_k), & if\ s_{\{k+1)} < s\_k' \end{cases} & if\ a_k^r = 1 \\ \begin{cases} 0, & if\ s_{\{k+1)} > s\_k \\ 1, & if\ s_{\{k+1)} < s\_k' \end{cases} & if\ a_k^r = 0 \end{cases}$$

Reward Function

The reward function specifies the reward associated with each transition. It depends on the human action to accept or reject the recommendation, or if the recommendation is rejected, the human decision to use the map or their previous knowledge. This model uses a negative reward process, which is the time taken and construction damage. A reward is associated with each time step, depending on the action the human takes. In this model, a value iteration method is used where the model is run for many iterations until the policy where the total reward cannot be increased. The ultimate objective of the model is to identify the policy that yields the highest cumulative reward.

Human's Action

In the proposed model, the set of potential human actions at a given time k is represented by a_k^h and is defined as $A = \{0,1\}$. Here, an action $a_k^h = 1$ signifies the human's selection of the correct component, while $a_k^h = 0$ indicates the opposite. This decision-making policy is assumed to be influenced by the human's perceived confidence in the agent's performance.

For each time step k, the term $\widehat{d_k}$ represents the human's estimation of the agent's likelihood of success, denoted as d_k. Depending on this estimate, the human might either accept to the agent's recommendation (i.e.,$z_k = 1$) or reject it (i.e., $z_k = 0$). When the human diverges from the agent's recommendation, they might consult a guide map or rely on personal judgment knowledge to make the decision. This decision is modeled by a Bernoulli random variable m_k, where $m_k = 1$ indicates using the guide map and $m_k = 0$ suggests otherwise. It's posited that $m_k = 1$ with a probability g_k. When the guide map is consulted ($m_k = 1$), the human invariably selects the correct component. Conversely, without the guide map ($m_k = 0$), the correct component is chosen with a probability $\widehat{d_k}$. The subsequent section defines the probabilistic framework governing human actions in this scenario.

$$\pi^d(a_k^h | z_k = 0, m_k = 1, a_k^r, \widehat{d_k}) = \begin{cases} 1, & if\ a_k^h = 1 \\ 0, & if\ a_k^h = 0 \end{cases}$$

$$\pi^d(a_k^h | z_k = 0, m_k = 0, a_k^r, \widehat{d_k}) = \begin{cases} \widehat{d_k}, & if\ a_k^h = 1 \\ 1 - \widehat{d_k}, & if\ a_k^h = 0 \end{cases}$$

$$\pi^d(a_k^h | z_k = 1, m_k = 0, a_k^r = 1, \widehat{d_k}) = \begin{cases} 1, & if\ a_k^h = 1 \\ 0, & if\ a_k^h = 0 \end{cases}$$

$$\pi^d(a_k^h | z_k = 1, m_k = 0, a_k^r = 0, \widehat{d_k}) = \begin{cases} 0, & if\ a_k^h = 1 \\ 1, & if\ a_k^h = 0 \end{cases}$$

When $z_k = 1$, we have $a_h^k = a_k^r$, whereas $z_k = 0$ does not necessarily mean $a_k^h \neq a_k^r$. Also, the case $z_k = 1$ with $m_k = 1$ is impossible.

Immediate rewards

The aim of the model is to complete the task without any construction damage and to complete it in the least amount of time. A rational reward function for this purpose can be a weighted sum of the construction damage loss δ_c (a wrong component damages the pipe layout) and the time cost δ_t at step k. The immediate reward is defined as

$$R(s_k, a_k^r) = \sum_{a_k^h} \sum_{(z_k, m_k} \left(-\omega_c f_c\left(a_k^h\right) - \omega_t f_t(z_k, m_k)\right) \mathbb{P}(a_k^h, z_k, m_k | a_k^r, s_k, \widehat{d_k}, g_k \tag{6}$$

where, $$f_c(a_k^h = \delta_c^1 \, 1\{a_k^h = 1\} + \delta_c^0 \, 1\{a_k^h = 0\}$$

and $$f_t(z_k, m_k) = \delta_t^{1,0} \, 1\{z_k = 1\} + \delta_t^{0,0} \, 1\{z_k = 0, m_k = 0\} + \delta_t^{0,0} \, 1\{z_k = 0, m_k = 1\}$$

The reward probabilities $\mathbb{P}\left(a_k^h, z_k, m_k \middle| a_k^r, s_k, \widehat{d_k}, g_k\right)$ are shown in Table 6 which are obtained as $\pi^d\left(a_k^h \middle| z_k, m_k. a_k^r, \widehat{d_k}\right) \pi(z_k | s_k) \mathbb{P}(m_k | g_k)$. Also sample values for $\delta_c^1, \delta_c^0, \delta_t^{1,0}, \delta_t^{0,1}, \delta_t^{0,0}$ are (0, 400, 110, 100, 90). In this combination, when the correct component is selected, there is no construction damage, and the reward is zero units, whereas if the humans select the incorrect component, construction damage loss is 400 units. When the user follows the agent's recommendation, there is a minimum time loss of 90 units, whereas if the user uses a map, the loss is 110 units, and if the user uses their knowledge to complete the task, the time loss is 100.

Table 6:

Reward Probabilities

Follow agent's recommendation	Use map	Human's action	Agent's action	
			$a_k^r = 1$	$a_k^r = 0$
$z_k = 1$	$m_k = 0$	$a_k^h = 1$	$\psi(s_k)$	0
		$a_k^h = 0$	0	$\psi(s_k)$
$z_k = 0$	$m_k = 1$	$a_k^h = 1$	$[1 - \psi(s_k)]g_k$	$[1 - \psi(s_k)]g_k$
		$a_k^h = 0$	0	0
	$m_k = 0$	$a_k^h = 1$	$[1 - \psi(s_k)](1 - g_k)\widehat{d_k}$	$[1 - \psi(s_k)](1 - g_k)\widehat{d_k}$
		$a_k^h = 0$	$[1 - \psi(s_k)](1 - g_k)(1 - \widehat{d_k})$	$[1 - \psi(s_k)](1 - g_k)(1 - \widehat{d_k})$

The pseudo-code of the algorithm can be found below:

Algorithm 1 Value iteration method to solve the POMDP model

Input: POMDP with states S, actions $A(s)$, transition model $P(s'|s, a)$, rewards $R(s)$), discount γ, initial belief b_0, initial state s_0

Output: Optimal policy π^*, Reward = 0

Initialize $\pi^* = \{\}$

for $j = 1, \dots, n_{steps}$ **do**

Find the best possible action $A(s_j)$ and maximum value $V(s_j)$ at step j as per (2) and (1), respectively

Generate human decisions, i.e. accepting or rejecting the recommendation, using guide map or not and human action based on own knowledge

Get reward R_j at step j based on the agent' recommendation and human decisions

Reward = Reward + R_j

Update agent's perceived trust and human's actual trust state level as per the belief update rule in (4)

Add the new state, i.e., $S_{\{j=1\}}$ to π^*

end for

Method

Participants

This study was approved by Clemson University's Institutional Review Board (IRB). We recruited 36 participants (21 male, 15 female) aged between 36 and 21 (M = 24.69, SD = 4.31) for this study who were students enrolled in undergraduate or graduate degrees in civil engineering, architecture, and construction science and management. They were recruited through email, word of mouth, and by announcing in the classes. Every participant went through four different conditions, but the order of the conditions was randomly assigned using a Latin Square design. Once they completed the study, they were remunerated with 3% of their overall grade towards the total grade or $10 for their time. Their participation was voluntary, and they were told they could discontinue the experiment anytime. Table 7 shows more about the demographic information.

Table 7:

Demographic Information

Variable (N = 36)	**N**	**%**
Gender		
Female	15	41.7
Male	21	58.3
Race		
African American	1	2.8
Hispanic/Latino	2	5.6
Asian	10	27.8
Caucasian/White	23	63.9
Degree Pursuing		
Graduate	12	33.3
Undergraduate	24	66.7

Apparatus and Materials

The experiment was conducted with the help of a simulation that was developed using the Unity game engine. The software simulated the interaction involving participants completing a pipe fitting task using augmented reality glasses. However, they did not actually wear an AR headset; instead, the interaction was replicated on a desktop computer as shown in Figure 5. This simulation was run on a Dell desktop computer with an Intel(R) Xeon(R) CPU E5-1620 v4 processor, a Quadro FX 5800 GPU, and an ultrawide LG monitor with a screen size of 38 inches. There was a pre-test questionnaire, a trust level questionnaire, a workload measurement, and an overall trust questionnaire. The pre-test and the overall trust questionnaire were administered using Qualtrics survey software. The trust level questionnaire was a 7-point Likert scale to measure trust in AI at a given time and was administered within the simulation. The overall trust questionnaire and workload assessment were administered after every condition using a validated self-reported 12-item set of Likert scales and the National Aeronautics and Space Administration-Task Load Index (NASA-TLX), respectively (Hart & Staveland, 1988; Jian et al., 2000). Figure 5 shows the experimental setup for this study.

Figure 5:

Experimental Setup

Procedure

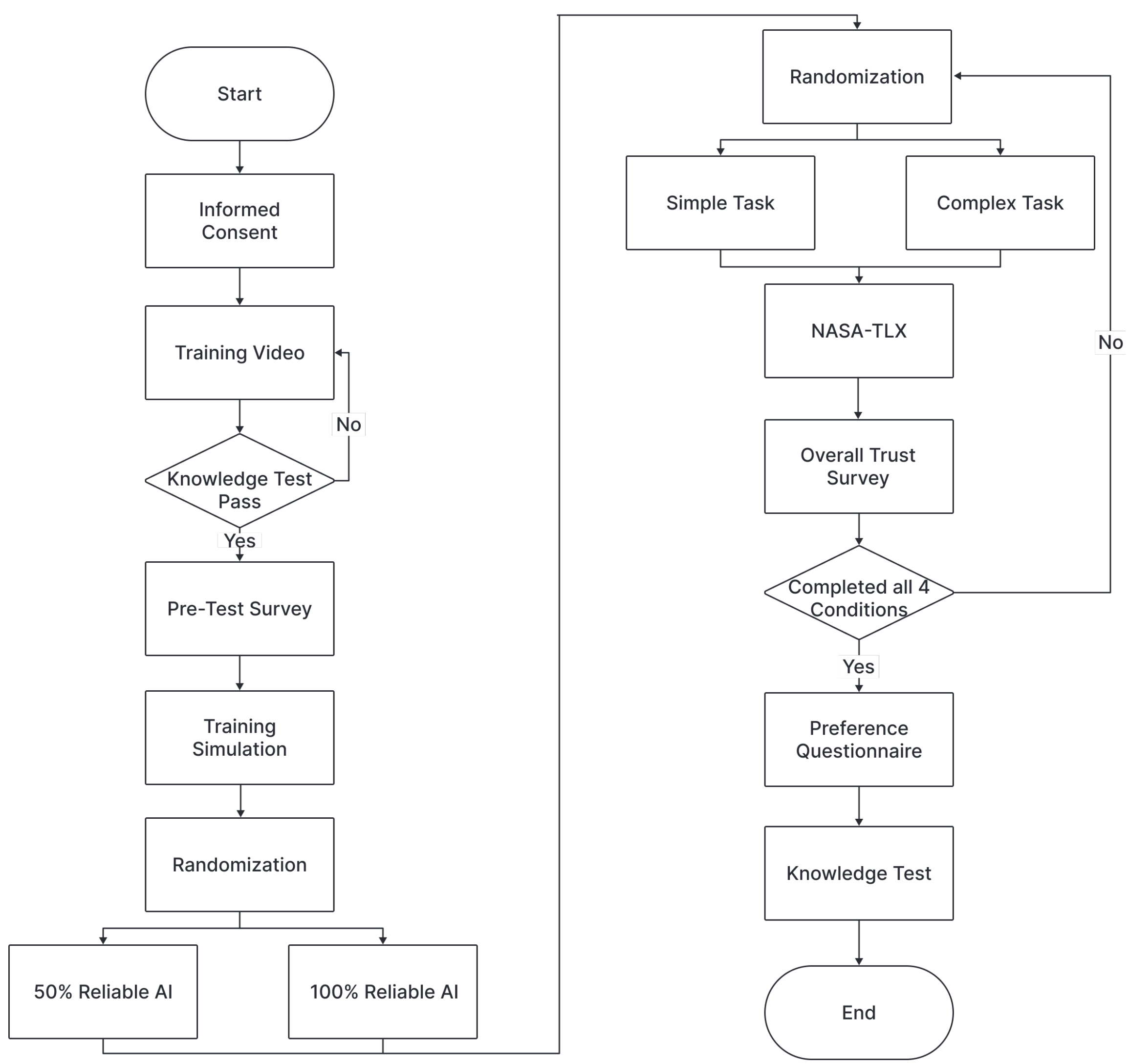
The participants were provided with the study material that consisted of the type of pipe fittings used in the study, what those components were used for, and the symbols of the components in a layout plan. Once they arrived for the study, they were seated and provided with the informed consent form to sign. Subsequently, they were asked to complete a pre-test questionnaire administered on Qualtrics on the computer. The pre-test consisted of a demographic questionnaire and trust questions to learn about the participants' baseline trust in technology (Schneider et al., 2017). After completing the pre-test questionnaire, participants were directed to view a training video. The video content reinforced the participants' understanding of heat pumps, including their constituent components, types of pipe fittings, and relevant plan symbols required to execute the task.

Furthermore, the video provided instructions on navigating the simulation, locating specific components, and how to install different components in the simulation. Then, they were asked to complete a short quiz that tested their knowledge of the symbols used for each component. This ensured that the participants were knowledgeable enough to read the layout plan provided. If the participants did not score at least 90% on the quiz, they were asked to watch the training video again until they met this criterion.

In the next step, the simulation was presented. The simulation was divided into two sessions: a training session and the actual task session. In the training session, the participants were asked to complete the task using a simple plan that involved only one heat pump along with two pipes and two different pipe fittings. In this session, there was no AI agent to assist them. This session was only intended to familiarize the participants with the simulation and the controls. Once they had completed the training session, they were taken to the actual task, where they were provided with a simple or complex task with an AI agent with low or high reliability, depending on the condition they were assigned to. The conditions were assigned to the participants based on a randomized Latin square design. The layout plan was provided to the participants before beginning the task. While completing the task, they could view the plan by clicking the “view map” button. The participants were asked to minimize the use of the plan and use the map only when they faced difficulties and were uncertain about how to proceed. This ensured their engagement with the recommender system rather than looking at the plan and completing the task. The simulation paused at every third step, and they were asked to rate their trust in the recommender system on a seven-point Likert scale. After each condition, they were asked to complete the NASA-TLX questionnaire, followed by the overall trust questionnaire. Once the participants finished the overall trust questionnaire for the fourth condition, they were asked to rank the four recommender systems based on their preference. This was followed by a knowledge quiz similar to the one they had taken in the initial part of the study. The entire experiment was carried out in a single session that lasted about 50-60 minutes. Figure 6 shows the flowchart of how the experiment was conducted.

Figure 6:

Flowchart of the Experimental Steps



Design

This study was conducted in a controlled space and used a 2*2 within-subjects design with each participant experiencing all four conditions.

Independent Variables.

Reliability of the AI Agent. This variable was studied at two levels: low and high. 50% of the AI agent's recommendations were correct for the low reliable condition, whereas, for the high reliable condition, 100% were correct. In both conditions, the participant was free to accept the agent's recommendation or move ahead with selecting components they thought were correct.

Complexity of the Task. This variable was also studied at two levels: simple and complex tasks. In the simple task, the participants were asked to complete a pipe layout that involved one heat pump and its associated fittings. For the complex task, four heat pumps of different capacities were provided, and they were asked to complete the layout connecting these heat pumps. This involved more pipe fittings and pipes in different directions, which made the task more complex compared to the simple one.

Dependent Variables.

Overall Trust. Trust is an essential factor to be considered when it comes to AI since it affects user acceptance, human-AI collaboration, and in decision-making (Hancock et al., 2011; Hoffman et al., 2018; Lee & See, 2004). This study used the most commonly used self-reported 12-item Likert scale to measure the trust in automation (Jian et al., 2000). This scale helps to understand the system characteristics that could influence the perception of the operators' trust. After each condition, the trust questionnaire was administered on Qualtrics to learn about the participant's subjective trust in the AI agent for that condition. After completing the questionnaire for the final condition, participants were asked an additional question: to rank the four AI agents in the order they would prefer to use them for completing a procedural task. This was done to understand which condition the participants favored most when using the AI agent.

Trust evolution. To investigate the evolution of trust during the interaction between a user and the recommender AI, a 7-point Likert scale question was utilized to measure the degree of trust at different stages of the interaction. The question was administered at regular intervals (every three steps) throughout the completion of the task in all four experimental conditions to capture changes in trust over time.

Task performance. This study evaluated task performance using completion time and the number of errors. The data were collected directly from the simulation since the time to complete the task and errors were tracked. The purpose of measuring these metrics was to investigate the effect of changing task complexity and AI agent reliability on task performance when assisted by an AI agent. If the participant selected a wrong pipe fitting or correct pipe fitting with a wrong size was counted as an error. The results of these measurements were used to determine if improvements in task performance were observed. The time required to complete a task in real-life settings may differ from that recorded in simulations. However, since all participants in this study were subjected to the same simulation, any differences in completion time can be used for performance comparison across varying conditions.

Workload. The varying reliability of the AI agent and the complexity of the task could affect the operator's workload while completing the procedural task. After every condition, the workload assessment was administered using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX)(Hart & Staveland, 1988).

Scenarios and Tasks. Heating ventilation and air conditioning (HVAC) installation is a common task in AEC, as it is an important component of building design and construction. The task selected for this study was the pipe fittings task associated with heat pumps in an HVAC installation. The task was designed in consultation with a subject matter expert and involved the completion of a pipe layout that included both the supply line and return line for a heat pump.

A heat pump is a device that uses electricity to transfer heat from a cooler place to a warmer place using a refrigerant. For this study, we selected a water source heat pump that uses water as the refrigerant to transfer heat from the indoor space to the cooling tower outside. A supply line provides cold water from the cooling tower to the heat pump, whereas a return line takes heat from the heat pump to the cooling tower.

The participants were provided with the plan of the pipe layout for each scenario. There were two different layouts: one for the simple task and another for the complex task. The simple task consisted of only one heat pump and the associated fittings, whereas the complex task contained four heat pumps of different capacities and associated fittings.

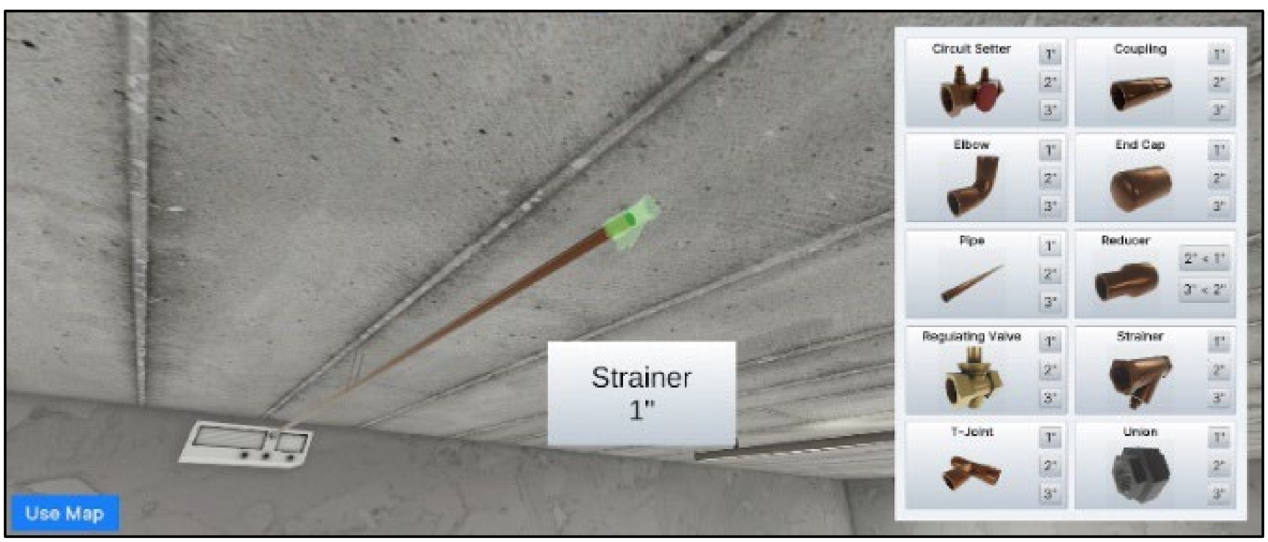
The task was completed on a computer with the help of an application that simulated a building where the HVAC installation was to be performed. The application simulated the environment such that the participant used AI-integrated AR technology while completing the task. The participants were asked to complete the pipe layout by deciding on the next component in relation to the current position and selecting the component from a library of fittings. An AI agent recommended the next fitting by visually showing it in a light green color and presenting its name and size. The participant was free to decide whether to install the fitting recommended by the AI agent or reject it and proceed with a different fitting.

Once the participant decided on the fitting, he could click on the fitting from the library of components, which will install the fitting at that location. This moves the point of view of the simulation to the following location where the component is required. This continued till the entire layout for that condition was completed. If, at any point in time, the participant was stuck and did not know how to move forward; they could click on the "use map" button on the left bottom part of the screen. This will open the layout plan, and they could determine which component goes at that point.

Figure 7 shows a screenshot from the simulation. The participants can select the component with a particular size from the library of pipe fittings available on the right side of the screen. The screenshot also shows how the recommender system offered suggestions. Here, AI recommended a pipe fitting which is a strainer of one inch. The recommendation is visually shown in light green color and by providing the name and the size of the fitting. On the bottom left of the screen, there is a "use map" button that shows the layout plan.

Figure 7:

Screenshot of the Simulation



Analysis. Data analysis was conducted with the help of statistical analysis software IBM SPSS Statistics for Windows, version 26.0.

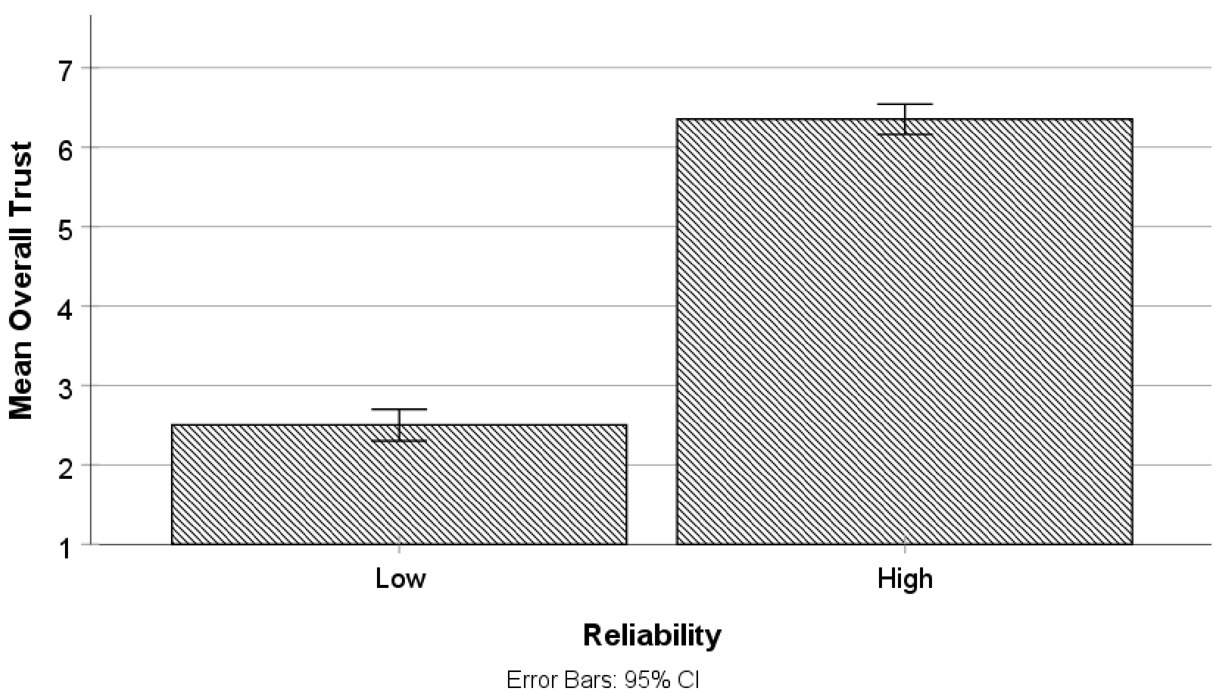
Results

Overall Trust

A two-way repeated measures ANOVA was run to determine the effect of the reliability of the AI agent and the complexity of the task on the overall trust in the AI agent. The two-way interaction between the complexity of the task and the reliability of the AI agent was not statistically significant for trust in AI, $F(1,34) = 1.98$, $p = 0.17$, partial $\eta^2 = 0.06$. However, the main effect of reliability was statistically significant, $F(1,34) = 492.20$, $p < 0.001$, partial $\eta^2 = 0.94$, while the main effect of complexity was not significant, $F(1,34) = 0.17$, $p = 0.69$, partial $\eta^2 = 0.01$. Subsequently, pairwise comparisons were conducted to examine the statistically significant main effect of reliability. Participants trusted the high reliable AI agent more than the low reliable one, with a mean difference of 3.92, 95% CI [3.56, 4.28], $p < 0.001$. Figure 8 shows the effect of the reliability of the AI agent on trust.

Figure 8:

Overall Trust for Two Levels of Reliability



Trust Development

Figures 9, and 10 were prepared to learn more about how trust was developed while completing the task. It was found that for the simple and complex tasks with high reliable agents, the participants started at a lower trust level and reached a consistent level in a few steps. The trend for the low reliable agent in simple and complex conditions was to stay at the trust level where they started. It was also found that for simple and complex tasks with a high reliable agent, the most number of people trusted the agent, with a rating of 7 in the early steps, whereas for simple and complex tasks with a low reliable AI agent, most of the participants started with a rating of 1,2, or 3 but the majority of the finished the task with either a rating of 1 or 2. A dependent t-test was conducted to understand if there was significant difference between the trust levels at the first step and the last step for each condition. For the simple task in the low reliable condition, it was found that there was a significant decrease in the trust level between the first step and the last step from 3.00 ± 1.80 to 2.08 ± 1.18 ($p = 0.007$); a decrease of 0.92 ± 1.90. For the simple task in the high reliable condition, it was found that there was a significant increase in the trust level between the first step and the last step from 6.03 ± 1.30 to 6.67 ± 0.68 ($p = 0.007$); an increase of 0.64 ± 1.33. In the case of complex task for the low reliable condition, it was found that there was there was no statistical difference between the trust levels of first and the last steps whereas for the high reliable condition, it was found that there was a significant increase in the trust level between the first step and the last step from 6.30 ± 1.30 to 6.89 ± 0.32 ($p < 0.001$); an increase of 0.69 ± 1.27.

Figure 9:

Trust Development of Simple Task

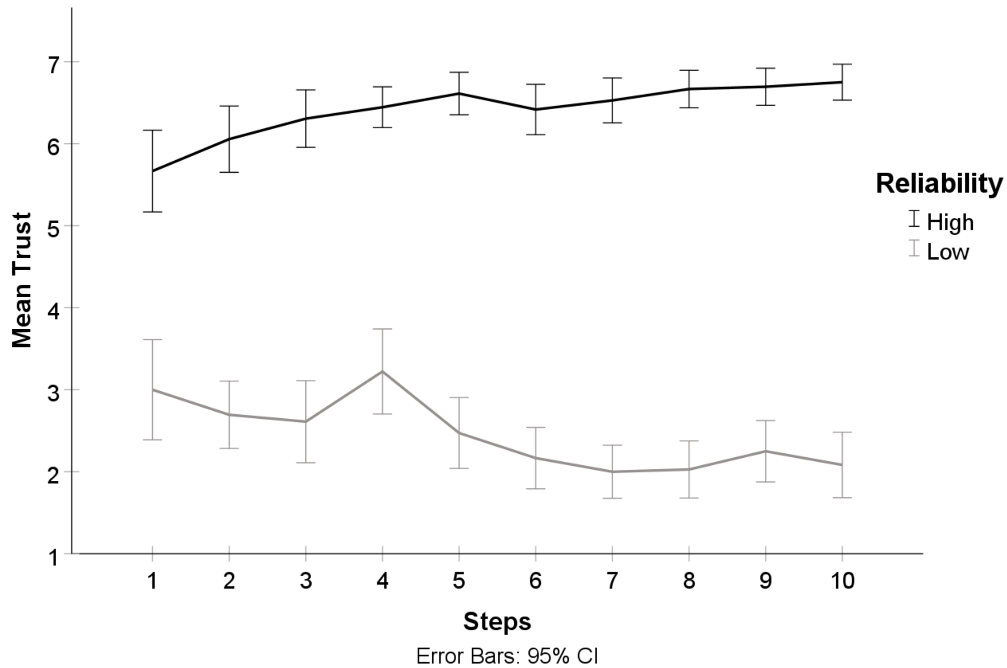
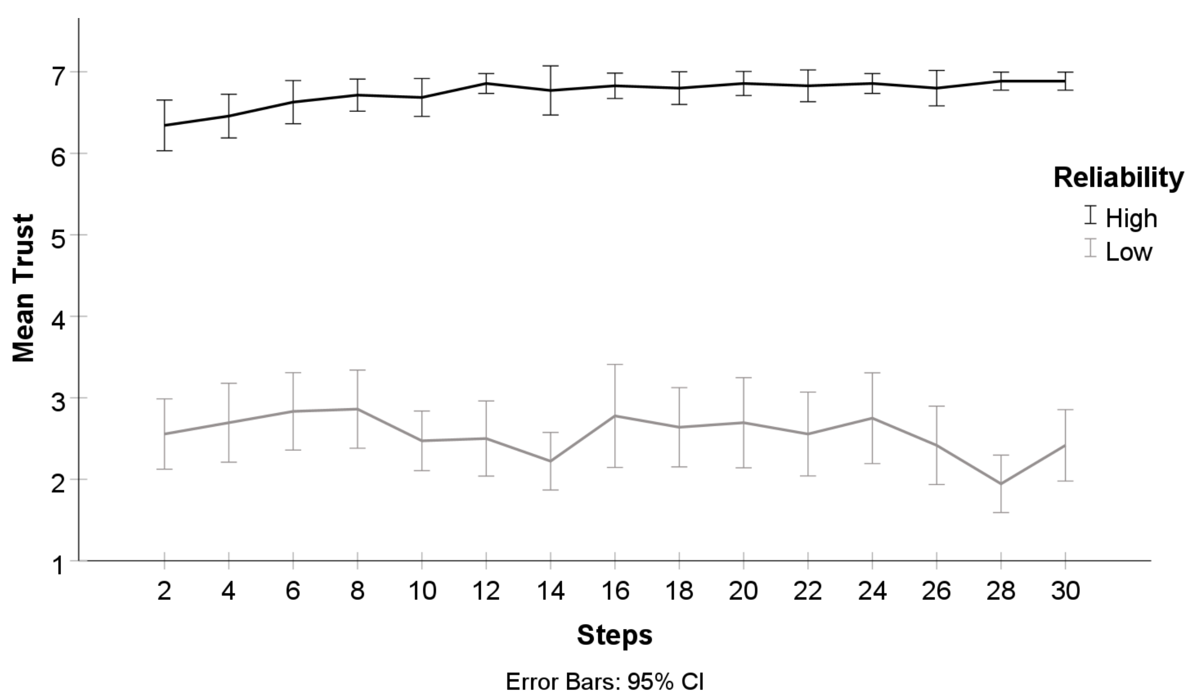


Figure 10:

Trust Development of Complex Task

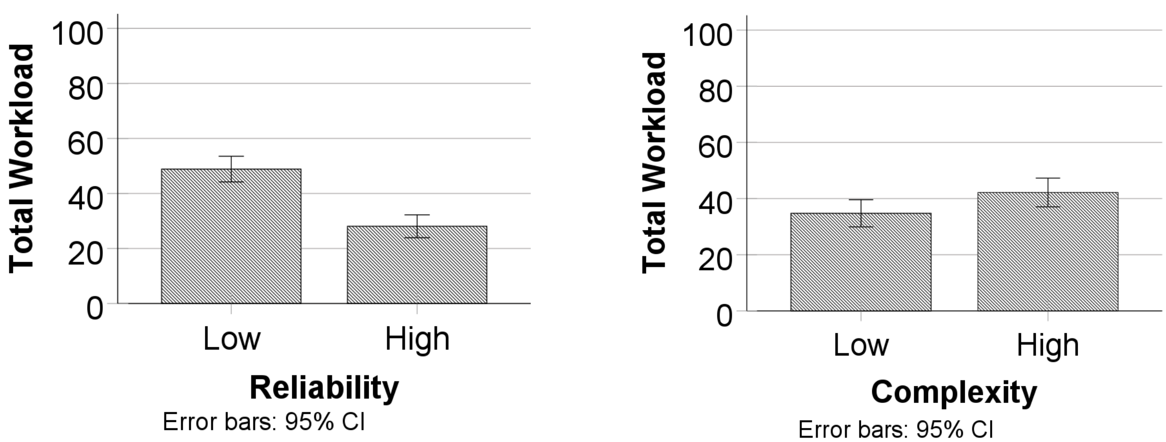


Workload

A two-way repeated measures ANOVA was run to determine the effect of the complexity of the task and the reliability of the AI agent on the users' workload. There was no statistically significant interaction between the complexity of the task and the reliability of the AI agent, $F(1, 35) = 1.73$, $p = 0.20$, partial $\eta^2 = 0.05$. However, the main effect of the reliability was statistically significant, $F(1,35) = 69.85$, $p < 0.001$, partial $\eta^2 = 0.67$, as was the main effect of the complexity, $F(1,35) = 14.21$, $p = 0.001$, partial $\eta^2 = 0.29$. Subsequently, pairwise comparisons were conducted to examine the statistically significant main effect of reliability and complexity. Participants perceived a higher workload when the AI agent exhibited low reliability than high reliability, with a mean difference of 20.78, 95% CI [15.73, 25.83], $p < 0.001$. Additionally, participants perceived the workload was higher in the high complexity condition than in the low, with a mean difference of 7.41, 95% CI [3.42, 11.40], $p < 0.001$.

Figure 11:

Workload for Two Levels of Reliability and Complexity

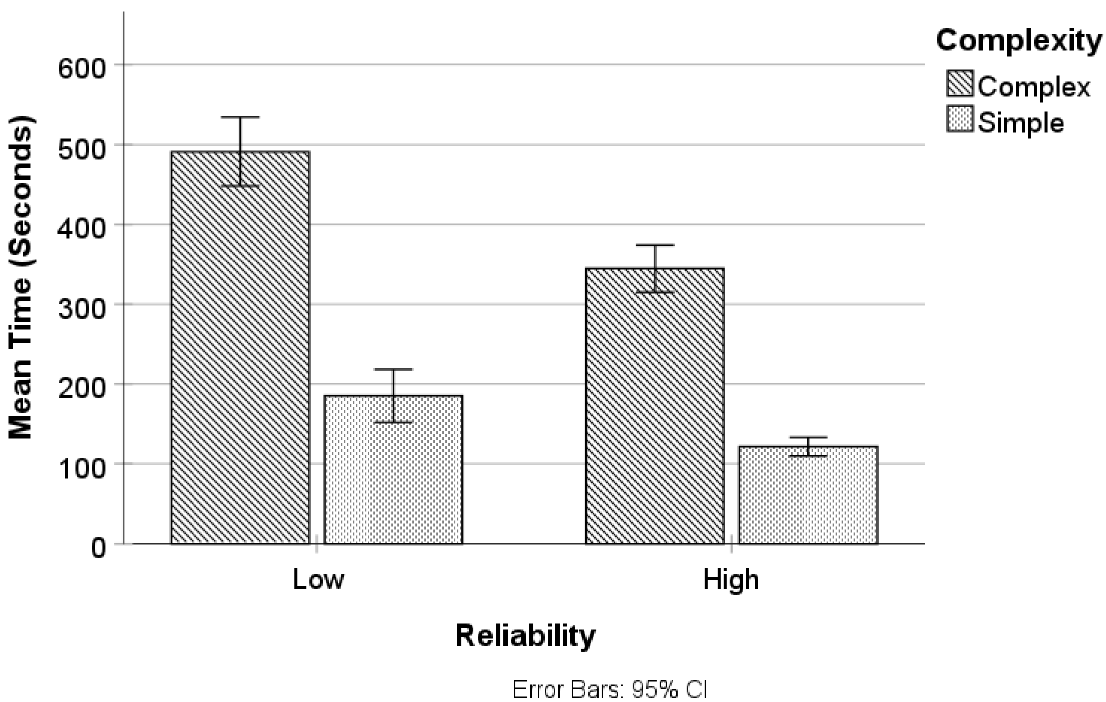


Performance

Task Completion Time. A two-way repeated measures ANOVA was conducted to determine the effect of the complexity of the task and the reliability of the AI agent on the time taken to complete the task. The two-way interaction between the complexity of the task and the reliability of the AI agent was statistically significant for the time taken to complete the task, $F(1,34) = 8.93$, $p = 0.005$, partial $\eta^2 = 0.21$. Therefore, simple main effects were run. The total time taken to complete was statistically significantly different in the high reliable condition (334.46 ± 64.16 seconds) compared to the low reliable condition (483.78 ± 129.59 seconds) in the complex task, $F(1,34) = 42.73$, $p <0.001$, partial $\eta^2 = 0.56$, a mean difference of 149.32 (95% CI, 102.90 to 195.75) seconds. The total time taken was also significantly different in the high reliable condition (119.16 ± 30.88 seconds) compared to the low reliable condition (174.44 ± 74.87 seconds) in the simple task, $F(1,34) = 16.98$, $p < 0.001$, partial $\eta^2 = 0.33$, a mean difference of 55.27 (95% CI, 28.01 to 82.53) seconds.

Figure 12:

Time Taken to Complete the Task



Number of Errors. A negative binomial regression analysis was conducted to examine the effects of the complexity of the tasks and the reliability of the AI agents on the number of errors made. The traditional Poisson regression was not applied because the error data was over-dispersed since the variance exceeded the mean. The results showed that the reliability of the AI agent had a significant effect on the number of errors made, such that participants who interacted with the low reliable agents had significantly higher error rates compared to those who interacted with high reliable agents ($B = -1.73$, $SE = 0.18$, 95% CI [-2.09, -1.38], Wald $\chi^2 = 90.57$, $df = 1$. $p < 0.001$). The results also showed that the complexity of the task had a significant effect on the number of errors made, such that the participants made more errors in the complex task than in the simple task ($B = 0.71$, $SE = 0.17$, 95% CI [0.37, 1.05], Wald $\chi^2 = 16.66$, $df = 1$. $p < 0.001$).

Preference of agents

A Friedman's test was conducted to analyze the ranked data for the preference of agents. There was a statistically significant difference in the ranking of agents which recommended the pipe fittings, $\chi^2(3) = 88.70$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction, resulting in a significance level set at $p < 0.008$ by dividing the significance level 0.05 by the number of comparisons, which is 6. Median inter quartile range ranking for the high reliable agent in the complex condition, high reliable agent in the simple condition, low reliable agent in the simple condition, and low reliable agent in the complex conditions were 1 (1 to 1), 2 (2 to 2), 3 (3 to 4), and 4 (3 to 4), respectively. There was a statistically significant difference in ranking where the participants preferred the simple condition with the high reliable agent over the complex condition with the low reliable agent ($Z = -5.40$, $p < 0.001$), the simple condition with the high reliable agent over the simple condition with a low reliable agent ($Z = -5.37$, $p < 0.001$), the complex condition with the high reliable agent over the simple condition with the high reliable agent ($Z = -3.36$, $p = 0.001$), the complex condition with the high reliable agent over the complex condition with the low reliable agent ($Z = -5.36$, $p < 0.001$), and the complex condition with the high reliable agent over the simple condition with the low reliable agent ($Z = -5.36$, $p < 0.001$). However, there was no significant difference in ranking between preferring the simple condition with a low reliable agent vs. the complex conditions with a low reliable agent ($Z = -1.76$, $p = 0.078$).

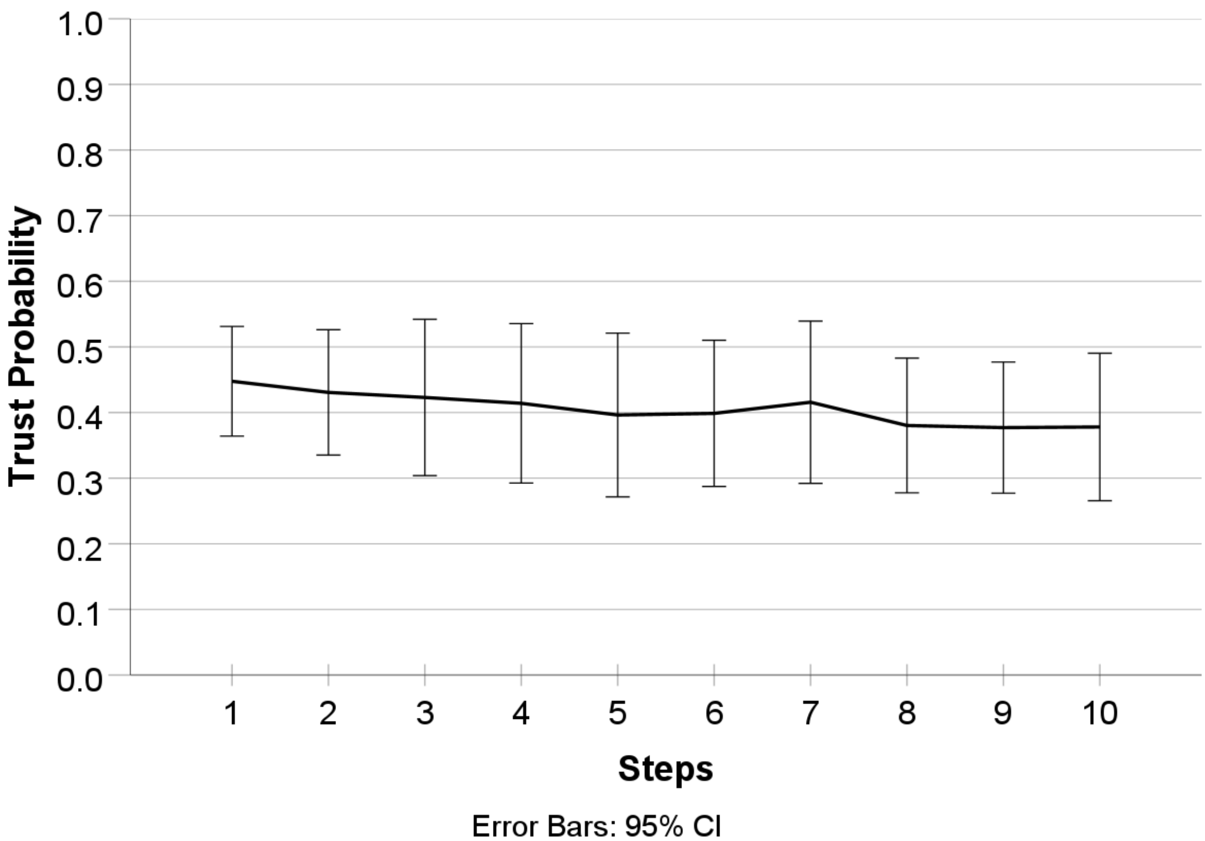
Trust in AI evolution using the POMDP model

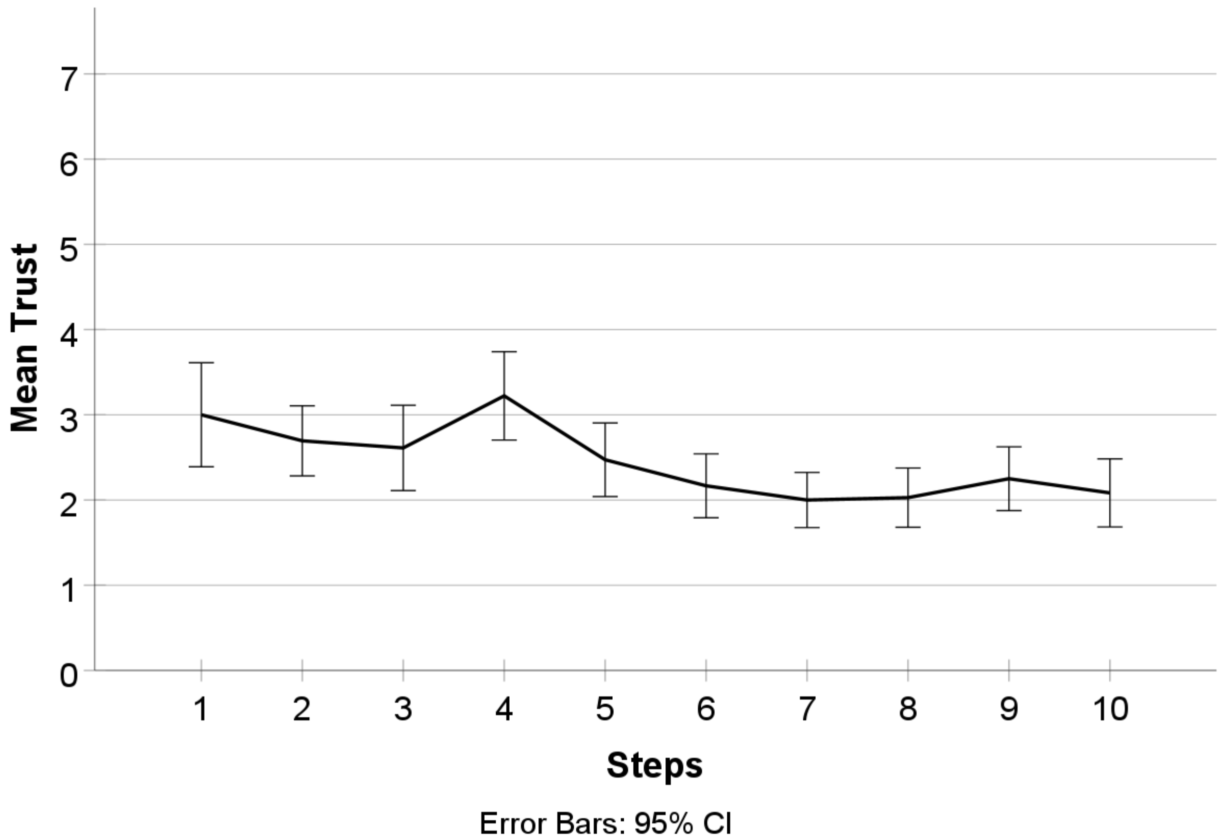
This section presents the results obtained from the proposed POMDP model for 10 iterations. These iterations converge the trust values and show the standard error for each step. In this work, the POMDP model was developed to model human trust while the AI assists the human in completing the pipe fitting task by providing recommendations. The human trust development was analyzed in scenarios with both low reliable and high reliable agents for simple and complex tasks. The trust dynamics were observed over 10 iterations, and the results are depicted in Figures 13 and 14 for the simple task and Figures 15 and 16 for the complex task. The trust development was measured using the state space $S = [0,1]$, which is the probability that represents all possible values of human's trust in the agent. The results were then compared to the experimental study to validate the model's effectiveness in simulating human trust during the interaction.

The initial trust level is affected by the shape parameters of the initial trust probability distribution. In the low reliable condition, the agents recommended the wrong components frequently, and this caused the trust level to decrease. As the steps progressed, the trust level dropped significantly due to the agent's frequent incorrect recommendations, leading to increased construction damages and increased time costs. Thus, a gradual decline of trust over time was noticed, aligning with the trust development trend reported by the participants during the experimental study. This similarity in trends validates the effectiveness of the model for representing human's trust when interacting with a reliable agent. The trust dynamics model, which is governed by the beta distribution, affected the actions of humans as they were less likely to follow the agent's recommendation as trust declined. This is because the recommender AI's belief of human's trust at each time step was updated based on the feedback that it received, which is human's action. Similarly, the human's trust was updated based on the performance of the recommender system.

Figure 13:

Trust Development for Simple Task with Low Reliable Agent- POMDP Model (Top) and Experimental Study (Bottom)

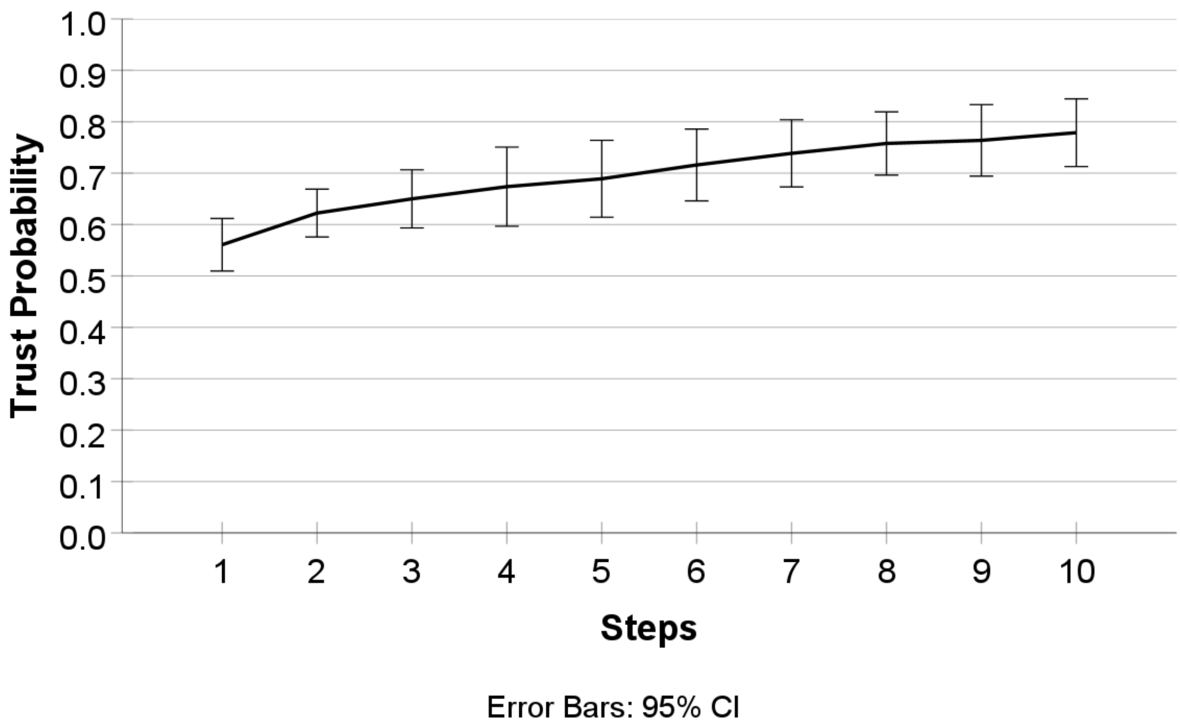


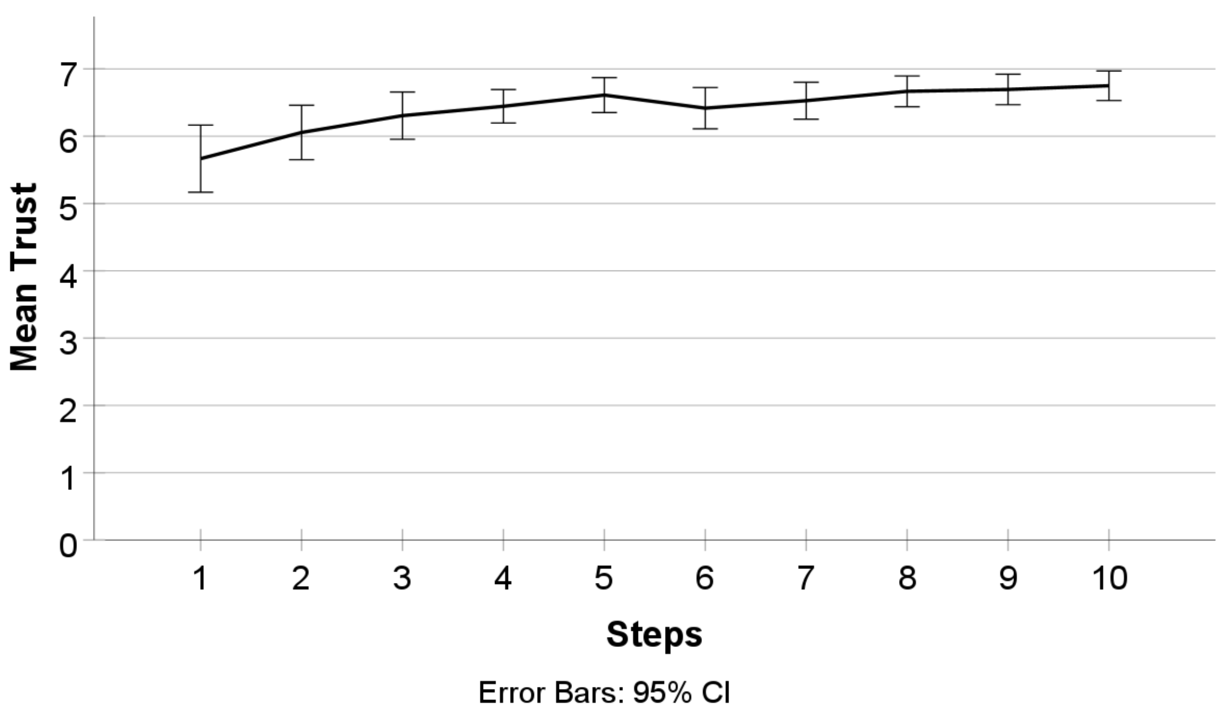


For the high reliable condition in the simple task, the AI agent provided the correct recommendation for all the steps. The consistent performance of the agent resulted in impacting the human's trust positively. The trust dynamics model, in this case, strengthened the positive trust trend due to positive experience gains. The increasing trend in trust for high reliable agents aligns well with the experimental study, demonstrating the model's validity in representing human trust development in scenarios with high reliable agents. The trust-behavior model reflected this trend, with humans becoming more inclined to follow the agent's recommendations as trust grew, leading to minimized. The trust-behavior model reflected this trend, with humans becoming more inclined to follow the agent's recommendations as trust grew, leading to minimized construction damages and time costs.

Figure 14:

Trust Development for Simple Task with High Reliable Agent- POMDP Model (Top) and Experimental Study (Bottom)

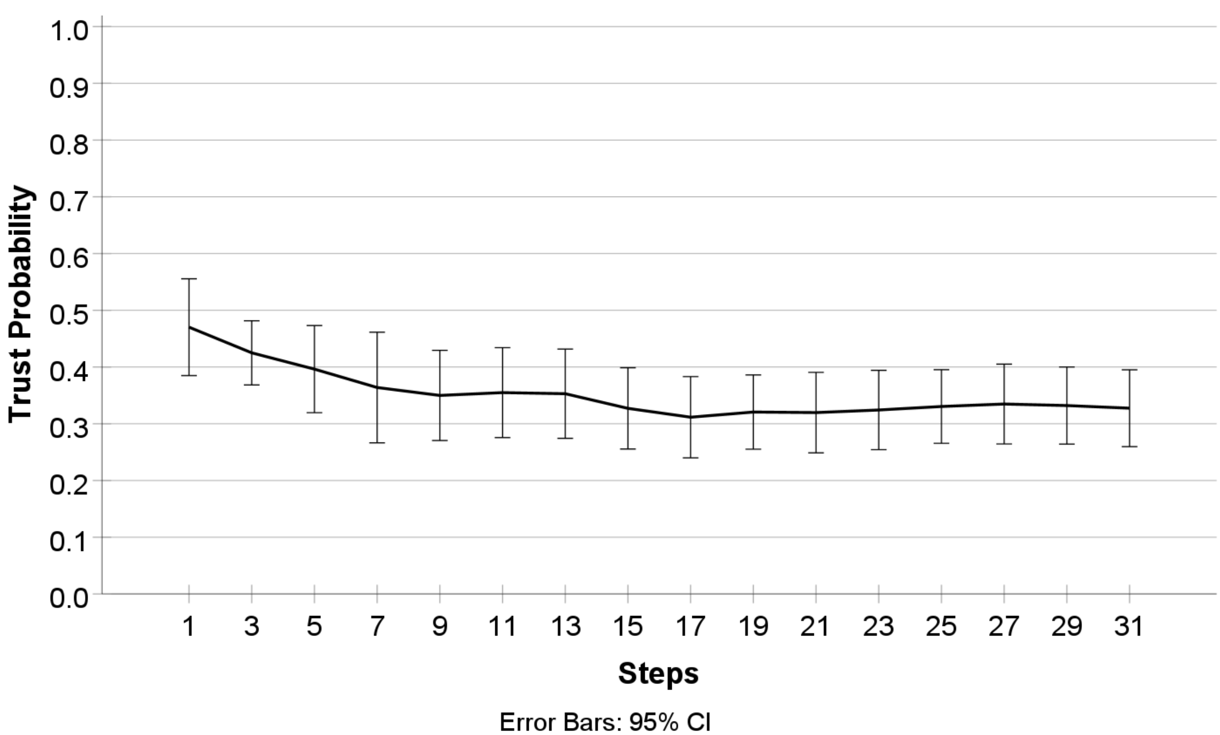


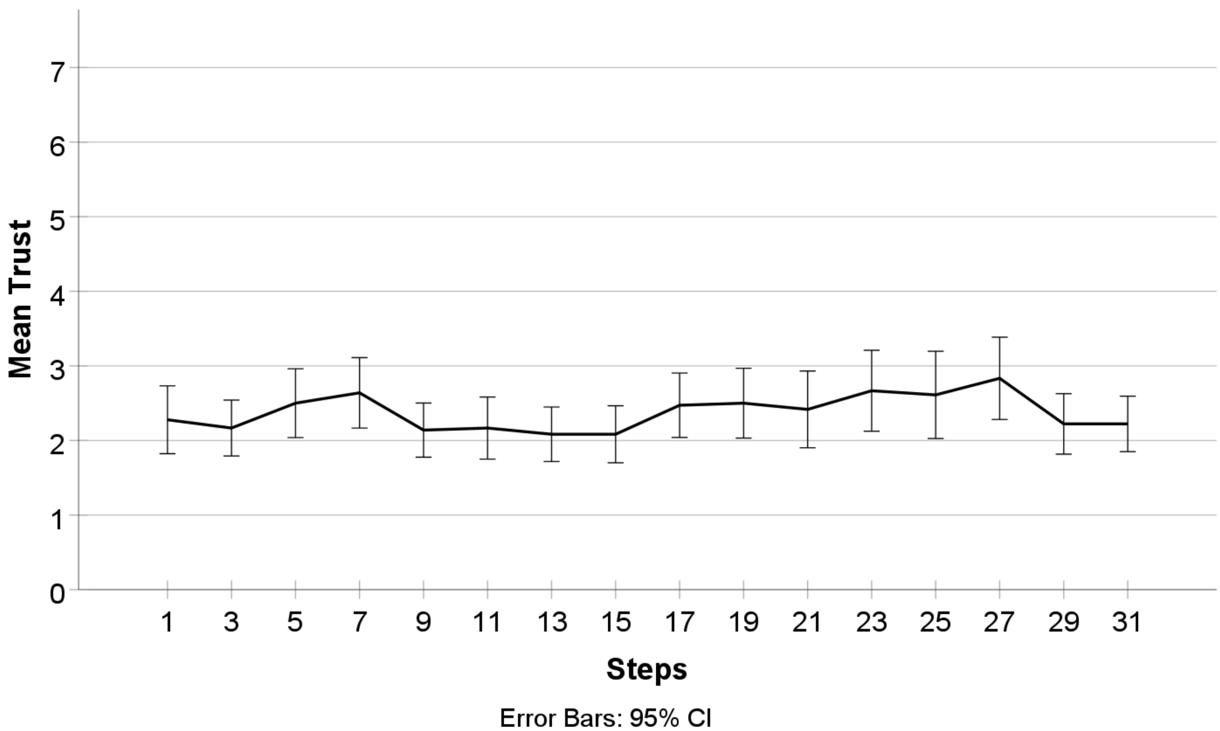


In the complex task, for the low reliable condition, the initial trust level, as in the simpler task, was influenced by the shape parameters of the initial trust probability distribution. However, trust development exhibited a different trend in this scenario. While the agent frequently recommended incorrect components, the trust level did not consistently decline, as observed in the simpler task. Instead, after approximately half of the interaction steps, the trust level seemed to stabilize. This stabilization, however, had a few fluctuations that are associated with the complexity of the task. These fluctuations in trust levels align closely with the experimental results, further validating the model's capability to represent human trust dynamics in complex scenarios. The trust dynamics model, governed by the beta distribution, played an important role in these fluctuations.

Figure 15:

Trust Development for Complex Task with Low Reliable Agent- POMDP Model (Top) and Experimental Study (Bottom)

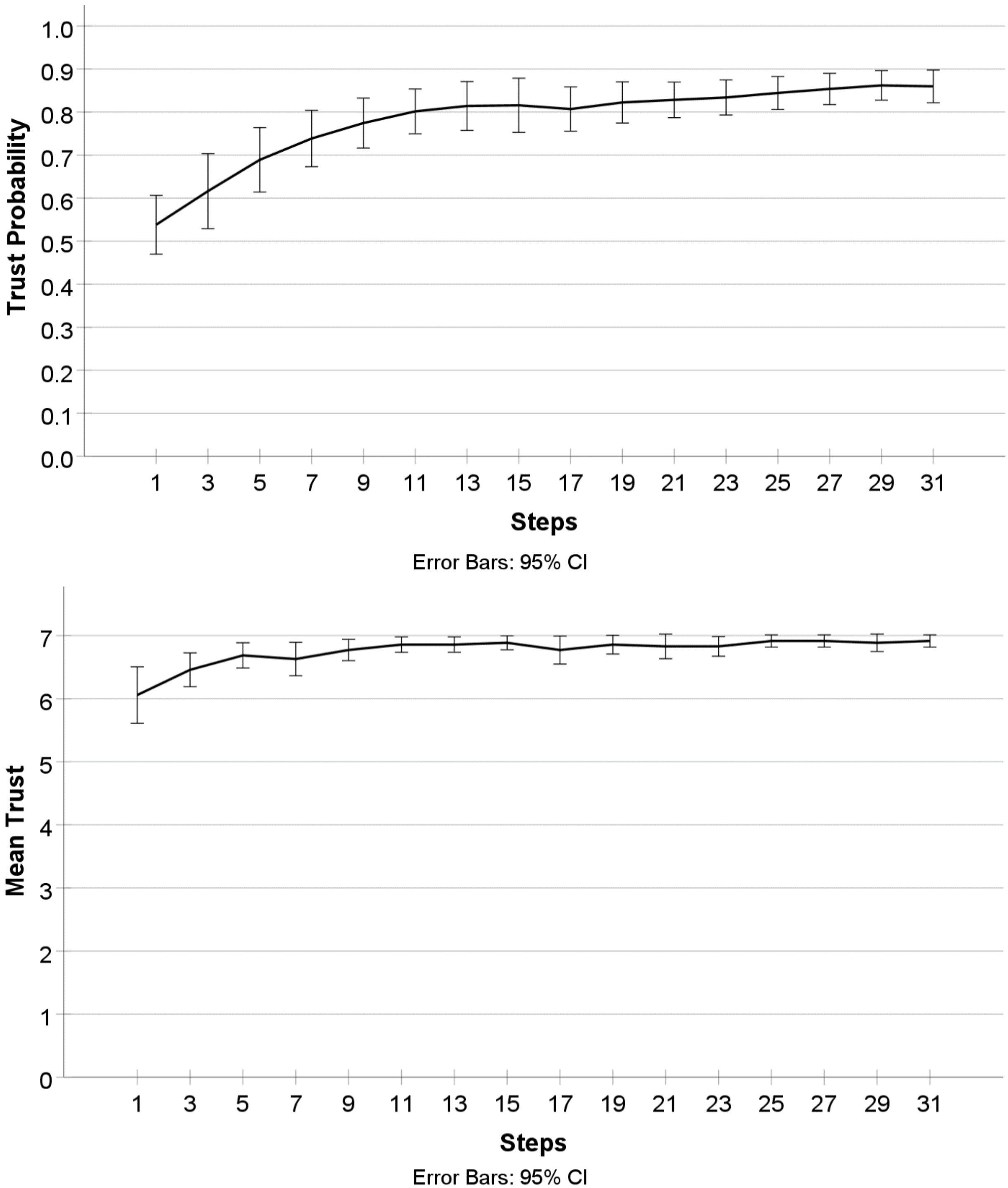




In the high reliable condition, the AI agent provided correct recommendations throughout the task. This consistent performance had an evident impact on trust development. The initial stages of the interaction saw a rapid increase in trust levels, reflecting the agent's reliability. However, as the interactions progressed, particularly after reaching approximately half of the steps, the trust increase rate began to stabilize or even reduce. This could be because the trust increased to almost the highest level. This stabilization in trust levels aligns with the findings from the experimental study.

Figure 16:

Trust Development for Complex Task with High Reliable Agent- POMDP Model (Top) and Experimental Study (Bottom)



Comparing the POMDP model and the experimental study suggests that the POMDP model can be used to represent the impact of the reliability of the recommender AI and the complexity of the tasks on human trust. The trust dynamics model, incorporating the beta distribution and the trust-behavior model relating human's actions to trust, played a crucial role in achieving results consistent with the experimental study. Considering construction damage loss and time cost, the immediate rewards effectively influenced human actions, reflecting realistic human behavior in response to the agent's performance.

Discussion

Trust in AI agents is an important factor affecting the user relying on the technology to complete the task. This study examined the effect of the AI agent's reliability and the task's complexity while completing a procedural task with the help of augmented reality technology. The results from the two-way repeated measures ANOVA suggest that for trusting the recommendations, the reliability of the AI agent is a more important factor than the task's complexity while designing systems that utilize AI systems and AR technology. There was no significant effect on trust due to the complexity of the task, whereas the reliability of the AI agent had a substantial influence on the perceived trust. This finding is consistent with the previous research, which has shown that reliability is an important system property that affects trust in automation (Parasuraman & Manzey, 2010). Participants trusted the high reliable condition more than the low reliable condition, which supports hypothesis 1 and is consistent with the findings from the previous research that found that the more errors the automation made, the lower the trust they had in automation for a route planning task (de Vries et al., 2003). This shows the importance of designing AI systems with high reliability for completing procedural tasks for the users to trust the AI agent. This study reinforces the concept that the relationship between AI agent reliability and operator trust, previously established in different domains, is also pertinent to scenarios where operators are engaged in procedural tasks supported by a recommender system within the framework of augmented reality technology. Future research could investigate ways to maintain appropriate trust using interventions such as explainability, which provides explanations for the recommendations or actions, and transparency, which shows the real-time status of what the system is doing, for varying AI reliability and complexity of the procedural tasks with the help of an augmented reality technology.

The complexity of the task did not have a significant main effect on trust. This shows that for the procedural task selected for this study, trust was not impacted by the complexity of the task alone. It means that the users might not trust the AI agent while completing a complex task since it could lead to many more complex errors if a mistake is made. Or in other words, for the user to use the AI to complete a complex task, they should have a good reason to have high trust in the AI (Al, 2023). But, from the results, it seems that the participants lost trust in a similar manner between the simple and complex tasks when the AI agent made errors. Future research could explore different task types to learn the effect of task complexity on perceived trust. There was no significant interaction between the reliability of the AI agent and the complexity of the task, which implies that the different levels of task complexity did not moderate the effect of the reliability of the AI agent on trust. Trust in AI was primarily determined by the reliability of the AI agent, regardless of whether the task was simple or complex. This lack of significant difference in trust between the complexity of the task and the lack of significant interaction could be specific to the task selected, and future research could look more into other domains for procedural tasks to confirm these findings.

The trust evolution while completing the task was investigated using a t-test between the trust levels at the first and the last step of each condition. The first trust questionnaire was asked after completing three steps of the task. This way, the participants had a chance to experience the reliability of the AI agent before the first questionnaire. For the simple and complex tasks with a high reliable agent, they started at a lower trust level than when they finished the whole task. From the t-test, we found a significant increase in trust level between the first and the last steps for both simple and complex tasks with high reliable agents. This shows the positive trend in trust as they interact with the recommender AI and the importance of high reliability on the trust not only for building it but also for maintaining the high trust level. The rapid development of trust in these scenarios, as seen in Figures 9 and 10, reflects the users' quick adaptation and growing confidence in the agent's capabilities. For the low reliable agents, in the simple condition, the t-test found that there was a significant decrease in trust levels between the first and the last steps. The participants started at a low trust level and ended at much lower levels of trust. This indicates that the participant had limited trust in the AI agent from the beginning, and their trust continued to diminish as they interacted with the agent over time. The decreasing trend of the trust rating could be because of the loss of confidence as they interacted more with the agent, suggesting that the trust was negatively impacted. This also shows that the users had minimal tolerance for errors when the recommender AI assisted the users with simple tasks. For the complex task with low reliable agents, there was no significant difference in trust level between the first and the last steps. This could be because since the participants knew they were in a complex condition, they started by not trusting the agent for a complex task and looking for a good reason to increase their trust. But, after interacting with the agent, the trust remained at the same level as there were errors in the recommendations provided.

The perceived workload while completing the procedural task was measured using the NASA-TLX survey. The results indicate that there was no significant interaction between the complexity of the task and the reliability of the AI agent. This means that these factors had an independent effect on the perceived workload for the context of this study. But there were significant main effects on workload by both the factors. The workload was higher for the complex task compared to the simple task, a result supporting previous research findings that complex task conditions resulted in a higher overall subjective workload (De Visser & Parasuraman, 2011). This result confirms that the tasks selected as simple and complex were of different complexity levels, as expected. The complex task could impose a higher cognitive load since it requires higher attention, information processing, and mental resource allocation, resulting in a higher perceived workload.

The two levels of reliability of the AI agent had a significant main effect on the perceived workload. This means that the reliability of AI agents can significantly affect users' workload while using AR technology to complete a procedural task. While the reliability of the AI agent was low, the perceived workload was high, supporting hypothesis three. This finding aligns with previous research that found that while interacting with a low-reliable agent, the users' cognitive load was higher (Daronnat et al., 2020). While the reliability of the AI agent was low, the participants had to find the solutions themselves rather than relying on the AI agent, which could have increased their cognitive load and decreased trust. They had to decide on the next component based on their working or long-term memory or could have clicked on the view map button to see the layout when stuck. Thus, they might not have relied on the AI agent for recommendations, increasing their perceived workload.

To evaluate the participants' performance while completing the tasks in different conditions, the total time taken to complete the task was measured. The two-way repeated measures ANOVA results suggest a significant difference in performance between the simple and complex conditions and between conditions with low and high reliable agents supporting hypothesis two. There was a significant difference in the time taken between the simple and complex conditions since the complex task had more steps and pipe fittings than the simple task. The interaction was also significant, indicating that the effect of the reliability of the AI agent on time taken depended on the complexity of the task. The difference in time taken to complete the task between the low and high reliable agents were more for the complex task indicating the importance of high reliable agents for complex task or tasks with higher cognitive workload. This decrease in time could be because the low reliable AI agent imposed a higher cognitive workload on the operator, and the higher workload deteriorates human performance (Xie & Salvendy, 2010). Complex tasks require more cognitive workload and decision-making capabilities. When the AI agent is reliable, they trust the recommendations and rely more on the agent, which could reduce their workload, leading to higher performance. However, for the low reliable agent, the participants did not trust the recommendations. They had to figure out the solutions themselves, leading to a higher workload and thus increasing the time taken to complete the task. This finding of higher task time completion for low reliable agents aligns with previous research that compared three levels of reliability and found that, as reliability increases, the task time decreases (Chavaillaz et al., 2016).

The participant's performance in completing the task was also measured using the number of errors they made while completing the task. A negative binomial regression analysis was conducted to gain some insights into the relationship between the complexity of the task and the reliability of the AI agent on the number of errors made. Consistent with our hypothesis two, the number of errors was significantly affected by the reliability of the recommender system. Specifically, the participants who interacted with the low reliable AI agent had significantly higher error rates compared to those who interacted with a high reliable agent. This aligns with previous research that found that as the reliability of the AI agents decreases, the participants make more errors while completing a visual search task (Shah & Bliss, 2017). While Shah & Bliss, (2017) focused on visual search tasks, the underlying principle of their findings—that decreased reliability in AI agents correlates with increased user errors—may extend to procedural tasks, as both require the careful execution of steps influenced by the agent's guidance, suggesting a potentially generalizable impact of AI reliability on task performance across different domains. This shows the importance of having high reliable recommender system to improve performance and reduce errors while completing procedural tasks. Additionally, the complexity of the task had a significant influence on the number of errors made while completing the task. While in the complex task, the participants made more errors than in the simple task. This could be because as the complexity of the task increases, the workload increases, resulting in low performance (Xie & Salvendy, 2010).

After completing all four conditions, the participants were asked to rank the AI agents that assisted them in completing the task. Our primary finding indicates a significant statistical differentiation in the ranking of agents considering the AI agent's reliability and the task's complexity. As expected, the participants preferred the high reliable agents for both simple and complex tasks. This is because the high reliable agent provided more correct recommendations compared to the low reliable agent and was more trustworthy. One important finding is that the participant preferred the high reliable agent for the complex task over the simple one. This shows the importance of high reliable agents for more complex tasks that require higher cognitive effort, especially when completing a procedural task. This shows that the participants had higher trust in the highly reliable agent and might have relied more on the agent. The participants might have trusted the high reliable agent in the simple task as well, but it was more important in the complex task since it might have reduced the extra workload the participants had in the complex condition. Another important finding is that the participants had no significant preference difference in selection between a simple and a complex task with a low reliable agent. This is an important insight as it indicates that when the reliability of the AI agent is low, the complexity of the task does not influence user preference. This again shows the importance of the reliability of the AI agent while providing recommendations for completing a procedural task.

The results from the POMDP model provide a good understanding of human trust development in the AI during the interaction, and it aligns closely with the trust development results from the experimental study. This validates the model and shows the effectiveness of the model in simulating real-world interactions. The model highlights the importance of reliability in establishing and maintaining trust in agents. It implies that prioritizing high reliability and the ability to make accurate recommendations is essential in an AI agent's design. The understanding gained from the trust dynamics can guide the creation of adaptive AI agents. Such systems, with the ability to change their behavior in response to changing trust levels, can promote smoother and more productive interactions in the fields of AEC, healthcare, and manufacturing. The model's focus on immediate rewards and the subsequent influence of trust on decision-making highlights the necessity to formulate strategies that can reduce potential damages and improve time efficiency while completing a procedural task with the assistance of AI agents.

The model comes with its own limitations as assumptions were made during its development, such as the initial trust probability distribution and the trust behavior model. The model also needs to be validated for other contexts and domains in order to generalize it. Future studies could investigate other factors that affect trust, such as environmental factors or other human or AI-related factors, for example, the self-confidence of the operator and the explainability and transparency aspects of the recommender system.

Conclusion

A POMDP model was developed to simulate human's trust in AI while completing a procedural task with the help of recommendations from the agent. An experimental study was conducted that investigated the impact of task complexity and AI agent reliability on trust, performance, workload, and user preferences in completing procedural tasks using AR technology. The trust development simulated from the model and that from the experimental study were compared to understand the effectiveness of the model. The findings offer valuable insights that can inform the design and implementation of AI systems and AR technology for procedural task completion.

The results highlight that the reliability of the AI agent significantly influences trust, with participants exhibiting higher trust in highly reliable AI agents compared to low reliable ones. Task complexity, conversely, does not independently affect trust, suggesting that reliability plays a more important role in determining users' trust levels. Analysis of trust development throughout the task completion process revealed interesting patterns. Participants using highly reliable AI agents experienced a gradual increase in trust, reaching a consistent level by the task's end. In contrast, those using low reliable AI agents maintained low levels of trust throughout the task. These findings highlight the importance of reliability in trust development in AI agents.

Workload analysis demonstrated that both task complexity and AI agent reliability significantly impact perceived workload. Complex tasks resulted in a higher perceived workload due to increased cognitive demands, while low reliable AI agents imposed a greater cognitive workload on participants. Designing AI systems with high reliability is crucial for reducing users' cognitive load and enhancing performance.

Both task complexity and AI agent reliability influenced performance measures, including task completion time and the number of errors. Complex tasks took more time to complete, and participants using low reliable AI agents also took more time to complete the task. Similarly, errors were higher in complex tasks and when interacting with low reliable AI agents. These findings highlight the detrimental effects of low reliability on task efficiency and emphasize the importance of AI agent reliability in minimizing errors. User preferences consistently favored high reliable AI agents, indicating their effectiveness in handling higher cognitive demands. Participants preferred high reliable agents for both simple and complex tasks, underscoring the significance of reliability in user acceptance and satisfaction.

The POMDP model results of the human's trust development was compared to the experimental results of trust development and found that both trust development trends were similar. This shows the effectiveness of the model in simulating human's trust while completing the procedural task with the assistance of an AI agent.

Overall, this study provides insights into the impact of task complexity and AI agent reliability on trust, performance, workload, and user preferences during procedural task completion with the assistance of AI agents and AR technology. These insights can guide the design and implementation of AI systems and AR technology, aiming to maximize user trust, enhance task performance, and reduce workload. Future research could explore additional factors and task types to further refine the understanding of the relationship between task complexity, AI agent reliability, and trust in the context of procedural task completion with the help of AR and AI technologies.

After investigating the effect of reliability and complexity of task on humans' trust in the recommender AI while completing a procedural task with the assistance of AR technology, it is evident that the reliability of the recommender AI is an important aspect to consider in such scenarios. Understanding the factors that could help maintain appropriate trust while interacting with recommender AI in such scenarios is important. Accordingly, the second study reported in Chapter 4 aims to investigate the effects of explainability and transparency on human's perceived trust levels when interacting with recommender AIs with varying reliability.

CHAPTER FOUR

INVESTIGATING THE EFFECTS OF ARTIFICIAL INTELLIGENT AGENTS' TRANSPARENCY AND EXPLAINABILITY ON USERS' TRUST FOR COMPLETING A PROCEDURAL TASK

Introduction

With the adoption of Construction 4.0, the AEC sector has seen significant transformations through technological advancements that reshape decision-making processes, enhance operational efficiency and pave the way for innovative construction methodologies (Shafei et al., 2022). Given the complexity of the AEC domain and the benefits of AI, such as efficiency and reduced errors, AI has huge potential if implemented correctly. AI can automate complex tasks, assist in decision-making, and improve productivity and efficiency in the AEC domain. Many complex problems in the AEC domain have been solved efficiently by AI, such as pavement crack detection, seismic safety evaluation, damage to concrete structures under complex constraints etc. (Lagaros & Plevris, 2022).

Integrating AI with AR technology could benefit in many ways. For example, while completing a physical task, AR can reduce the cognitive load of the users by displaying the required information without affecting the user's focus (Sahu et al., 2021). The combination of AR and AI technologies could assist humans considering the benefits of both technologies, which have already been practiced in domains such as manufacturing and healthcare (Chen et al., 2019; Sahu et al., 2021). For example, an AR- and AI-assisted surgical navigation system was developed and tested for accuracy and feasibility (Siemionow et al., 2020). The AEC domain could also benefit from this combination in addressing the challenges caused by the tasks in complex and dynamic contexts. One example could be the use of these two technologies to enhance the safety of highway workers by providing them with notifications based on real-time danger predictions (Sabeti et al., 2021).

Trust the users have in the AI system affects the adoption and acceptance of the AI technology (Asan et al., 2020; Chin et al., 2024). Trust in AI is a complex multidimensional construct that is affected by various factors, including AI reliability, explainability, and transparency of the system (Bedué & Fritzsche, 2022; Hoff & Bashir, 2015; Ryan, 2020; Schraagen et al., 2021). Perceived reliability is an important aspect in understanding the issues associated with human performance due to its relationship with the trust users have in AI. As the perceived reliability increases, the trust should also increase as the users have confidence in AI. However, the calibration of trust is not always perfect, and this results in over-trust and mistrust of the AI, which can lead to inefficiency. Over-trusting the AI can lead to complacency or automation bias whereas mistrust can lead to underutilization of the system. Trust can be characterized as dispositional trust, situational trust and learned trust and this study focused on characterizing situational trust while users interact with a recommender system (Hoff & Bashir, 2015).

AI technology is driven by machine learning models that could be powerful and complex (Adadi & Berrada, 2018). These complex models are less transparent and difficult for users to understand. These are called ‘black box’ models since they are not transparent (Kok et al., 2023), where the internal processes and the rationale behind the decisions are not accessible to the users. This lack of transparency of the black box models might lead to deception for the users, and increasing the transparency could increase the users' trust in AI as they can better understand AI’s actions and comprehend the decisions made (Kaplan et al., 2023). There is a growing emphasis on developing transparent and explainable AI (XAI) to solve the trust issues of black box models.

The concept of XAI is to provide transparency of the models and explanations of the decisions of the black-box models (Ali et al., 2023).

Transparency and explainability are important factors that could improve the trust a user has in the AI agent. Transparency of an AI system provides real-time status of what the system is currently doing (Arrieta et al., 2020), whereas explainability provides an explanation or reason for the actions or recommendations provided by the AI (Dosilovic et al., 2018). The explainability aspect aims to make the actions of the AI agent clear and understandable for the users, thereby improving trust in the system. The concept of levels of explanation in AI, as explained by Mueller et al., (2021), categorizes explanation types based on the complexity of the explanations that are provided for the recommendations. This ranking suggests that higher levels typically offer more relevant information for user understanding. These are eight levels of explanation that range from providing no explanation to providing a detailed explanation of the diagnoses of failures, as shown in Table 8. At a basic level, the AI presents basic features such as heat maps and decision boxes, which show visually the factors that influenced AI to make the decisions, whereas, at an advanced level, they could show instances of successful decisions or the details of the mechanisms behind how AI made the decisions. Furthermore, the higher levels could show the reasoning process of the AI's decision by presenting instances of failures to showcase the system's limitations and by providing comparisons to help users understand the important factors of the decision-making process. At the highest level, AI could provide detailed diagnoses of failures so that the users can understand the reasons behind the mistake.

***Table 8**:*

Levels of Explanation Scorecard (Mueller et al., 2021)

1	Null	No explanation
2	Features	Heat maps, bounding boxes, and linguistic features
3	Successes	Presents instances of AI generating recommendations
4	Mechanisms	Global description of how the AI works
5	AI Reasoning	Details of how AI is making decisions
6	Failures	Presets instances of failure
7	Comparisons	Provides comparisons for users to understand the process
8	Diagnoses of failures	Provides descriptions of AI failures

Explanations can be provided in different ways, such as textual, visual, audio, or haptic modalities. Textual explanations are often more intuitive and easier for end-users to understand, as they provide clear, natural language descriptions of a model's reasoning, making them accessible to users without technical expertise. It could potentially articulate complex model behaviors in an understandable way (Holzinger et al., 2022). Similarly, providing visual explanations could provide an intuitive understanding of the model along with improving trust in AI. The visual explanations can highlight the key aspects of the environment that helped AI to provide a recommendation. Using a hybrid model of visual and textual explanations could potentially leverage from the benefits of both methods. Previous research has shown that using a hybrid model of explanation improved the user's trust, understanding, and satisfaction. For example, Alam & Mueller, (2021) found that the use of a combination of visual and textual explanation were more effective in terms of trust and satisfaction than just using visual explanations for a medical diagnostic system. Another study investigating the influence of user expertise on understanding different methods of explanations (visual, textual and a combination of visual and textual explanations) found that using visual and textual explanations in a hybrid form resulted in a higher understanding of the explanations, especially for lay users (Szymanski et al., 2021). In this study, we have provided explanations and transparency in a hybrid model of visual and textual explanations to make it easier for the users to understand the reason behind the AI's recommendation.

There is a research gap in the literature that investigates the effect of explainability and transparency of the AI's recommendation while using AR technology to complete a procedural task. This study aims to fill this research gap by conducting an experimental study that investigates the effect of explainability and transparency of the AI agent's recommendations on users' perceived trust in the recommender AI agent while using AR technology to complete a pipe-fitting task. The task selected is one of the most common tasks in the AEC industry, which is completing a pipe fitting task for an HVAC system. One of the primary challenges in HVAC pipe fitting tasks is the complexity of selecting the appropriate components and of the correct size from a vast array of options, leading to time-consuming decision-making, potential errors, and inefficiencies in system installation and maintenance. These issues could potentially be addressed to an extent with the help of the recommender system that provides recommendations for the HVAC system installation. If an AI is trained well with the previous dataset on how the workers select the components and their size depending on the problem at hand, it could recommend the next component for similar scenarios in the future. The primary objective of this study is to identify factors that could improve the users' situational trust in the AI agent while completing the procedural task. Other variables investigated in the study include the task performance, the workload, and the users' preference for the AI agent after the users went through all the conditions. The findings from this research will be useful for designing AI agents in the future while using AR to complete a procedural task.

Research questions:

1. How do transparency and explainability in an AI agent influence users' trust in the AI while completing a procedural task?
2. How does the user's trust in recommender AI vary when integrated with transparency and explainability features, across both low and high levels of AI reliability?
3. How does the presence of transparency and explainability in an AI agent affect performance while completing a procedural task?
4. How does the user's perceived workload vary when they interact with recommender AIs with different features of transparency and explainability?
5. How does the preference for the recommender AI change when integrated with transparency and explainability features?

Hypotheses:

1. It is hypothesized that users will exhibit higher trust in AI agents that possess both transparency and explainability compared to agents with neither or only one of these features.
2. It is hypothesized that trust will be greater in AI agents with high reliability as opposed to those with low reliability.
3. The presence of transparency and explainability in AI agents, in combination with their reliability level, is anticipated to affect not only users' trust in AI but also their task performance. It is expected that users will require more time to complete the tasks when either transparency or explanation, or both, are presented.
4. It is hypothesized that the perceived workload will be lower when interacting with the recommender AI having both transparency and explainability features integrated.

5. It is hypothesized that the participants prefer the recommender AIs that provide either transparency, explainability or both over the one that provides only recommendations.

Method

Participants

The study was approved by Clemson University's Institutional Review Board (IRB2023-0356). 36 participants (19 males and 9 females), aged between 20 and 32 (M = 25, SD = 3.13) were recruited using fliers, emails, word of mouth, and by announcing in classes. The participants were civil engineering, architecture, construction science and management, industrial engineering and mechanical engineering students. Upon completion of the study, they were renumerated with either $10 or 1% of the course credit for the participating class for their time.

Table 9:

Participant Demographics

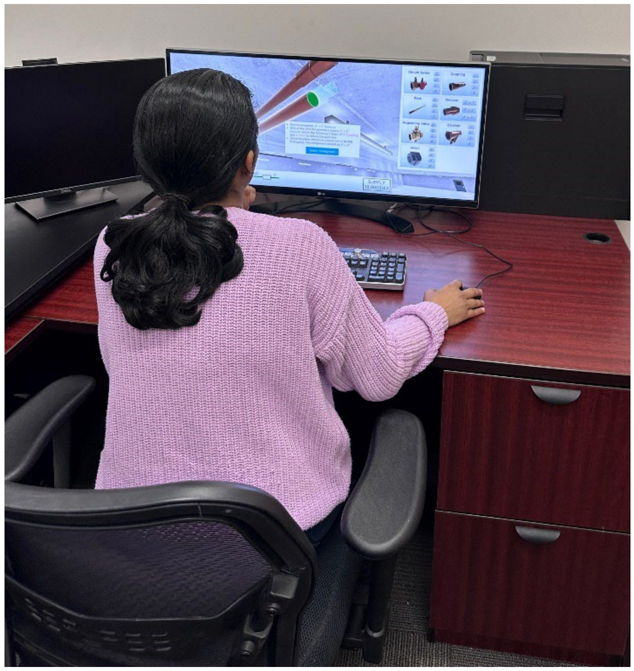
Variable (N = 28)	N	%
Gender		
Female	9	32.1
Male	19	67.9
Race		
American Indian or Alaska Native	1	3.6
Hispanic/Latino	2	7.1
African American	3	10.7
Caucasian/White	7	25
Asian	15	53.6
Degree Pursuing		
Undergraduate	9	32.1
Graduate	19	67.9
Specialization		
Electrical Engineering	1	3.6
Industrial Engineering	6	21.4
Civil Engineering	9	32.1
Mechanical Engineering	12	42.9

Experimental Setup

The experiment was conducted on a computer using a simulation developed for the study using Unity 3D. Each participant completed pipe fitting tasks in a simulated environment. The simulation emulates a user wearing augmented reality glasses while installing the pipe fittings, through which an AI assistant recommended the next component and the transparency and the explainability factors of the recommendation based on the condition. This simulation was run on a Dell desktop computer with an Intel(R) Xeon(R) CPU E5-1620 v4 processor, a Quadro FX 5800 GPU, and an ultrawide LG monitor with a screen size of 38 inches. The participants went through a pre-test questionnaire, a trust development questionnaire (within the simulation), a workload measurement and overall trust questionnaires. The pre-test (Appendix A), and the overall trust questionnaire (Appendix C) were administered using Qualtrics survey software. The trust development questionnaire was a 7-point Likert scale to measure trust in AI at a given time and was administered within the simulation. The overall trust questionnaire was administered before the update and after completing the task in every condition using a self-reported 8-item set of 5-point Likert scales (Hoffman et al., 2018) and the workload assessment tool (Appendix D) was administered after every condition using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) (Hart & Staveland, 1988). Figure 17 shows the experimental setup.

Figure 17:

Experimental Setup



Stimuli Development. The task selected for the experiment is a pipe fitting task associated with an HVAC unit. A subject matter expert in the domain of HVAC was consulted to develop the stimuli and to understand the mental model of the HVAC installer while completing the task. Detailed sessions were conducted to understand the factors that a worker considers before deciding on what the next component should be while completing the task. These factors included the thought process or the mental process that the installers follow to select a particular component at a particular location. In the absence of a predefined plan, the installers usually refer to the previous two components they completed and look at the final component or heat pump that they are connecting to the pipe. These factors were used to develop the scenarios and the explainability aspect of the AI.

In this experiment, since the study investigated the effective implementation of the recommender AI while using AR technology to complete a procedural task, we try not to over-clutter the visual space with the explanations. Instead, the study used the levels of explanation that the associated task and scenario warrants. A combination of levels 2, 4, 5, and 7 (features, mechanisms, AI reasoning, and comparisons) provides a detailed understanding of how the AI system provides recommendations. Providing basic features like the bounding boxes helps participants gain an initial understanding of the AI's analysis. Providing a global description and comparison, along with the reasoning behind why a particular recommendation was provided, will help them understand AI's thoughts much more deeply. Figures 18, 19, 20, and 21 show screenshots of different levels of AI features where the AI agent provides the recommendation. The participants were able to select the pipe fitting from the library of fittings on the right side. They were also free to remove the previous component if they had made a mistake by clicking the remove previous component button. In Figure 18, the AI provides only the recommendations, whereas Figure 19 shows the AI recommendation along with the transparency, and Figure 20 shows the AI recommendation along with the explainability. For the condition with both transparency and explainability, the transparency screen is followed by the explainability one, as shown in Figure 21.

Table 10:

Sample of Transparency and Explanation Provided in the Simulation

Feature	Details (Sample)
Transparency	• Looking for the previous two components • Looking for the type of heat pump getting connected to • Searching for previous data with similar situations
Explanation	• Recommendation: 3” coupling • 90% of the time the plumbers used a 3” coupling within the following 3 steps of 3” coupling and 3” pipe to continue the pipeline • Since the pipes should be connected to 9,000 BHU pump, this component should be 3”

Figure 18:

AI with no Transparency or Explanation

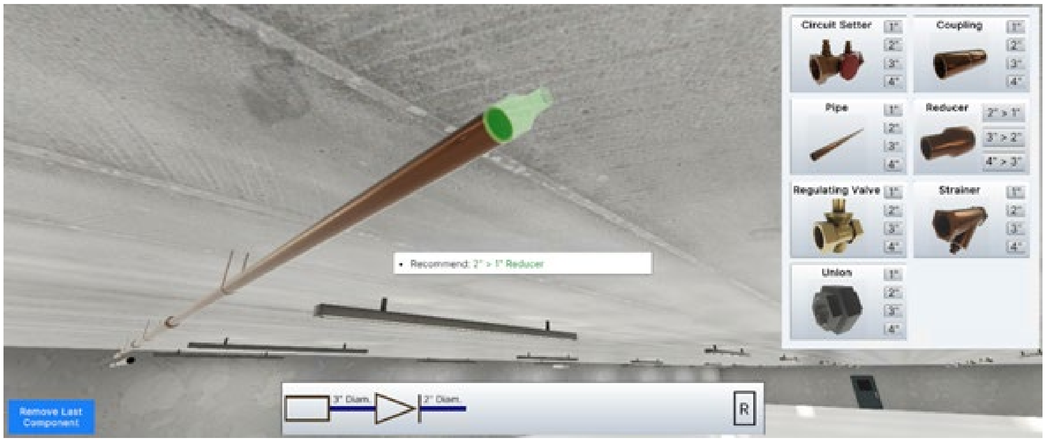


***Figure 16**:*

AI with Transparency

Figure 17*:*

AI with Explanation

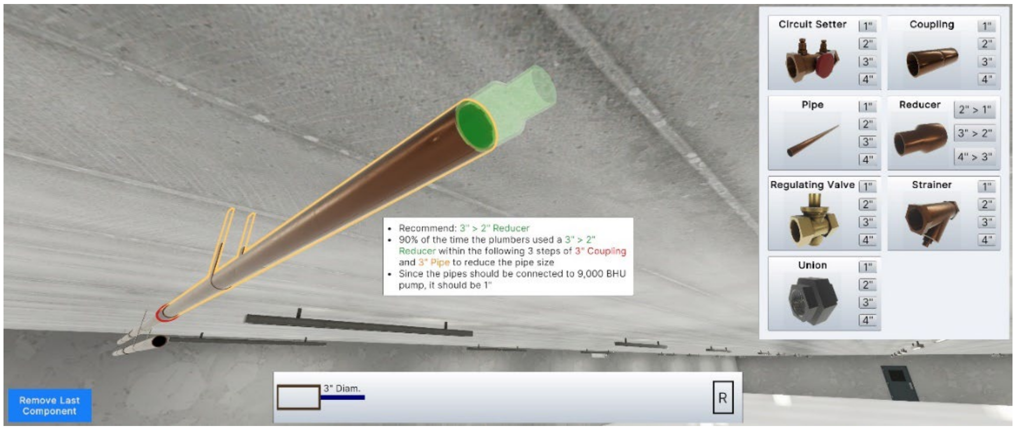
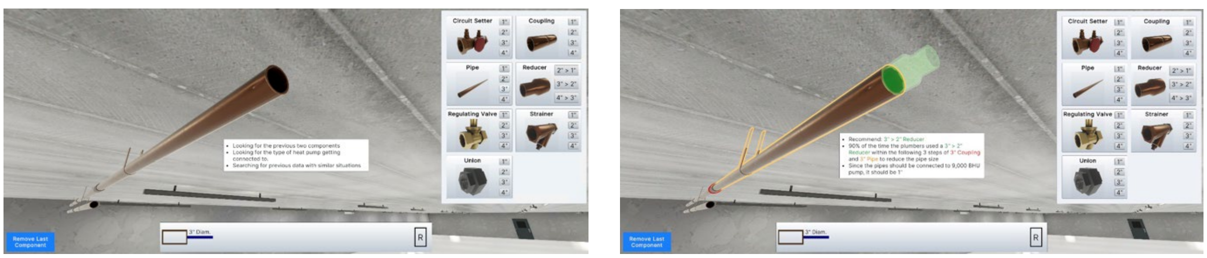


Figure 18*:*

AI with both Transparency and Explanation



Procedure

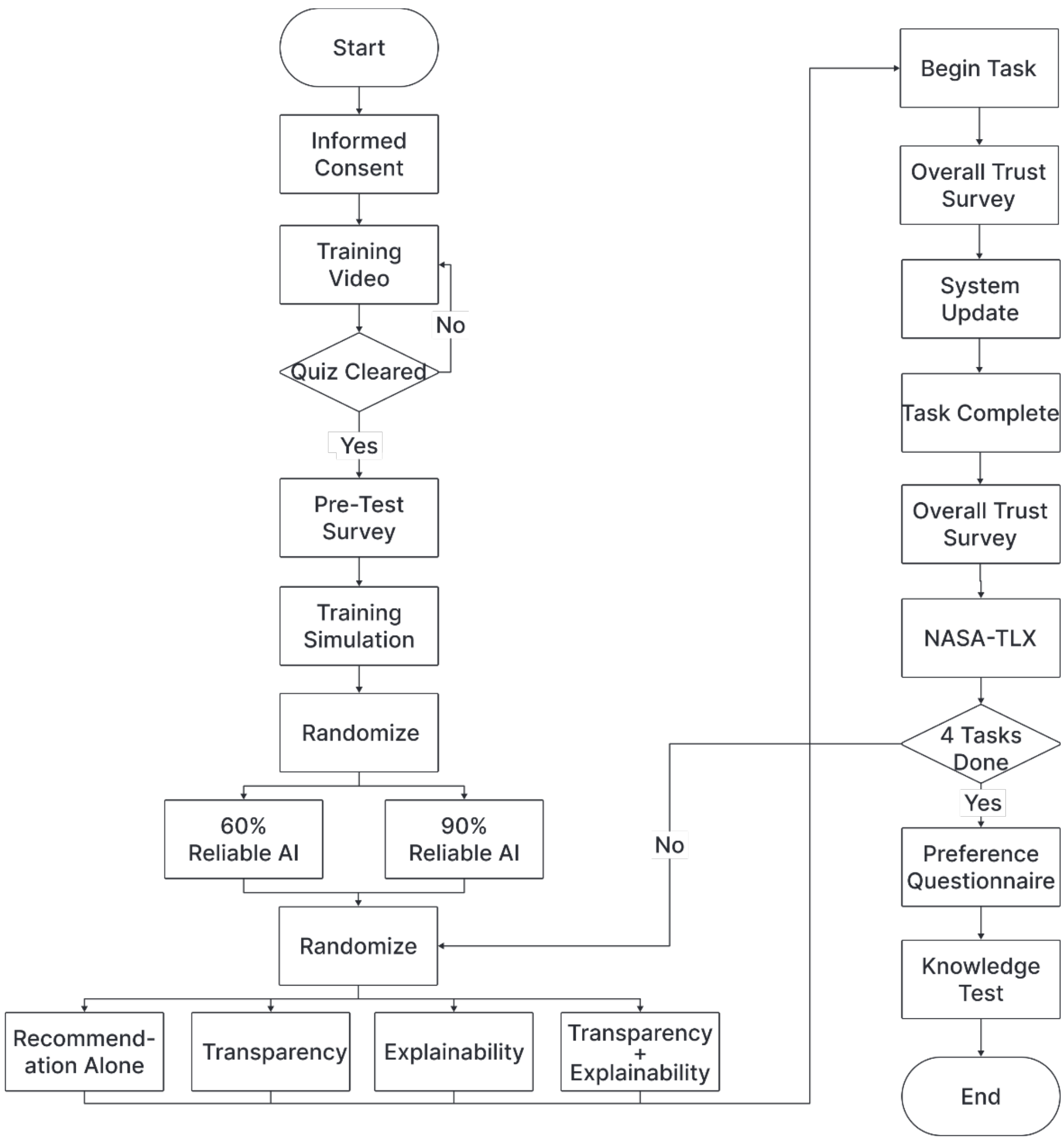
The participants were provided with study material and a training video prior to coming for the study, which included the concepts of the type of pipe fittings, their use, and the symbols used in the layout. The study material also involved the important components associated with the heat pump, such as the regulator valve, y strainer, etc., and the location/order where they usually occur. The video also contained instructions on navigating the simulation and how to complete the task. Once they were present for the study, they had the consent form to sign. Then they were asked to complete a pre-test questionnaire on Qualtrics, which included the demographic questionnaire. After completing the questionnaire, they were provided with the training video one more time that reinforced the content of the study material. Once they completed watching the video, they were asked to take a short quiz to ensure they were familiar with the concepts. If the participant did not score at least 80%, they were asked to watch the training video until they scored 80% on the quiz. Only 7% (2 out of 28) of participants did not score 80% on the quiz, which was attained by both the participants the next time they took the quiz after watching the training video one more time.

The next step was to open the simulation and was taken to the training session, where the participants were able to complete a simple pipe fitting task based on a plan, similar to the one in the Study 1. This ensured they were familiar with the simulation before starting the study session. Once the training was completed, the participant was taken to the study session of the simulation, and a reliability condition was selected based on the group to which the participant was assigned, which was randomized. They clicked the continue button to start the task. Initially, they were provided with an AI agent with 60% reliability with just the recommendations, without any explanation or transparency, regardless of their assigned condition. Trust development questions appeared every two steps which were presented on a 7-point Likert scale to understand their perceived trust level at that point in time during the interaction. There was a total of ten steps before the update and ten after.

Once they completed half of the task, they were notified of a system update and forced to click the update button. They were asked to complete an overall trust questionnaire just before the update using an 8-questionnaire survey on a 5-point Likert scale, administered on Qualtrics. Once the update was done, they had either a 60% or 90% reliable agent based on their assigned group. They were also provided with one of the four conditions of presenting the explanations/transparency based on the balanced Latin square design. Once they completed the task, they were asked to complete another overall trust questionnaire and the workload questionnaire using the NASA-TLX instrument. Then, they continued to the next condition and followed similar steps for the next three tasks. After the final task, the overall trust questionnaire had a preference questionnaire to select which AI agent they preferred. Before ending the study, they took the knowledge test the same as the one they took in the initial stages of the study. Figure 23 shows the flowchart of the procedure.

Figure 19:

Procedure Flowchart



Study Design

The study was a 2*4 mixed-subject design in which the participants were separated into two groups. Each participant in both groups experienced four different conditions. The tasks involved installing pipe fittings for an HVAC system. The participants were given a situation where the layout was half complete, and they had to complete the rest. They used pipes and different pipe fittings to connect the existing pipe structure to the heat pump. There were four different layouts (similar in difficulty level) for the four conditions to avoid the learning effect, which was randomized for the four conditions using a balanced Latin square. The participants were provided with a pre-test questionnaire, two overall trust questionnaires before the update and at the end of the four tasks, a trust development questionnaire at every two steps, workload measurement, and a preference questionnaire where they provided their preferred AI agent that they would like to work with. The pre-test, overall trust questionnaires and the preference questionnaire were administered using Qualtrics, whereas the trust development questionnaire was administered within the computer simulation. The trust development question was asked at every two steps of the task, whereas the overall trust questionnaires were asked at two different points during the task, one before the update and one after they completed the task.

Independent Variables. The independent variables were the reliability of the AI agent with two levels (60% and 90%) and the AI features with four levels (no transparency and explainability, transparency, explainability, and both transparency and explainability).

The reliability of the AI agent was manipulated as a between-subject variable, whereas the presence of explainability/transparency was manipulated as a within-subject variable. The reliability of the AI agents was based on the percentage of correct recommendations provided. For the low reliable AI, the reliability was set to 60% which means out of the total number of recommendations, 40% of the recommendations were incorrect. For the high reliable AI, the reliability was set to 90%, where only 10% of the recommendations were incorrect.

The first level of AI features variable did not provide transparency or explanations along with the recommendation. The transparency condition provided additional information on what the process AI is going through before providing the recommendations, whereas the explainability condition provided the justification for why a particular recommendation was provided. The final level was a combination of transparency and explainability features where the AI process was followed by the justification of the explanation.

Dependent Variables. The dependent variables measured are overall trust, trust evolution, task performance, and workload.

Trust. In this study, the trust the users have in the recommender AI was measured as a self-reported score. It was measured in two ways: One, using a trust development survey by administering a 7-point Likert scale that was asked every two steps, and second, the overall trust was measured using an 8-item questionnaire on a 5-point Likert scale. To measure trust in AI, different trust scales have been previously used in the literature (Cahour & Forzy, 2009; Jian et al., 2000; Merritt, 2011). Measurement scales for trust in XAI typically include assessments of both trust and reliance on the system's recommendations. These scales need to be sensitive to capturing both positive and negative aspects of trust and should consider the users' experiences and perceptions over time to gauge the appropriate level of trust and reliance on XAI systems (Hoffman et al., 2018). Some of the scales previously used are context-specific and cannot be generalized for every domain. In the context of XAI, Hoffman et al., (2018) recommended using a trust scale with 8 items of a 5-point Likert scale ranging from “I strongly agree” to “I disagree strongly," which was administered in this study. This overall trust questionnaire was administered two times in one task, one before the update and one after completing the task.

Task Performance. The performance was measured using the total time taken by the users to complete the task, measured in seconds, and also by the total number of errors made by the participants while completing the task. The purpose of measuring the performance was to investigate the effect of the different AI factors on task performance while completing the procedural task while completing it with the help of a recommender AI.

Workload. The workload was to investigate the effect of the AI features introduced, on the users' workload while completing the procedural task with the help of the recommender AI. This was measured once for every condition after they completed the task using the NASA-TLX instrument (Hart & Staveland, 1988).

Preference. After completing all the tasks, a preference questionnaire was asked to understand the preferred AI agent they would like to work with.

Data Analysis

Data analysis was conducted with the help of statistical analysis software IBM SPSS Statistics for Windows. For the overall trust, workload, and time taken, two-way mixed measures ANOVA was used, whereas negative binomial regression was used to analyze the number of errors. A Friedman's test, followed by Wilcoxon signed-rank tests, were used to analyze the preference data.

Results

Overall trust

The overall trust data was collected using a trust scale with 8 items of a 5-point Likert scale questionnaire (Hoffman et al., 2018), which was administered two times, one before the update and another after completing the task. The value collected from this survey shows the trust the users had in the recommender AI they were working with to complete the task. A higher value means higher trust in the AI, and a lower value means low trust.

A two-way mixed measures ANOVA was run to examine the influence of the reliability of the AI agent and the AI features (Recommendation, transparency, Explainability, Explainability Transparency combined) on the overall trust in the AI agent. The two-way interaction between the AI features and the reliability of the AI agent was not statistically significant, $F(3,78) = 0.70$, $p = 0.57$, partial $\eta^2 = 0.03$ (as shown in Figure 23), indicating that the impact of the AI features on overall trust did not differ significantly between high and low reliability levels of the AI agent. However, the main effect of AI features was statistically significant, $F(3,78) = 7.29$, $p < 0.001$, partial $\eta^2 = 0.22$ (as shown in Figure 24). The pairwise comparisons revealed significant differences: the baseline condition, which was recommendation alone, was perceived as significantly less trustworthy than the AI with transparency (mean difference = -0.47, 95% CI [-0.83, -0.11], $p = 0.005$), explanation (mean difference = -0.46, 95% CI [-0.76, -0.16], $p = 0.001$), and with both transparency and explanation (mean difference = -0.40, 95% CI [-0.76, -0.05], $p = 0.018$). The main effect of reliability was statistically significant as well $F(1,26) = 22.76$, $p < 0.001$, partial $\eta^2 = 0.47$ (as shown in Figure 25). Participants trusted the high reliability AI agent more than the low reliability one, with a mean difference of 1.02, 95% CI [0.580, 1.459], $p < 0.001$.

Figure 20:

Effect of AI Features and Reliability on Trust

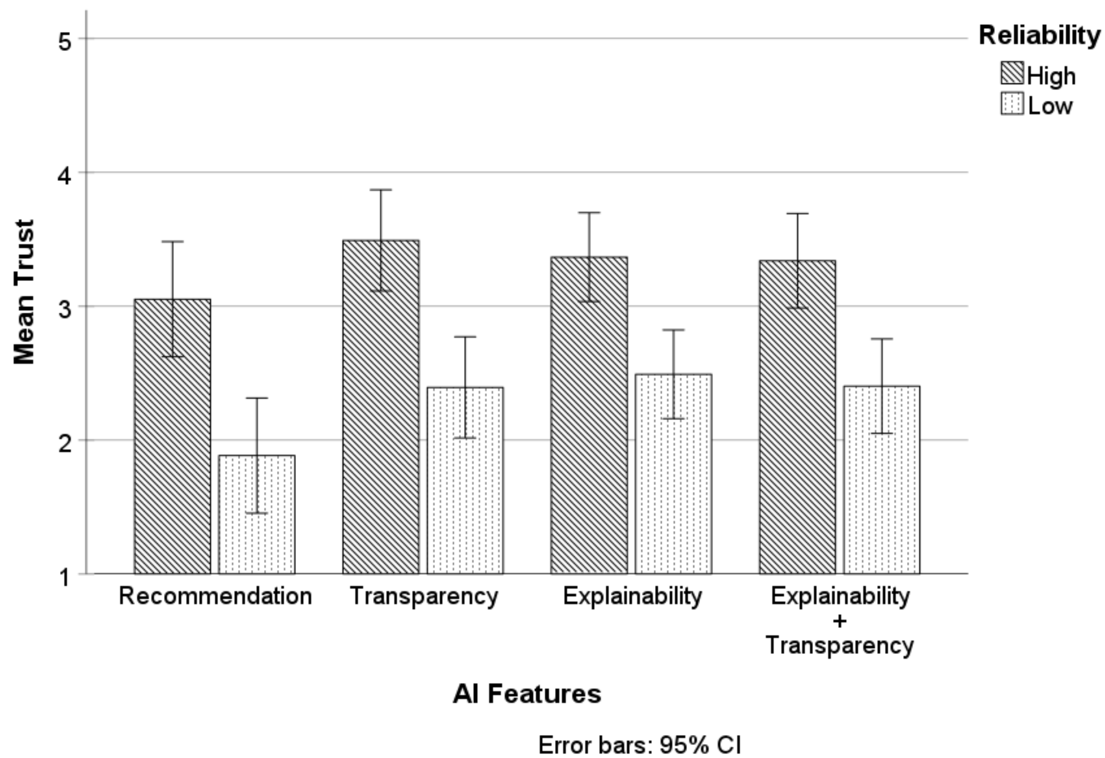


Figure 21:

Main Effect of AI Features on Trust

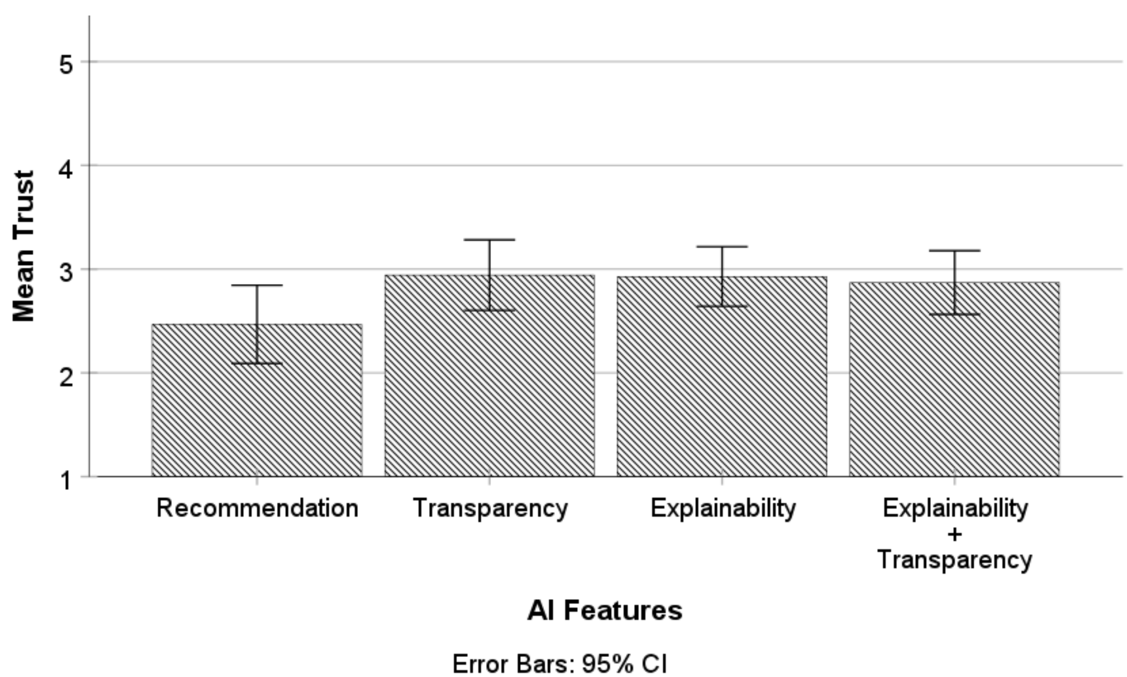
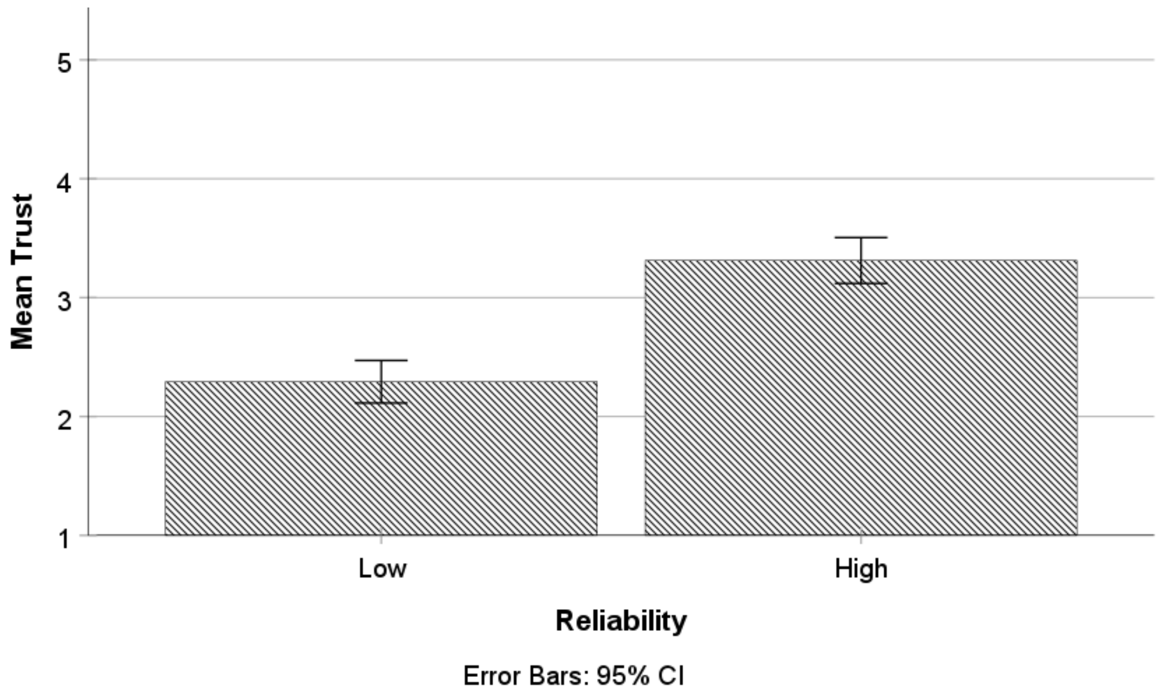


Figure 22:

Main Effect of Reliability on Trust



Performance

Task Completion Time

The total time to complete the task was measured within the simulation. This measures the total time taken by the participants to complete each task. The higher values of time taken mean they took longer to complete the task and vice versa.

A two-way mixed ANOVA was conducted to determine the effect of AI features and the reliability of the AI agent on the time taken to complete the task. The interaction between the AI features and the reliability of the AI agent was not statistically significant, $F\ (3,78) = 0.76$, $p = 0.519$, partial $\eta^2 = 0.03$ (as shown in Figure 26), indicating that the impact of the AI features on completion time did not differ significantly between high and low reliability levels of the AI agent.

The main effect analysis shows that the total time taken to complete the task was statistically significantly different between the different levels of AI features $F\ (3, 78) = 39.28$, p

< .001, partial η^2 = 0.60 (as shown in Figure 27). The pairwise comparisons show that the time taken to complete the task was significantly shorter in the recommendation condition compared to the transparency condition (mean difference = -59.58 seconds, 95% CI[-69.10, -50.06], p <0.001), explainability condition (mean difference = -21.93 seconds, 95% CI[-30.88, -12.98], p <0.001), and with both explainability and transparency condition (mean difference = -67.10 seconds, 95% CI[-76.32, -57.88], p <0.001). The main effect of the AI agent's reliability on time taken was not statistically significant, $F(1, 26) = 1.86$, $p = 0.184$, partial η^2 =0.07 (as shown in Figure 28), indicating no significant difference in overall completion times between the high and low reliability conditions.

***Figure 23**:*

Effect of AI Features and Reliability on Time Taken

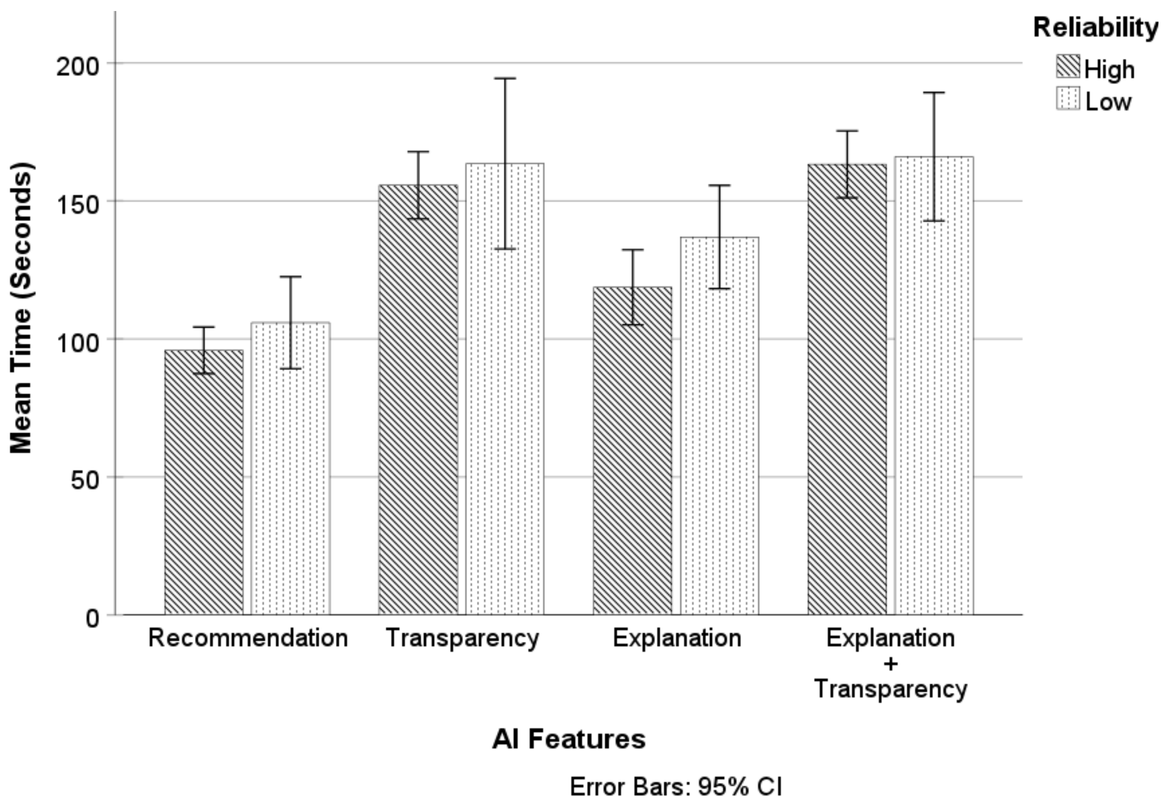


***Figure 24**:*

Main Effect of AI Features on Time Taken

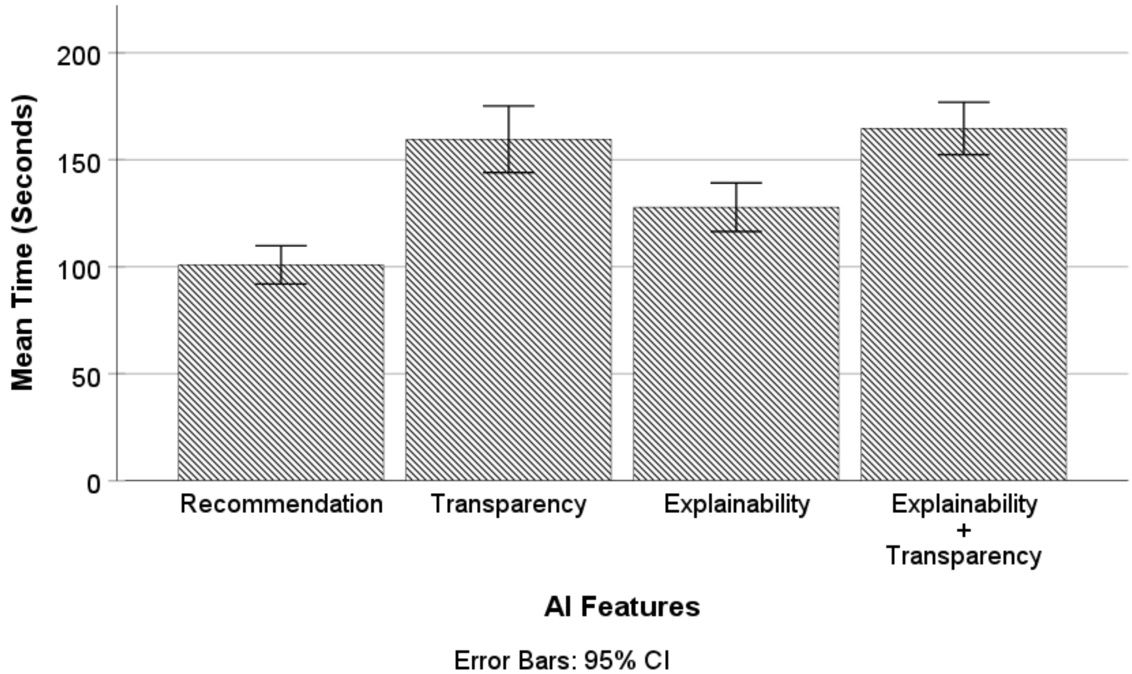
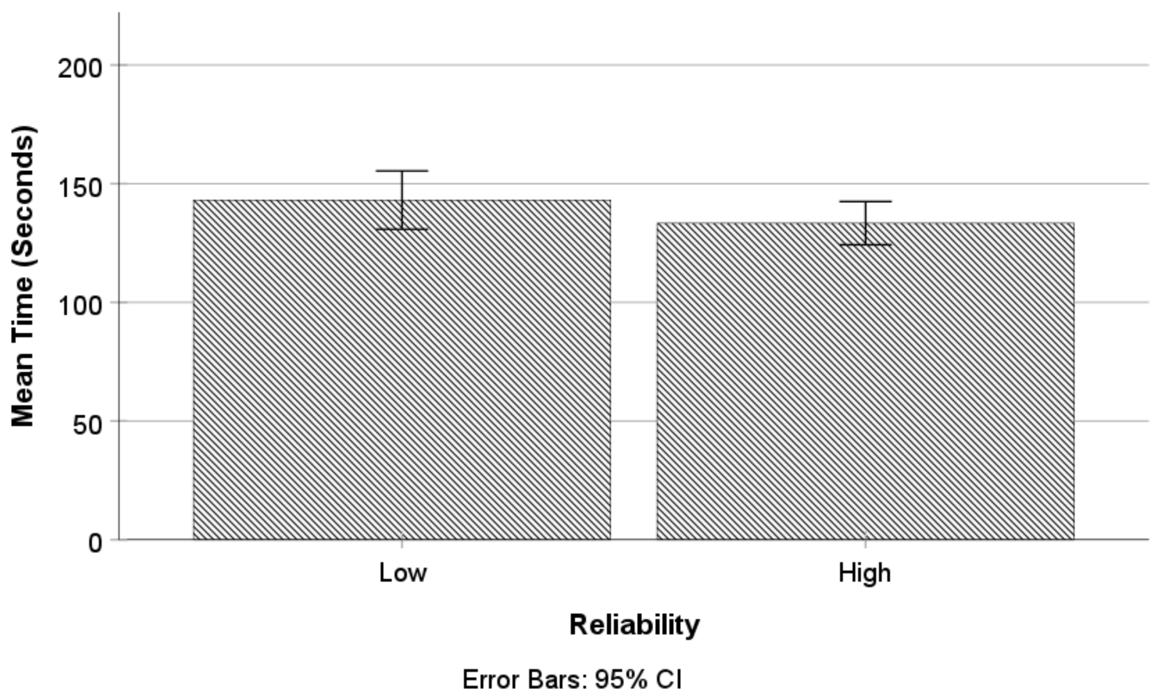


***Figure 25**:*

Main Effect of Reliability on Time Taken



Number of Errors

The number of errors refers to the number of times a user made a mistake while completing the task. This error could be the wrong selection of the size of the component or the component itself. This was collected from the simulation where the final numbers were produced after completion of the experimental study.

A negative binomial regression analysis was conducted to examine the effects of the reliability of the AI agents and the AI features (recommendation, explainability, transparency, and explainability combined with transparency) on the number of errors made. The negative binomial, instead of the traditional Poisson regression, was used since the variance exceeded the mean resulting in over-dispersion. The analysis revealed that the reliability of the AI agent did not have a significant effect on the number of errors made, with participants interacting with low reliable agents showing no statistically significant difference in error rates compared to those interacting with high reliable agents (B = -0.00, SE = 0.33, 95% CI [-0.66, 0.65], Wald $\chi^2 = 0.00$, df = 1, p = .99). Similarly, the different levels of AI features also did not yield a significant difference in the number of errors made, as indicated by the overall test for AI features effects (Wald $\chi^2 = 1.48$, df = 3, p = .686).

Task Completion

This refers to whether the participant completed the task correctly or not. If a mistake was made, it was considered not completed. A binomial logistic regression was performed to ascertain the effects of the reliability of the AI agent, AI features, and the number of errors on the likelihood of participants completing the task. The logistic regression model was statistically significant, $\chi^2(5) = 72.94$, $p < 0.001$. The model explained 90.4% (Nagelkerke R^2) of the variance in task completion and correctly classified 97.3% of cases. The number of errors showed a negative association with task completion, but this was not statistically significant ($p = 0.992$), indicating no reliable effect of errors on task completion under the model tested. There was also no significant effect of the reliability of the AI agent or AI features on the likelihood of task completion.

Hits

Another way to measure performance is to evaluate operator responses to AI recommendations by utilizing Signal Detection Theory (SDT), focusing on the operators' ability to correctly identify correct recommendations (hits) and accurately reject incorrect ones (correct rejections). These two metrics; hits and correct rejections, serve as fundamental indicators of the AI's effectiveness in supporting decision-making and the operators' proficiency in utilizing AI assistance. Since the overall accuracy can be effectively measure using hits and correct rejections, the false alarms and misses were not considered for the analysis. Hits reflect the instances where operators correctly recognized and accepted correct AI recommendations, indicating effective decision-making in alignment with the AI. Conversely, correct rejections represent scenarios where operators correctly identified and disregarded incorrect recommendations, showcasing their ability to critically evaluate AI's recommendation. A Generalized Linear Mixed Model (GLMM) with a negative binomial distribution was conducted to explore the effects of the AI agent's reliability and AI features (recommendation, transparency, explainability, and explainability combined with transparency) on the number of times participants accepted a correct recommendation from the AI agent. The negative binomial was used instead of the Poisson regression since the data was over-dispersed with the variance exceeding the mean. The analysis of fixed effects revealed a significant effect of the model, $F(4, 107) = 259.63$, $p < 0.001$, indicating that the predictors, as a set, significantly predicted the number of hits. Within the model, reliability emerged as a significant predictor $F(1, 107) = 1034.92$, $p < 0.001$, whereas AI features did not significantly predict the number of hits $F(3, 107) = 1.20$, $p = 0.313$. The number of hits for the high reliable condition was 1.51 (95% CI, 1.47 to 1.55) times the number of hits in the low reliable condition, a statistically significant result, $p < 0.001$.

Figure 26:

Mean Hits for Different Reliability

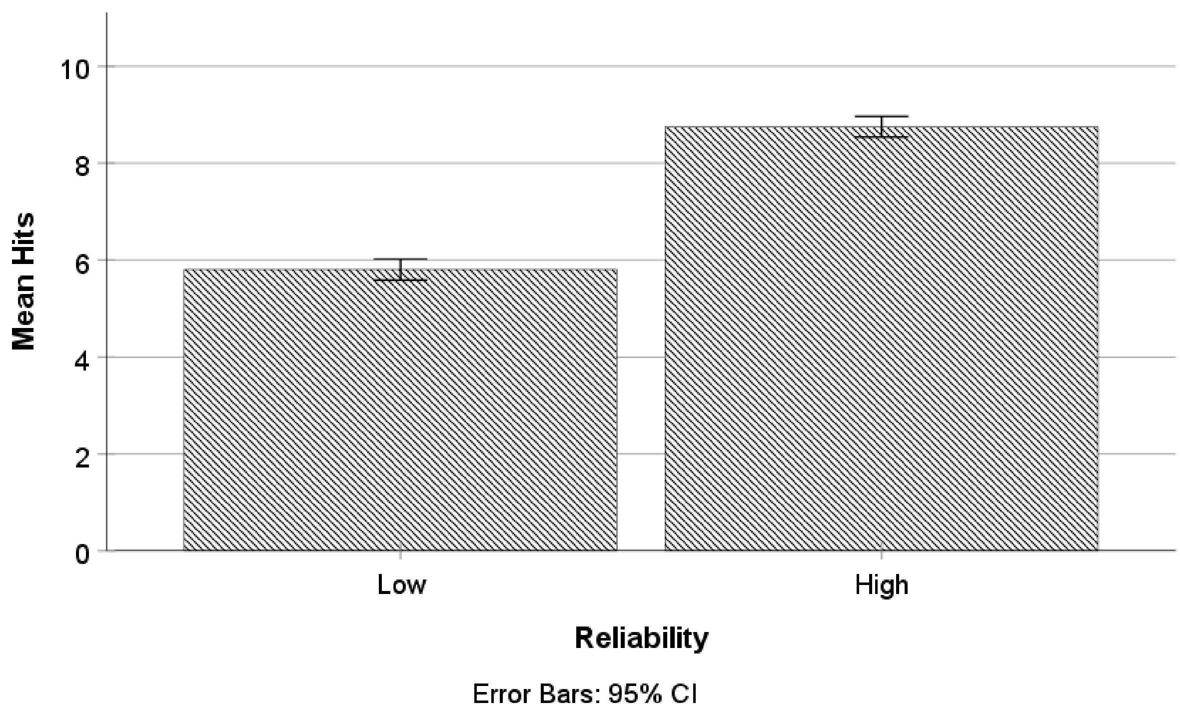
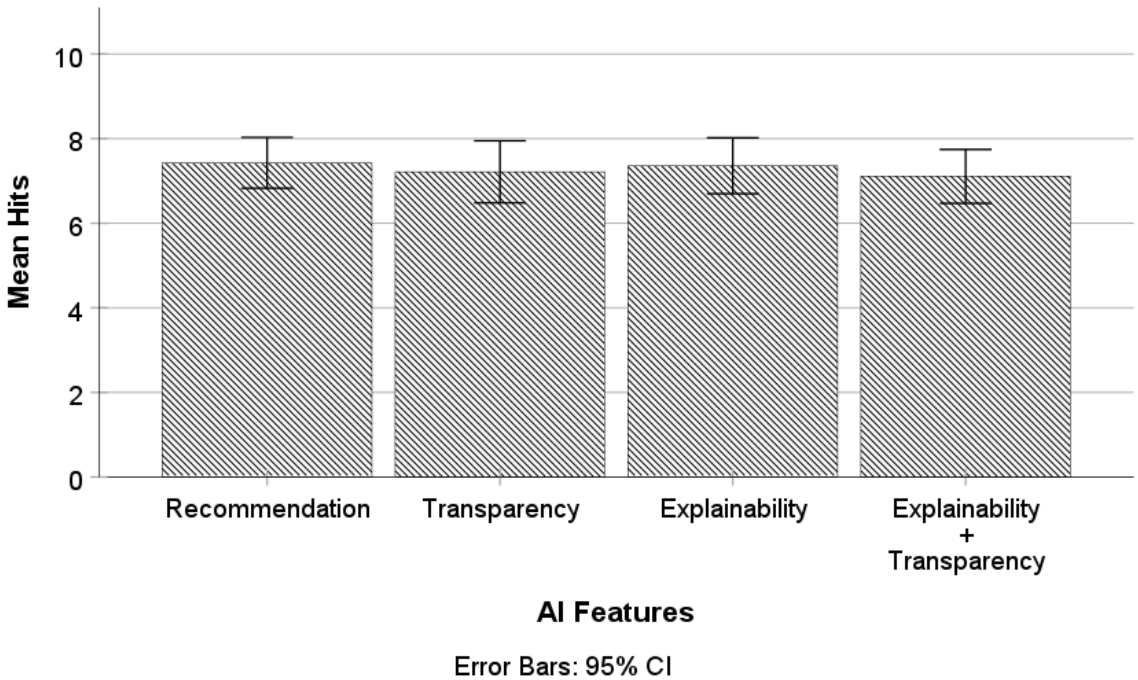


Figure 27:

Mean Hits for Different AI Features



Correct Rejection

A GLMM with a negative binomial distribution was conducted to explore the effects of the AI agent's reliability and AI features (recommendation, transparency, explainability, and explainability combined with transparency) on the number of times participants correctly rejected an incorrect recommendation from the AI agent. The negative binomial was used instead of the traditional Poisson regression due to over-dispersion of the data since the variance exceeded the mean. The analysis of fixed effects revealed a significant effect of the model, $F(4, 107) = 365.00$, $p < 0.001$, indicating that the predictors, as a set, significantly predicted the number of correct rejections. Within the model, reliability emerged as a significant predictor, $F(1,107) = 1459.77$, $p < 0.001$, whereas AI features did not significantly predict the number of correct rejections, $F(3,107) = 0.07$, $p = 0.974$. For the high reliable condition, the number of correct rejections was 0.24 (95% CI, 0.23 to 0.26) times the number of correct rejections in the low reliable condition, a statistically significant result, $p < 0.001$.

Figure 28:

Mean Correct Rejections for Different Reliability

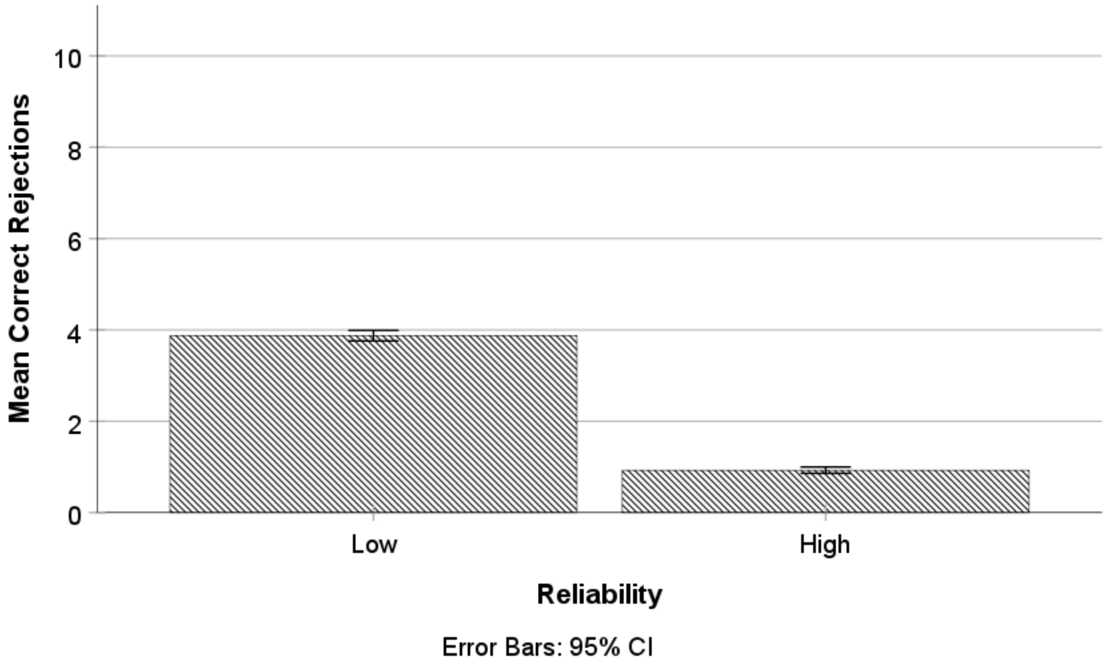
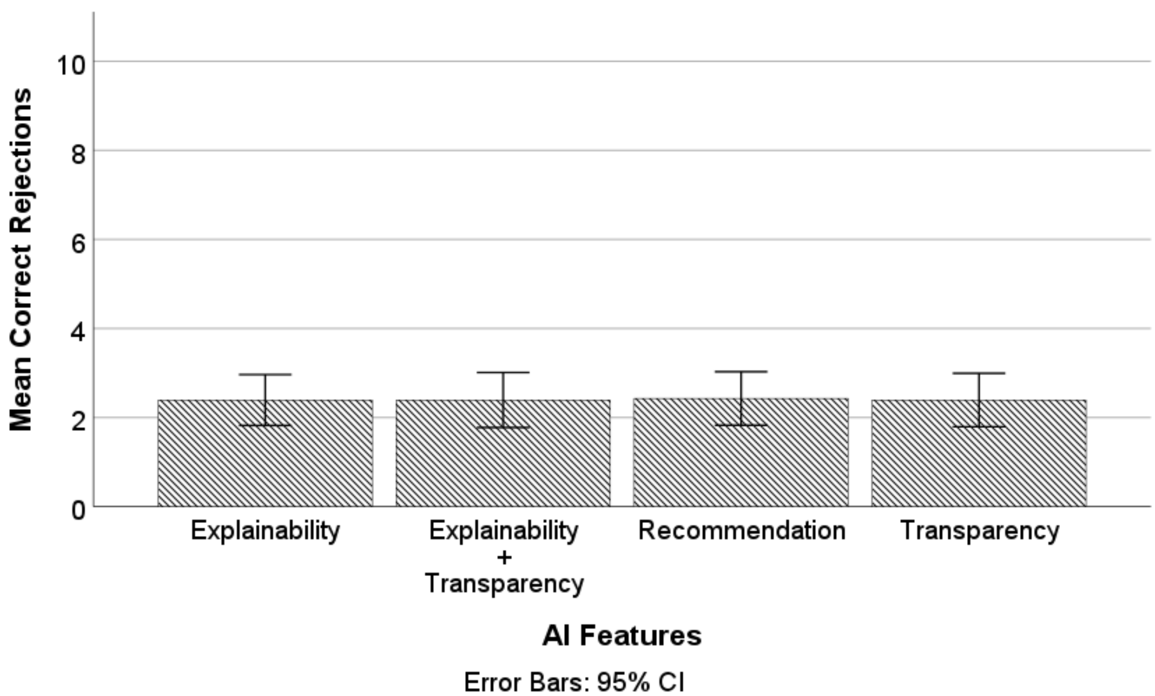


Figure 29:

Mean Correct Rejections for Different AI Features



Workload

The workload was measured using a NASA-TLX survey that was administered after completion of each task. The total workload is measured as a sum of different scales which are mental demand, physical demand, temporal demand, performance, effort, and frustration level. A higher value of total workload shows that the user's workload was higher.

Two-way mixed measures ANOVAs were conducted to determine the effect of the AI features (recommendation, transparency, explainability, explainability transparency combined) and the reliability of the AI agent on different subscales of workload. There was no statistically significant interaction between the different levels of AI features and the reliability of the AI agent for any of the sub-scales and the total workload. There was also no statistically significant main effect of AI features on different sub-scales and total workload. There was also no statistically significant main effect of the reliability on different sub-scales and total workload except for the sub-scale, temporal demand, $F(1,26) = 7.99$, $p = 0.009$, partial $\eta^2 = 0.24$. A pairwise comparison revealed that participants perceived higher temporal demand in the high reliable condition compared to the low reliable condition, with a mean difference of 4.26, 95% CI [1.16, 7.35], $p = 0.009$.

Figure 30:

Effect of AI Features on Workload

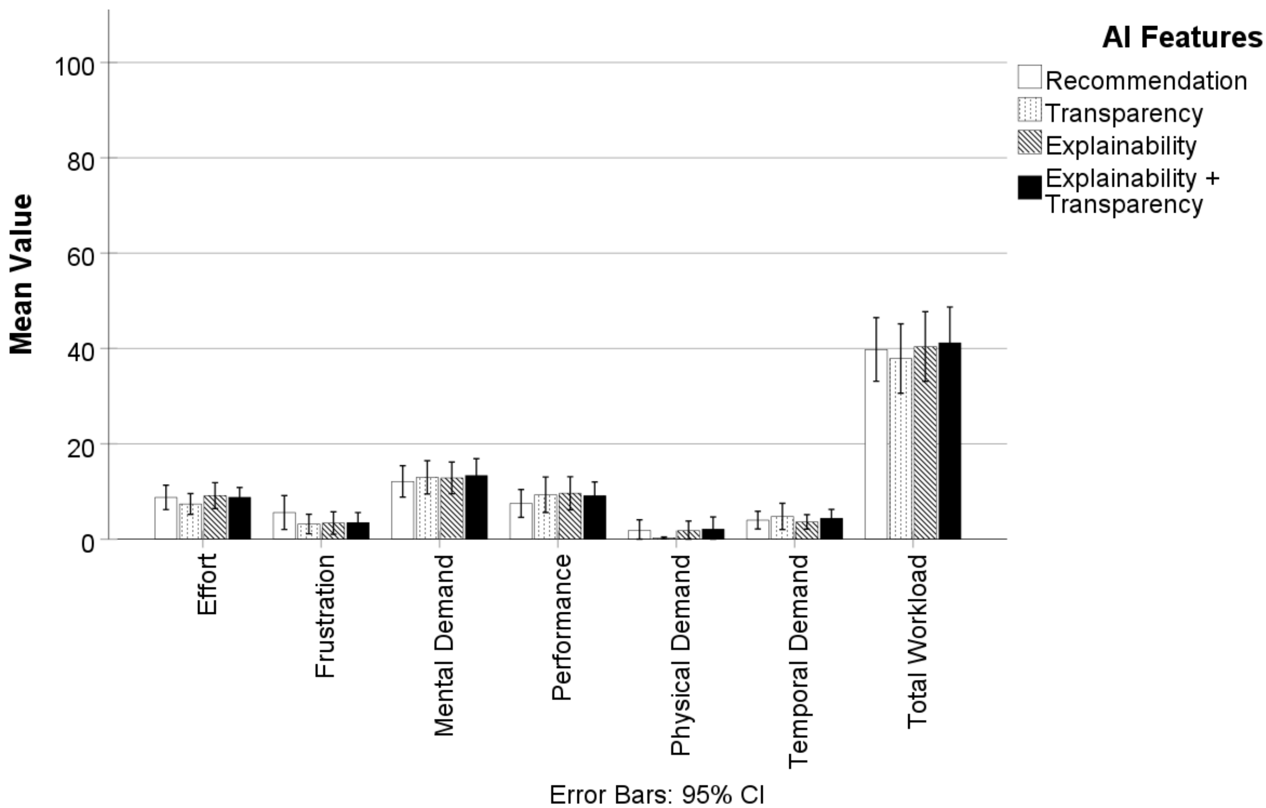
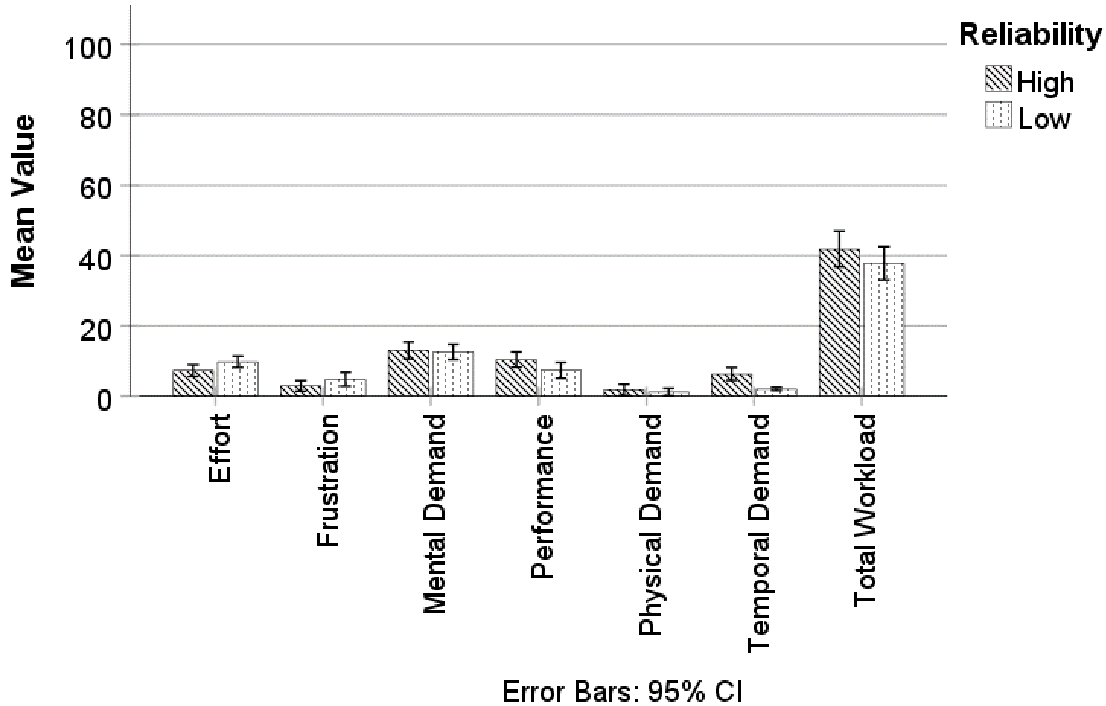


Figure 31:

Effect of Reliability on Workload



Preference

A Friedman's test was conducted to analyze the ranked data for the preference of agents. There was a statistically significant difference in the ranking of agents which recommended the pipe fittings, $\chi^2(3) = 19.84$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction. The median interquartile range for the recommendation, transparency, explainability and the combined condition were 4 (3 to 4), 3 (2 to 3), 2 (2 to 3) and 1 (1 to 2.75), respectively. A ranking of 1 indicates the highest preference, while 4 indicates the lowest. There was a statistically significant difference in the ranking, where participants preferred the transparency condition over the recommendation condition ($Z = -2.81$, $p = 0.005$), the explainability condition over the recommendation condition ($Z = -2.99$, $p = 0.03$), and the explainability and transparency combined condition over the recommendation condition ($Z = -3.13$, $p = 0.002$). However, there was no significant difference in ranking between the explainability and transparency condition (Z = -0.45, p = 0.65), the explainability transparency combined condition and transparency condition (Z = -1.67, p = 0.096), the explainability transparency combined condition and explainability condition (Z = -1.75, p = 0.08).

Discussion

Situational trust that the users have in AI is affected by external and internal factors during the user's interaction with the system (Hoff & Bashir, 2015). External factors refer to the factors such as the type of automation, the complexity of the automation, task difficulty etc. whereas the internal factors refer to the user factors such as self-confidence, expertise in the task, attentional capacity etc. These factors collectively affect the relationship between trust and the reliance that the user has on the AI.

The interaction between AI reliability levels and the AI features did not yield a statistically significant difference in trust following the AI update. This outcome indicates that the effect of AI features on trust in AI does not differ significantly between low and high reliable conditions. In other words, the presence or absence of transparency or explainability did not affect how the users trusted high and low reliable AI. This could be specific to this task and could be explored further for more complex tasks in the future to understand if there could be any difference in the effects of different complexities of tasks in AEC.

Consistent with the results from the first study and existing literature (Hoff & Bashir, 2015; Kaplan et al., 2023; Lee & See, 2004) and aligning with the second hypothesis, the main effects indicate that the higher reliability in AI agents significantly enhances users' trust. The relationship between the trust the user has in the system and their reliance on the system is affected by different internal and external factors. One such factor that affects the strength of the relationship is the user's ability to compare automated performance to manual. In this experiment, the task was such that the users could identify when the AI provided a wrong recommendation resulting in clearly understanding the performance of AI. When the reliability of the recommender AI was high, and since the users understood the performance of the system, the users might have relied more on the recommendation provided by the AI thus increasing their trust in AI. This result underscores the foundational role of reliability in trust formation and suggests that ensuring high reliability in AI systems is crucial for their acceptance and effective use in complex domains like the AEC sector.

The analysis of within-subject effects revealed significant differences between the recommendation condition and transparency alone, explanation alone and both transparency and explanation combined conditions, with the recommender condition yielding lower trust, aligning with a part of the first hypothesis. The users trusted the AIs that provided information to the user on how or why a recommendation was provided compared to not providing them. This reinforces the important role of transparency and explanation in AI systems in maintaining appropriate trust, aligning with previous literature that suggests that transparency and explanation improve the trust the users have on AI (Alufaisan et al., 2021; Arrieta et al., 2020; Panganiban et al., 2020). The increase in trust when transparency and explanation were provided with the recommendation could be because the users could understand the AI process well and the justifications provided were able to convince the user to accept a particular recommendation. The visual-textual information or distractors provided are known to increase the trust the users have in AI (Phillips & Madhavan, 2011). The effect of these distractors depends on the level of interaction they have with the user such that, if the interaction is high the users might trust the AI more and conversely, the trust could decrease or could have no effect. In this study, the transparency and explanation were provided in such a way that the users had to interact with both features to move to the next step (to select the component). This higher level of interaction could be one of the reasons for increased trust in the three conditions with transparency or explanation or both compared to the condition with no transparency or explanation. Future research could further explore this aspect by investigating the effect of different ways of providing explanations and their interactions, such as adaptive explanations where transparency and explanations are provided only when the user needs them, on trust for completing a similar task in the AEC domain. Since AEC deals with tasks of different complexities, future research could also investigate different complexities of the task, such as low vs high, while incorporating the adaptive explanation aspect of AI to understand its effect on trust.

For task completion time, there was no significant interaction between the different transparency/explanation conditions and the agents' reliability suggesting that the efficiency of task completion is similar for the different transparency/explanation conditions across the two levels of reliability. As expected, the task completion time was lower in the recommendation condition, where no transparency or explanations were present compared to the other three conditions, supporting hypothesis three. The reason for this could be mainly because the participants had to read through the explanations and the transparency aspects of the AI in the other three conditions, whereas, in the recommendation condition, they can select the component right after viewing the recommendation provided by the AI. The participants might have taken more processing time on the transparency and explainability aspects when these elements were present. Also, prior research has shown that response time can increase as more information is displayed to the user since their information processing demand increases (Helldin, 2014; Zuk & Carpendale, 2007). Further research is required to investigate effective ways to implement transparency and explanations without increasing the task completion time. Specifically, they could investigate the effects of providing transparency and explanation along with the recommendation in an adaptive manner. This way, users get required information when they need it and not continuously, without affecting their trust in AI. By reducing the number of times the transparency and explanation is provided, it can be hypothesized that the total time to complete the task could go down significantly.

The findings from the negative binomial regression analysis focusing on the impact of reliability and AI features on the number of errors made found that both the reliability and AI features did not significantly influence the number of errors. This result diverges from Study 1 and previous research that investigated the effect of the reliability of automation aid on human performance while completing a visual search task (Shah & Bliss, 2017), which showed that the number of errors varies based on the reliability of the AI, such that as the reliability decreases, the number of errors increases. The number of errors did not vary between the four different levels of AI features as well. This could be because the particular task used in this study was not sufficiently sensitive to the variation in AI reliability or the task used was simple enough to show potential differences in error rates. In addition to that, the analysis that investigated the likelihood of completion of the task found that it was independent of the reliability of the AI agent, the AI features and the number of errors made. This further affirms that the simplicity of the task might have led to such a result.

There was no significant interaction between the reliability and AI features in relation to multiple components of the workload construct suggesting a uniformity in perceived workload across the four levels of AI features and two levels of reliability of the recommender AI. The absence of a statistically significant effect of different levels of AI features on the total workload and the different subscales of workload suggests that providing transparency and explanation did not add any workload to the user while completing the task. This result aligns with previous research that shows workload did not increase as a function of transparency (Mercado et al., 2016).

This could be because of the way the transparency and explanation were provided in this experiment, which was good enough not to increase the participant's workload while completing the procedural task. This finding is important because, the additional visual and textual components that were presented on the screen did not increase the workload of the users, but it increased the trust the users have in the AI. In the case of reliability variable, it influenced the temporal demand such that the users felt high temporal demand when they interacted with the high reliable AI compared to the low reliable AI. This could be due to the fact that the higher reliable AI made the users rely more on the AI thus increasing the reliance that could have slowed down the user's decision-making process. This slowing down might have added a sense of temporal demand.

Incorporating SDT to analyze operators' responses to AI recommendations, this study identified the AI agent's reliability as an essential factor in enhancing decision-making accuracy, reflected through significant effects on hits and correct rejections. Despite varying AI features (recommendation, transparency, explainability, and their integration), the recommender AI's reliability, rather than the specific nature of the information presented, predominantly influenced operators' ability to accurately accept correct recommendations and reject incorrect ones. This observation suggests that the role of transparency and explainability in AI systems, which is important for building user trust, may not directly affect users' decision accuracy in distinguishing whether AI recommendations are correct or not, echoing the results from a previous study that investigated the effect of transparency (Loft et al., 2023). This could be explored further to understand how different levels of transparency and recommendation could have an impact on decision accuracy, and it could also look at the different difficulty levels of tasks. For different levels of reliability, the number of times the participants accepted the correct recommendation provided by the AI agent was statistically higher in the high reliable condition as compared to the low reliable condition. This is because the number of correct recommendations provided by the AI was higher in the high reliable condition compared to the low reliable one. The high reliable AI had a reliability of 90% whereas the low reliable AI was set to 60%. From Figure 29, it can be seen that the high reliable AI had a mean hit of nine out of ten recommendations, whereas the low reliable one had six hits out of ten recommendations. This shows that the user's perceived reliability of the AI was in line with the actual reliability of the AI. This is also evident from Figure 31, where the mean correct rejection of high reliable AI was one, whereas, for low reliable AI, it was four. This shows that the users had appropriate trust in AI since the perceived reliability matched the actual reliability of the system.

After completing the final task, the participants were asked to rank the AI agents that assisted them in completing the task. The study findings from Friedman's test indicate that the participants preferred the conditions with transparency, explanation or both with transparency and explanation along with the recommendation over the ones with just the recommendation, supporting hypothesis five. This could be due to the fact that the participants could understand the process that AI goes through and the reason for a particular recommendation, and they trusted them, as discussed earlier. The post hoc analysis showed that even though most of the participants ranked the agent with both transparency and explanation, there was no statistical difference in preference between the condition with explanation, transparency and the condition with both transparency and explanation. This aligns with the overall trust results that showed that the recommendation condition had lower trust compared to all three other conditions of AI features. The users trusted the AI agent in a similar fashion when transparency or explanation or both of them wer provided, but the trust was low for the recommendation condition. This might have been reflected in their preference for AI agents as well.

Conclusion

This study explored the effects of the reliability of AI agents and its transparency and explainability on users' trust, workload, and performance while completing a procedural task. The study used a 2*4 mixed-subject experimental design, to understand the dynamics between AI reliability, transparency, and explainability, revealing interesting insights into their collective impact on user trust and performance.

The findings from this study suggest the importance of integrating transparency and explainability into recommender AI systems, illustrating how such features could potentially enhance users' trust. This is an important finding, as trust serves as an important aspect of user reliance on AI for decision-making support. However, this integration comes with a complex trade-off: while users prefer AI agents that offer transparency and explanations, preferring these over just recommendations, this preference is accompanied by an increase in task completion time. Interestingly, this increase does not correlate with increased workload, suggesting that the added time may be due to the users engaging more deeply with the information provided rather than experiencing it as an increased cognitive load.

This relationship between enhanced trust and increased task completion time, without the addition of workload, highlights the necessity for a strategic approach to the implementation of transparency and explainability within AI systems. It indicates toward investigating when to provide these features to ensure that the appropriate trust is maintained without compromising efficiency. The next study investigates this research question to understand when it is beneficial to provide transparency and explanation to maintain appropriate trust.

This study contributes significantly to the literature on effective human-AI collaboration, identifying the essential role of transparency and explainability in building trust while acknowledging the practical challenges associated with these features. The findings could serve as a basis for designers and researchers to explore ways to implement transparency and explanation with the recommender AI to maintain appropriate trust.

CHAPTER FIVE

EVALUATING THE EFFECT OF ADAPTIVE EXPLANATIONS AND TRANSPARENCY ON MAINTAINING TRUST IN ARTIFICIAL INTELLIGENT AGENTS

Introduction

Human's trust in AI is important in effectively implementing the technology (Asan et al., 2020). Trust in AI is a complex concept that significantly influences user acceptance, reliance, and adoption of AI technologies (Bach et al., 2022; Lee & See, 2004; Lockey et al., 2021). Trust can be conceptualized as dispositional, situational, and learned trust (Hoff & Bashir, 2015). Situational trust is affected by various internal factors such as the operator's expertise, and self-confidence and external factors such as system complexity, task difficulty and environmental risks. Previous research has shown the importance of adding transparency and explanation aspects of AI to improve trust (Papenmeier et al., 2022). The explanations justify the underlying mechanisms, rationales, and decision processes of AI, enabling users to comprehend the actions and outputs of AI models. With the advancements in machine learning, AI systems are becoming increasingly complex, and it becomes difficult to explain (Laato et al., 2021). These are black box models that are opaque to the end user about how the AI makes recommendations or decisions, and it affects human trust in AI and, in turn, results in the disuse or misuse of the technology (Pynadath et al., 2018; von Eschenbach, 2021). To tackle this issue, explainable AI (XAI) has been a prime research topic to avoid the opaqueness of the black box models while maintaining their performance (Meske et al., 2022). The role of explanations in AI, a crucial component of XAI, has been instrumental in enhancing user understanding, trust in AI, and effective management of AI systems (Zeng et al., 2018).

However, explanations are a cognitive knowledge transfer process between humans and XAI systems (Miller, 2019). Previous research indicates that unnecessary explanations can cause cognitive overload and a subsequent decline in performance (Herm, 2023; Szymanski et al., 2021). This may make it more difficult for the user to process and understand the information being presented as explanations. Also, the user's reliance on the AI recommendation might be affected when unnecessary explanations are provided, and this might lead to a false sense of trust in the system and potentially negative outcomes, especially when the recommendations are incorrect (Jakubik et al., 2022).

Sometimes providing explanations continuously with every AI recommendation may not be necessary and can cause potential disadvantages, especially in time-sensitive tasks or while providing explanations/transparency on the AR glasses, since it might disconnect the user from the real world more than required. A previous study found that most participants did not require explanations when the recommendations matched the user's expectations (Riveiro & Thill, 2022). When the recommendation from the AI aligns with that of the user, the explanation becomes a piece of redundant information that could increase the task completion time and the cognitive load to perceive and comprehend the explanation. They could ignore explanations that are not necessary, but that will take up a part of the user's attention. Further, providing explanations for recommendations that do not require them might increase the system's computational resources, resulting in decreased system performance and making the real-time decision-making process difficult. However, providing an explanation at the right time is also important as it might overburden the user when the explanation is provided when it is not needed, and it will be important to provide explanations when the AI is making mistakes since that is when the users will stop relying on the AI (Alam & Mueller, 2021). Thus, developing a system that could adapt to the situations and provide explanations, rather than providing at every decision point, would be helpful for the user to complete the task more efficiently.

Adaptive XAI that provides explanations for the recommendations only when the user needs it aligns with the concept of personalized XAI. Personalized XAI considers the user's domain knowledge, trust propensity, and persistence in providing explanations (Millecamp et al., 2019). By adapting explanations to the user's needs, AI systems can enhance user understanding and trust without overwhelming them with unnecessary information. Previous studies have mainly examined the "what" part of XAI where they investigated the effects of visual and textual explanations (Alam & Mueller, 2021; Szymanski et al., 2021). However, there is a gap in the literature that investigates the "when" part of providing explanations and transparency along with the recommendations. This study tries to fill this gap by investigating the "when" aspect of providing explanations and transparency. This study investigates the concept of adaptive explanations and transparency, focusing on providing them depending on their necessity.

Research questions:

1. Does providing explanations in an adaptive manner, instead of continuously, decrease the trust in AI?
2. How is the performance affected when the explanations are provided in an adaptive manner?
3. How is the workload affected when the explanations are provided in an adaptive manner instead of continuously?
4. How does the preference of the recommender AI vary when the transparency and explanations are provided continuously versus in an adaptive manner?

Hypotheses:

1. Providing transparency and explanations in an adaptive manner will not reduce the trust the users have in the recommender AI compared to providing it continuously.

2. When the explanations are provided continuously compared to the other conditions, the task time will be high, whereas the number of errors will not see any significant difference between conditions.
3. When the explanations are provided in an adaptive manner, the workload will not increase compared to providing them continuously.
4. It is hypothesized that the preference for the AI with adaptive explanations will be more than that for the one with continuous explanations.

Method

Participants

The study was approved by Clemson University's Institutional Review Board (IRB2023-0356). The participants were recruited using fliers, emails, word of mouth, and by announcing in classes. The participants were civil engineering, architecture, construction science and management, industrial engineering, and mechanical engineering students. A total of 30 participants (24 male and 6 female), aged between 22 and 48 (M = 28, SD = 5.8), were recruited for the study. Upon completion of the study, they were renumerated with either \$15 or 3% of the course credit for the participating class for their time.

Table 11:

Participant Demographics

Variable (N = 28)	**N**	**%**
Gender		
Female	6	20
Male	24	80
Race		
Prefer not to answer	1	3.3
Caucasian/White	9	30
Asian	20	66.7
Degree Pursuing		
Undergraduate	6	20
Graduate	24	80
Specialization		
Mechanical Engineering	1	3.3
Electrical Engineering	3	10
Architecture	3	10
Civil Engineering	5	16.7
Industrial Engineering	9	30
Construction Science and Management	9	30

Experimental Setup

The experimental setup was similar to that of the study 2. A computer simulation that emulated a user wearing augmented reality glasses while completing a pipe fitting task was developed for the experiment. This was conducted on a Dell desktop computer with an Intel(R) Xeon(R) CPU E5-1620 v4 processor, a Quadro FX 5800 GPU, and an ultrawide LG monitor with a screen size of 38 inches, where an AI agent recommended the next steps to complete the task. The task was the same as that used in the previous experiment, where the participants had to complete a half-complete pipe layout associated with a heat pump. The explanations/transparency were provided along with recommendations for all the steps or a few of them based on the condition they were assigned to. The pre-test questionnaire (Appendix A) and the overall trust questionnaire (Appendix C) were administered using the Qualtrics survey software, whereas the trust development questionnaire was provided within the simulation after every two steps. The workload assessment tool (Appendix D) was administered after every task using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) (Hart & Staveland, 1988). The experimental setup can be seen in Figure 35.

***Figure 32**:*

Experimental Setup

The task selected for the experiment is the same as study 2, which is completing a pipe layout associated with an HVAC unit. This was developed after consultation with subject matter experts. The levels of explanations used in this study were a combination of levels 2, 4, 5, and 7 (features, mechanisms, AI reasoning, and comparisons) similar to the previous study. The three conditions that the participants went through were one which provided transparency and explanations for all the recommendations provided by the AI, one where transparency and explanations were provided only when AI’s confidence was less and the final condition where transparency and explanations were provided at the first, middle and the last step. Figure 36 shows the transparency and explanations, and Table 12 provides a sample of transparency and explanation provided.

Figure 33:

Transparency and Explanations

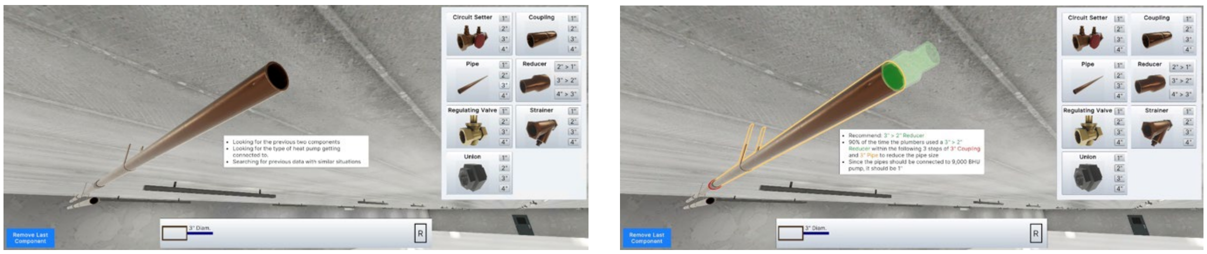


Table 12:

Sample of Transparency and Explanation Provided in the Simulation

Feature	**Details (Sample)**
Transparency	• Looking for the previous two components • Looking for the type of heat pump getting connected to • Searching for previous data with similar situations
Explanation	• Recommendation: 3” coupling • 90% of the time the plumbers used a 3” coupling within the following 3 steps of 3” coupling and 3” pipe to continue the pipeline • Since the pipes should be connected to 9,000 BHU pump, this component should be 3”

Procedure

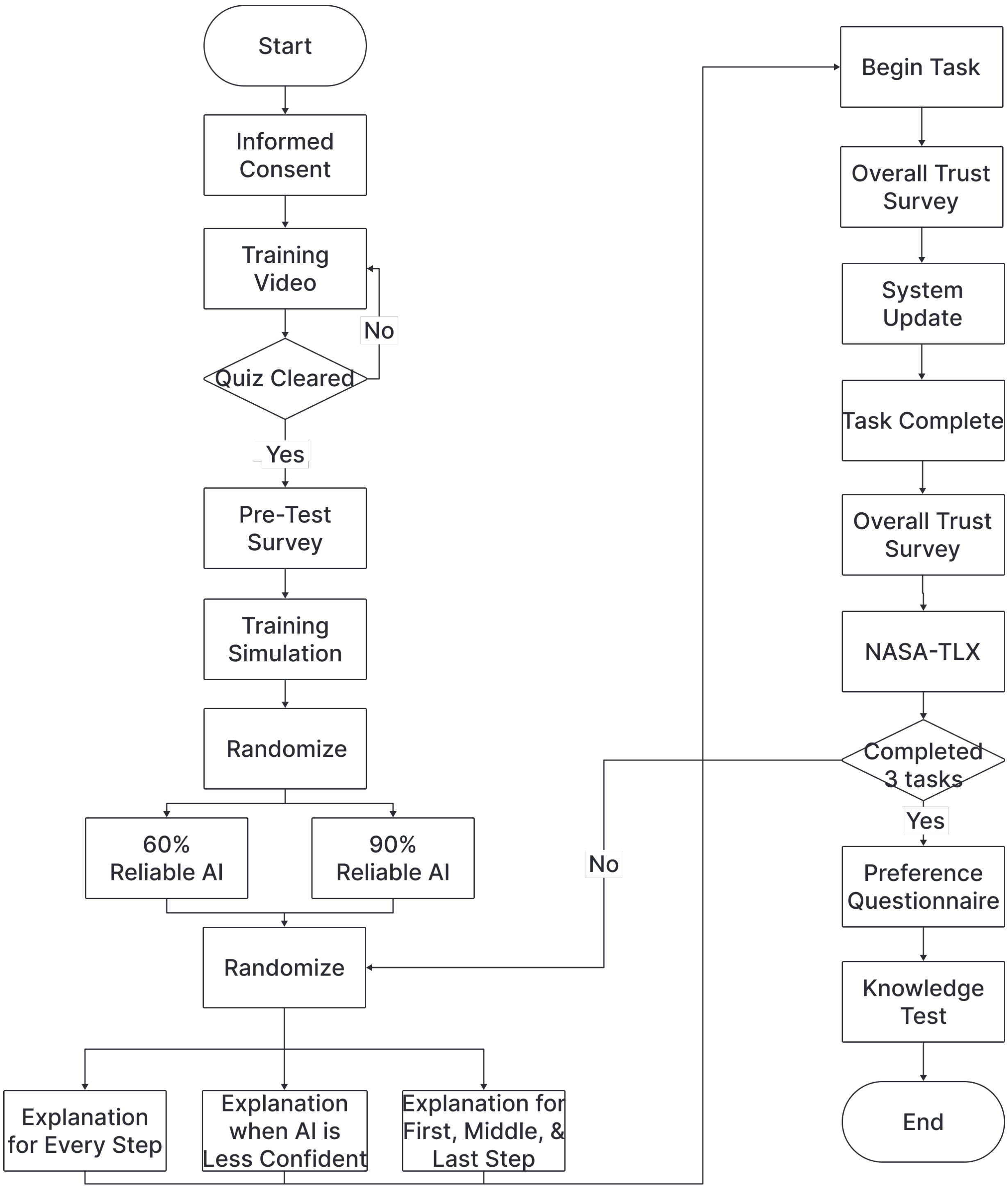
The task for this experiment was to complete an incomplete pipe layout associated with a heat pump. The participants were briefed before each task about the task they were about to complete. They were asked to use the library of components in the simulation to complete the layout. The AI provided the recommendations, along with explanations, either continuously or only when needed, based on the condition they were assigned to.

To ensure the participants were familiar with the pipe fitting task and associated components, they were provided with study material a day before the experiment. The study material comprised the basic use of different components associated with the heat pump that was used for the task. Once they arrived for the study, after signing the consent form, they were asked to watch a training video that helped them understand the different uses of pipe fittings and instructions on how to complete the task in the simulation. This video also reinforced the study material contents. After that, they were asked to complete a short quiz to ensure they understood the associated components of the heat pump and the pipe fitting task they were about to complete. To move forward with the study, they had to score at least 80% on the quiz; upon failing, they watched the video until they got the required score. Only 7% (2 out of 30) of participants had to retake the quiz to get an 80% score on the quiz.

Once the researcher ensured that the participants had enough knowledge to complete the task, they received a pre-test questionnaire that included the demographic questionnaire and their baseline trust in AI. Once they completed this step, they were provided with a training session in the simulation to familiarize themselves with how to complete the task in the simulation. Once that was done, they were provided with the task to complete, where AI provided recommendations for the next component to select. In the simulation, they received the 7-point Likert scale trust development questionnaire every two steps. In the middle of the task, they received an update. Before the update, every participant in every task received an AI with low reliability that provided only the recommendations. After the update, they were randomly assigned to the condition of explanation provision. Before they interacted with the updated AI, they received a trust questionnaire to collect the overall trust in AI (Hoffman et al., 2018). After each task, they were asked to complete the overall trust questionnaire again and the NASA-TLX survey. After the final task, the overall trust questionnaire included an additional provision to provide a preference ranking of the AI agents they worked with. After that, they were asked to retake the quiz that was taken initially one more time. Figure 37 shows the flowchart of the procedure.

Figure 34:

Procedure Flowchart



Study Design

The study is a 2*3 mixed-subject study, where the participants were randomly assigned to one of the conditions of reliability. All the participants went through the three conditions of provisions of explainability and transparency and were randomized using a balanced Latin square design. There were three different layouts for the three tasks to reduce the learning effects, but they were of similar difficulty in order to reduce the variability between the conditions.

Independent Variables. The independent variables were the reliability of the AI agent varied in two levels: low and high, and explanation/transparency provision (further mentioned as explanation provision) with three levels: continuous, when the agent's confidence is low (60% confidence), and at the first, middle, and last steps (further mentioned as a spaced condition). The explanation provision was randomized as a within-subject variable in three levels. In the first level, transparency and explanations were provided along with all the recommendations after the update. In the second level, transparency and explanations were provided only when AI's confidence was low for a particular recommendation. This provided information helps the users to make an informed decision while the confidence of AI for that recommendation is low (incorrect recommendation). In the third level, transparency and explanations were provided in the first, middle and final steps. By providing transparency and explanations at the first step ensures that the users understand the process and justification for a recommendation provided by the AI. After a few interactions, this is further reinforced by providing this information in the middle and the final steps of the task.

The reliability of the AI agent was manipulated as a between-subject variable with two levels: low reliable AI and high reliable AI. The low reliable AI had a reliability of 60%, which means that only 60% of the recommendations provided by the AI were correct. The high reliable

AI had a reliability of 90%. The incorrect recommendations were randomized to achieve the required reliability.

Dependent Variables. The dependent variables measured were overall trust, trust evolution, task performance and workload.

Trust. The trust evolution was measured using a 7-point Likert scale that is provided every two steps during the task. The overall trust was measured using an 8-questionnaire survey on a 5-point Likert scale (Hoffman et al., 2018).

Task Performance. The performance of the users while completing the task was measured using two metrics: total time taken to complete the task and the number of errors made while completing the task. Both these metrics were measured within the simulation.

Workload. The workload was measured using the NASA-TLX survey (Hart & Staveland, 1988) and was administered at the end of each task.

Preference. After completing the three tasks, the participants were asked to complete a ranking survey to select their preference for the AI agent they worked with while completing the task.

Data Analysis

Data obtained from the experimental study was analyzed using the statistical analysis software IBM SPSS Statistics for Windows. For the overall trust, workload and the time taken, two-way mixed measures ANOVA was used, whereas negative binomial regression was used to analyze the number of errors. A Friedman's test was used to analyze the preference data.

Results

Overall Trust

The overall trust data was collected using a survey of 8 items with a 5-point Likert scale (Hoffman et al., 2018), which was administered two times, one before the update and one after completing the task. This was a self-reported trust rating, which reflects the trust the user had in the recommender AI while completing the task. The higher the value of trust, the more they trust the AI compared to the one with a lower value.

A two-way mixed measures ANOVA was conducted to examine the influence of the AI agent's reliability and the explanation provision on the trust in the AI following an update. The interaction between the explanation provision and the AI agent's reliability was not statistically significant, $F(2, 56) = 1.04$, $p = 0.362$, partial $\eta^2 = 0.04$ (as shown in Figure 38) indicating that the effect of the explanation provisions on trust post-update did not significantly vary between high and low reliability levels of the AI agent. The main effect of the explanation provision on trust was also not significant, $F(2, 56) = 1.01$, $p = 0.373$, partial $\eta^2 = 0.04$ (as shown in Figure 40). However, the main effect of reliability on trust was significant, $F(1, 28) = 18.43$, $p < 0.001$, partial $\eta^2 = 0.40$ (as shown in Figure 39). Subsequently, pairwise comparisons were conducted to examine the statistically significant main effect of reliability. This indicated that the trust in AI post-update was significantly higher in the high reliable than in the low reliability condition, with a mean difference of 0.88, 95% CI [0.46,1.31].

Figure 35:

Effect of Explanation Provisions on Trust

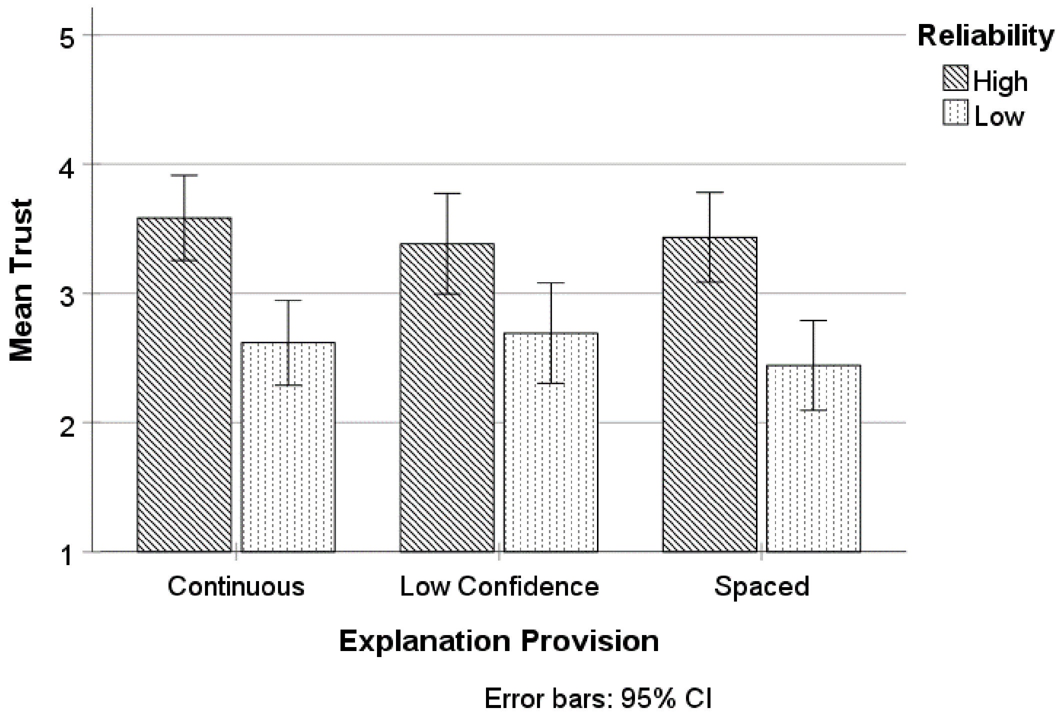


Figure 36:

Main Effect of Reliability on Trust

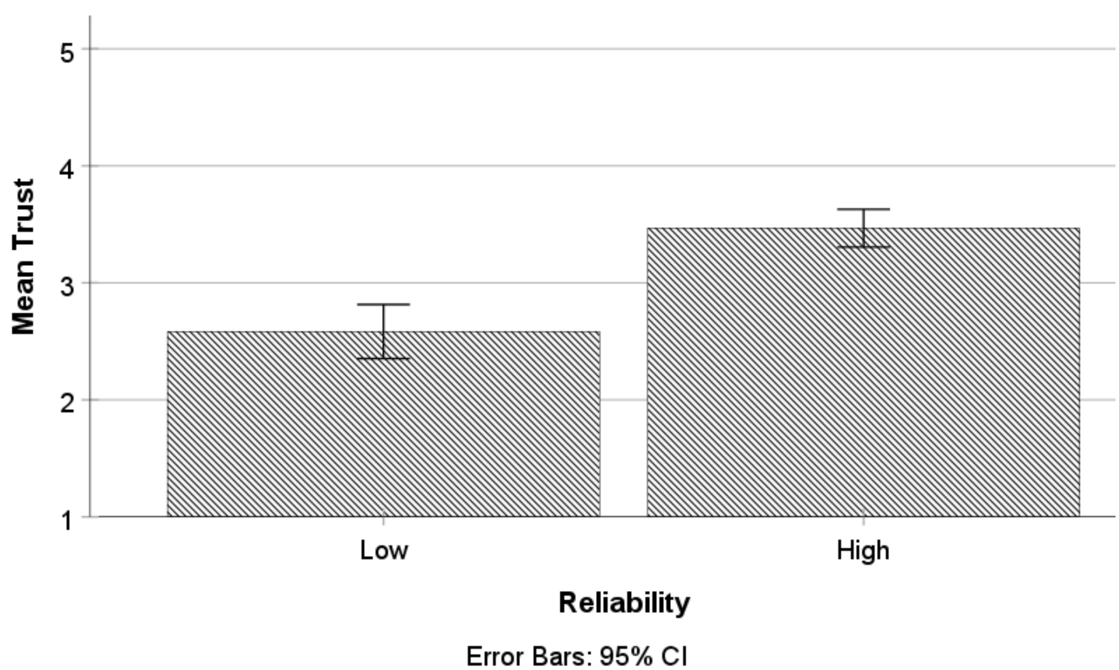
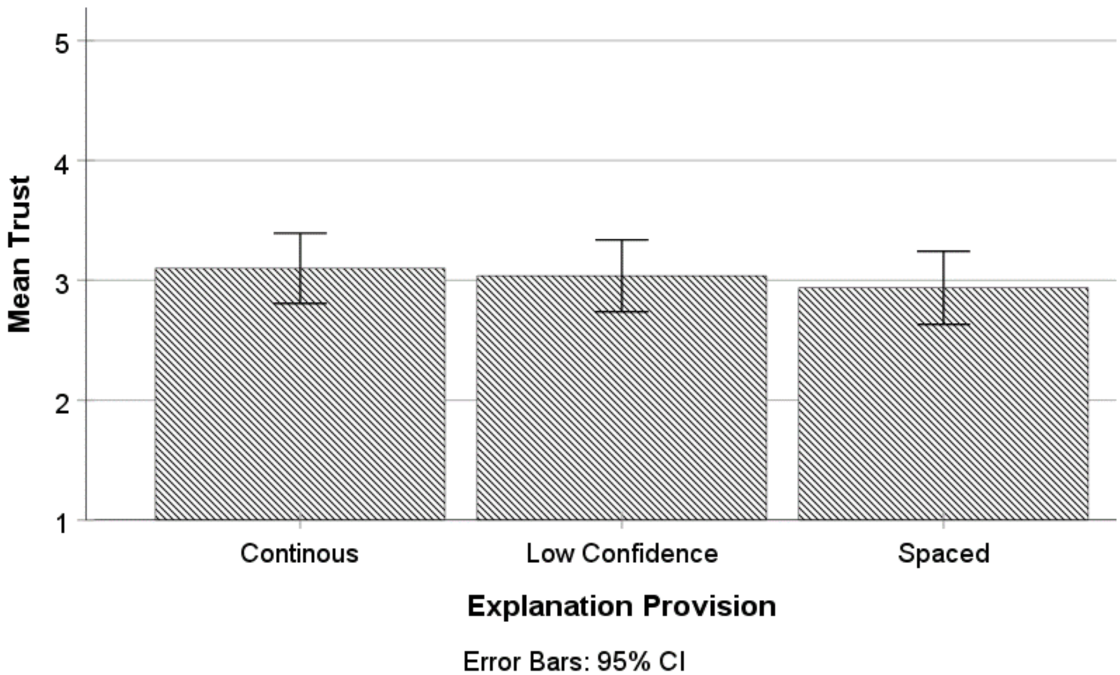


Figure 37:

Main Effect of Explanation Provision on Trust



Performance

Task Completion Time

The total time taken to complete the task was measured within the simulation. The higher value represents that they took more time to complete the task. A two-way mixed measures ANOVA was conducted to determine the effect of the explanation provision and the reliability of the AI agent on the time taken to complete the task. The interaction between the explanation provision and the reliability of the AI agent was not statistically significant, $F (2,56) = 0.68$, $p = 0.511$, partial $\eta^2 = 0.02$ (as shown in Figure 41), indicating that the impact of the explanation provision on completion time did not differ significantly between high and low reliable AIs. The total time taken to complete the task was statistically significantly different between the different conditions of explanation provisions $F (2, 56) = 4.57$, $p = 0.014$, partial $\eta^2 = 0.14$ (as shown in Figure 42). The pairwise comparison reveals that the time taken to complete the task was significantly high when the transparency and explanation were provided continuously, compared to when they were provided when AI's confidence was low, with a mean difference of 15.76 seconds, 95% CI [-14.63, 50.51], p = 0.028. Similarly, the time taken was significantly high when the transparency and explanation were provided continuously compared to when they were provided in a spaced manner, with a mean difference of 18.15 seconds, 95% CI [-20.78, 63.05], p = 0.049. The main effect of the AI agent's reliability on time taken was not statistically significant, F (1, 28) = 0.83, p = 0.370, partial η^2 = 0.03 (as shown in Figure 43), indicating no significant difference in overall completion times between the high and low reliable conditions.

Figure 38:

Effect of Explanation Provision and Reliability on Time Taken

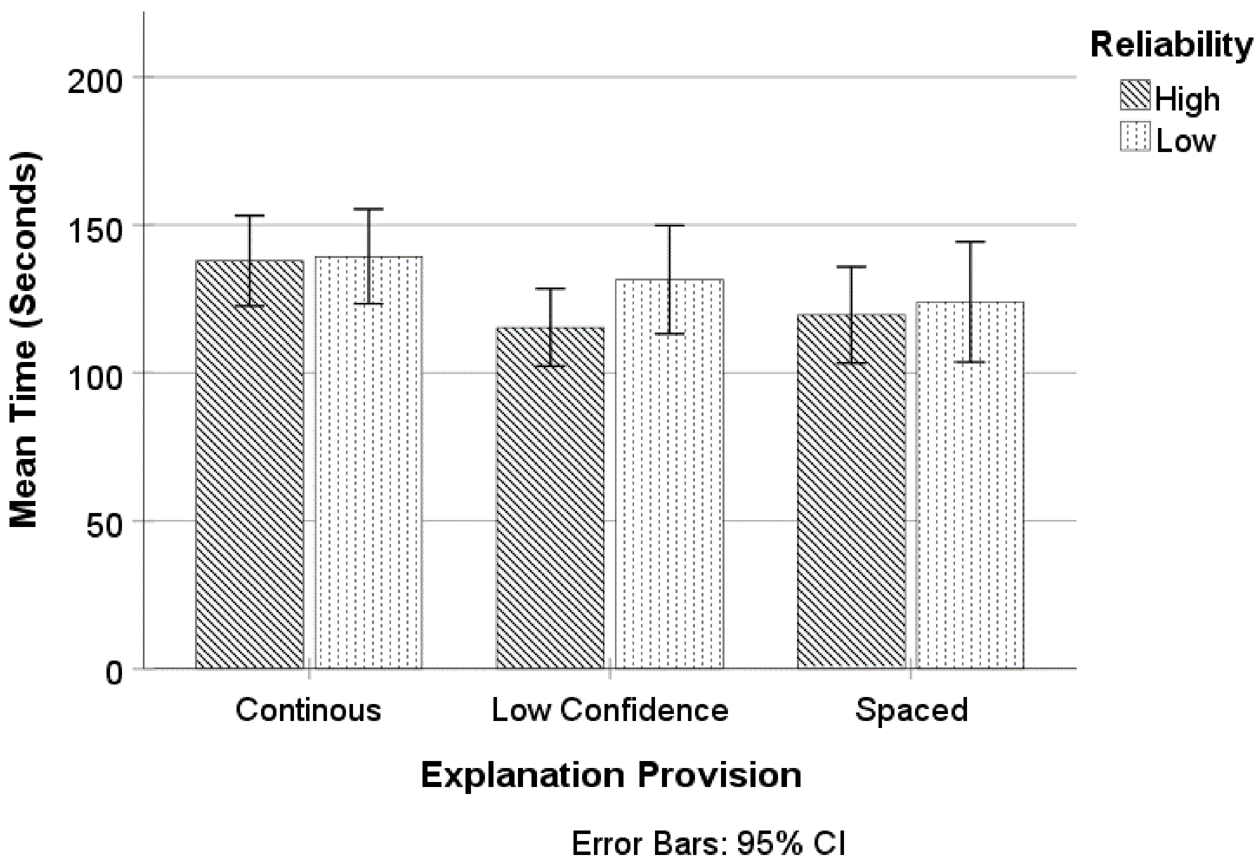


Figure 39:

Main Effect of Explanation Provision on Time Taken

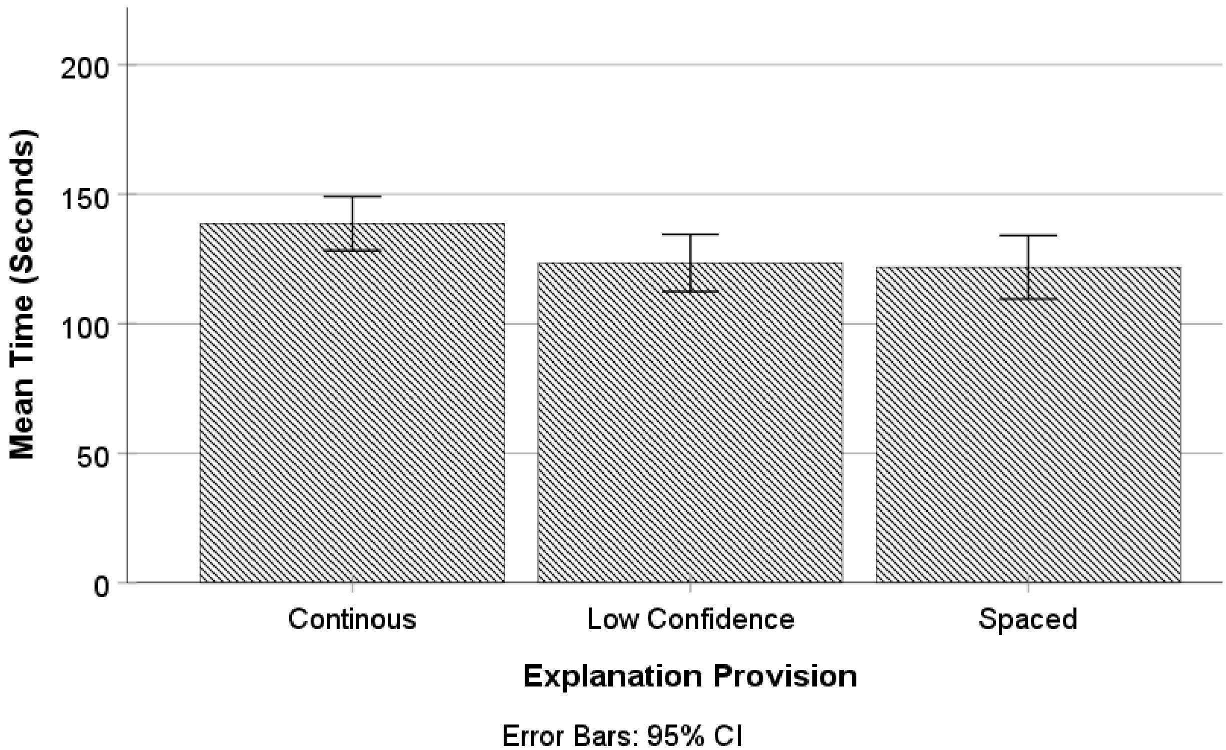
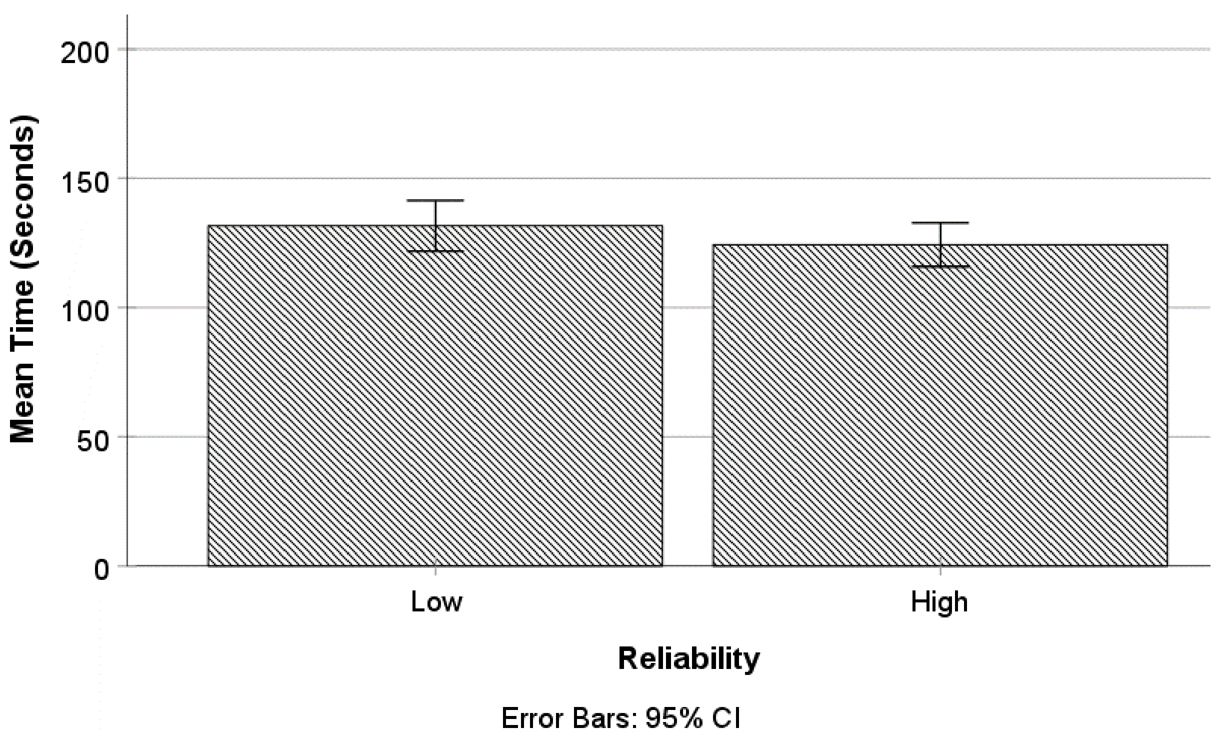


Figure 40:

Main Effect of Reliability on Time Taken



Number of Errors

The number of times the participants made a mistake while selecting the pipe fittings was counted as an error. The total number of errors was recorded within the simulation. A negative binomial regression analysis was conducted to examine the effects of the reliability of the AI agents and the explanation provision on the number of errors made. The traditional Poisson regression was not applied since the data was over dispersed with the variance exceeding the mean. The analysis revealed that the reliability of the AI agent did not have a significant effect on the number of errors made, with participants interacting with low reliable agents showing no statistically significant difference in error rates compared to those interacting with high reliable agents ($B = -0.81$, $SE = 0.56$, 95% CI [-1.70, 0.09], Wald $\chi^2 = 3.14$, $df = 1$, $p = .076$). Similarly, the explanation provision also did not have any significant difference in the number of errors made (Wald $\chi^2 = 3.46$, $df = 2$, $p = 0.177$).

Task Completion

Task completion refers to whether a participant completed the pipe layout without making any mistakes. If a mistake was made, it was considered incomplete. A binomial logistic regression was performed to investigate the effect of explanation provision, reliability of AI agent and the number of errors on the likelihood of participants completing the task. The logistic regression model was statistically significant $\chi^2(4) = 27.95$, $p < 0.001$. The model explained 76.5% (Nagelkerke R^2) of the variance in task completion and correctly classified 97.8% of cases. The number of errors showed a negative association with task completion but was not statistically significant ($p = 0.996$), indicating no significant effect of the number of errors made on task completion. The explanation provision and the reliability of the AI agent also showed no significant effect on task completion.

Hits

Hits refers to the instances where users correctly recognized and accepted a correct recommendation provided by the recommender AI. A Generalized Linear Mixed Model (GLMM) with a Poisson distribution was conducted to explore the effects of the AI agent's reliability and explanation provision on the number of times participants accepted a correct recommendation from the AI agent. The analysis of fixed effects revealed a significant effect of the model, $F(3, 86) = 18.88$, $p < 0.001$, indicating that the predictors, as a set, significantly predicted the number of hits. Within the model, reliability emerged as a significant predictor $F(1, 86) = 56.52$, $p < 0.001$(as shown in Figure 44), whereas the explanation provision did not significantly predict the number of hits $F(2, 86) = 0.06$, $p = 0.941$ (as shown in Figure 45). The expected number of hits for the low reliable condition was 0.65 (95% CI [0.54, 0.79]) times the number of hits in the high reliable condition, a statistically significant result, $p < 0.001$.

Figure 41:

Mean Hits for Reliability

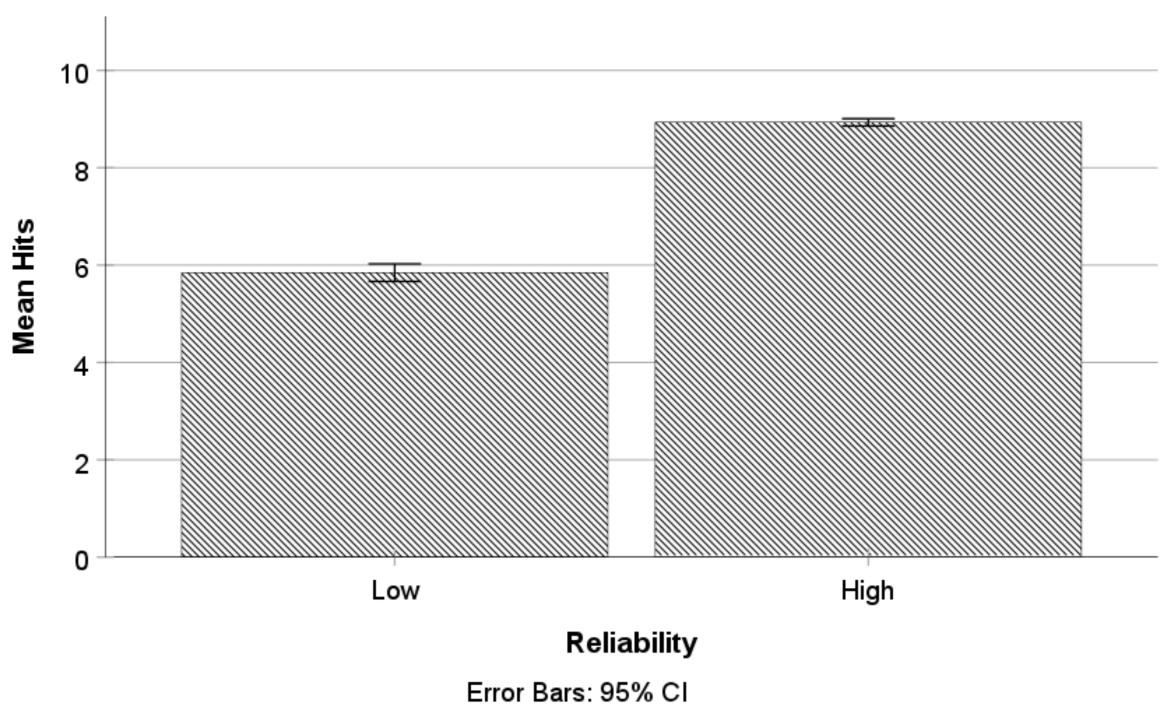
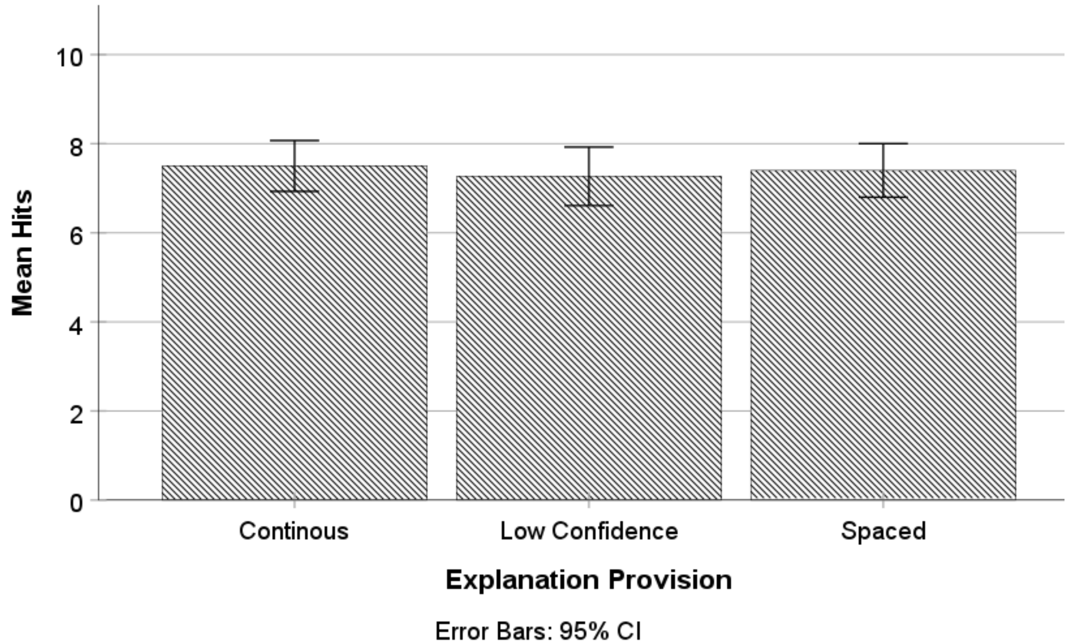


Figure 42:

Mean Hits for Explanation Provision



Correct Rejection

Correct rejection refers to the instances where the users correctly rejected the wrong recommendation provided by the AI and selected the correct pipe fitting for a particular location in the task. A GLMM with a Poisson distribution was conducted to explore the effects of the AI agent's reliability and explanation provision on the number of times participants correctly rejected an incorrect recommendation from the AI agent. The analysis of fixed effects revealed a significant effect of the model, $F(3, 86) = 16.81$, $p < 0.001$, indicating that the predictors, as a set, significantly predicted the number of correct rejections. Within the model, reliability emerged as a significant predictor, $F(1, 86) = 50.44$, $p < 0.001$ (as shown in Figure 46), whereas the explanation provision did not significantly predict the number of correct rejections, $F(2, 86) = 0.00$, $p = 0.998$ (as shown in Figure 47). For the low reliable condition, the number of correct rejections was 3.90 (95% CI

[2.66, 5.70]) times the number of correct rejections in the high reliable condition, a statistically significant result, $p < 0.001$.

Figure 43:

Mean Correct Rejections for Reliability

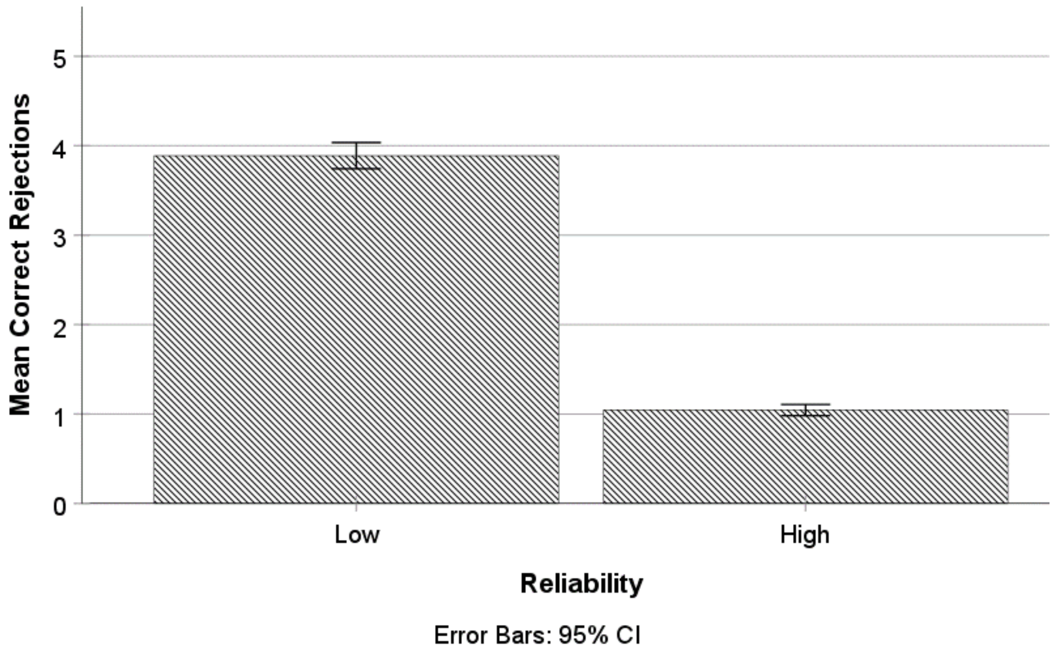
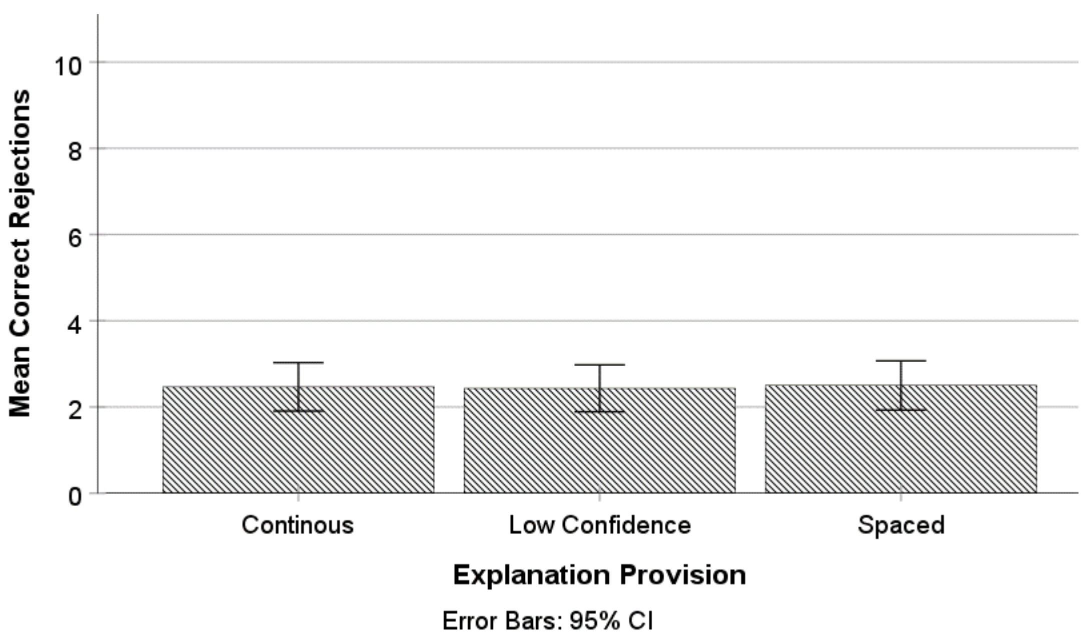


Figure 44:

Mean Correct Rejection for Explanation Provision



Workload

The workload was measured using the NASA-TLX survey after completion of each task. A two-way mixed measures ANOVA was conducted to determine the effect of the reliability of the AI agent and explanation provision on the different factors that contribute to users' workload, which includes mental demand, physical demand, temporal demand, performance, effort, and frustration level. There was no statistically significant interaction between the reliability of the AI agent and the explanation provisions, for any of these factors and for the total workload. The explanation provisions had no statistically significant effect on different sub-scales of workload except for the performance, $F (2, 56) = 3.62$, $p = 0.033$, partial $\eta^2 = 0.11$ and temporal demand sub-scales, $F (2, 56) = 4.75$, $p = 0.012$, partial $\eta^2 = 0.15$ (as shown in Figure 48). A pairwise comparison revealed that there was a significant difference in performance subscale between spaced and the low confident condition, such that the participants perceived that they performed poorer when the transparency and explanation were provided in a spaced manner compared to when provided when AI's confidence was low, with a mean difference of 3.54, 95% CI [0.24, 6.85], $p = 0.036$. Similarly, the temporal demand was high for the low confidence condition compared to the spaced condition, with a mean difference of 2.56, 95% CI [0.38, 4.73], $p = 0.017$. there was no statistically significant main effect of reliability on different sub-scales and total workload (as shown in Figure 49).

Figure 45:

Effect of Explanation Provision on Workload

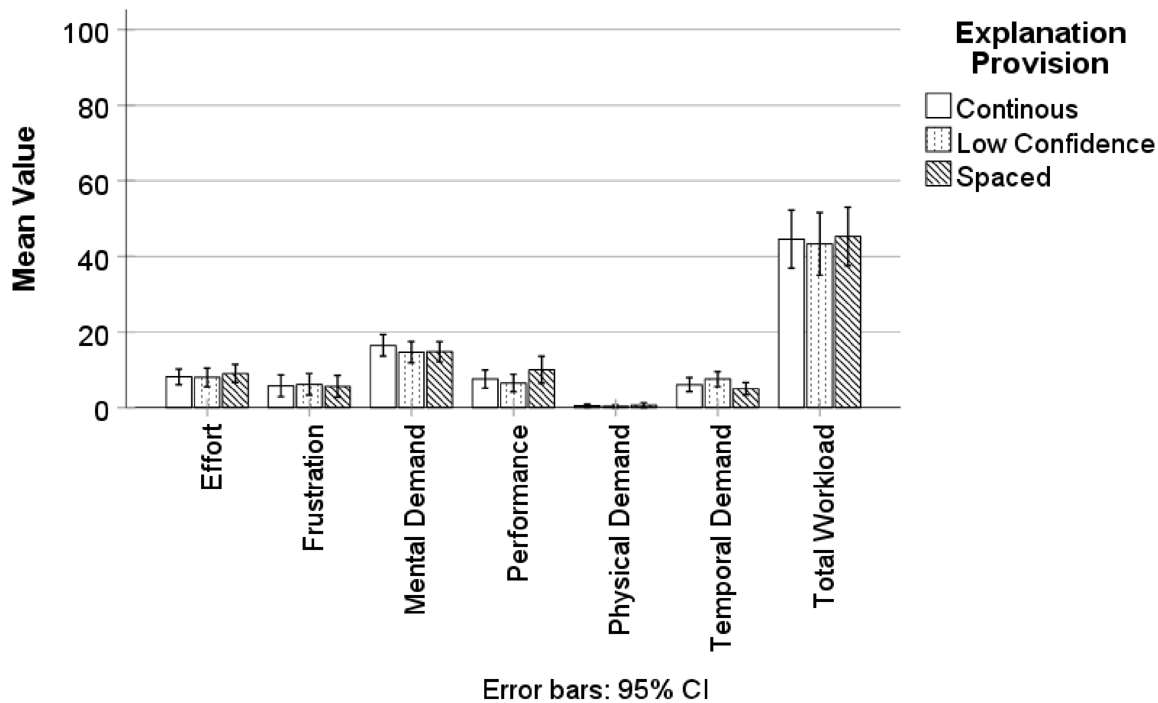
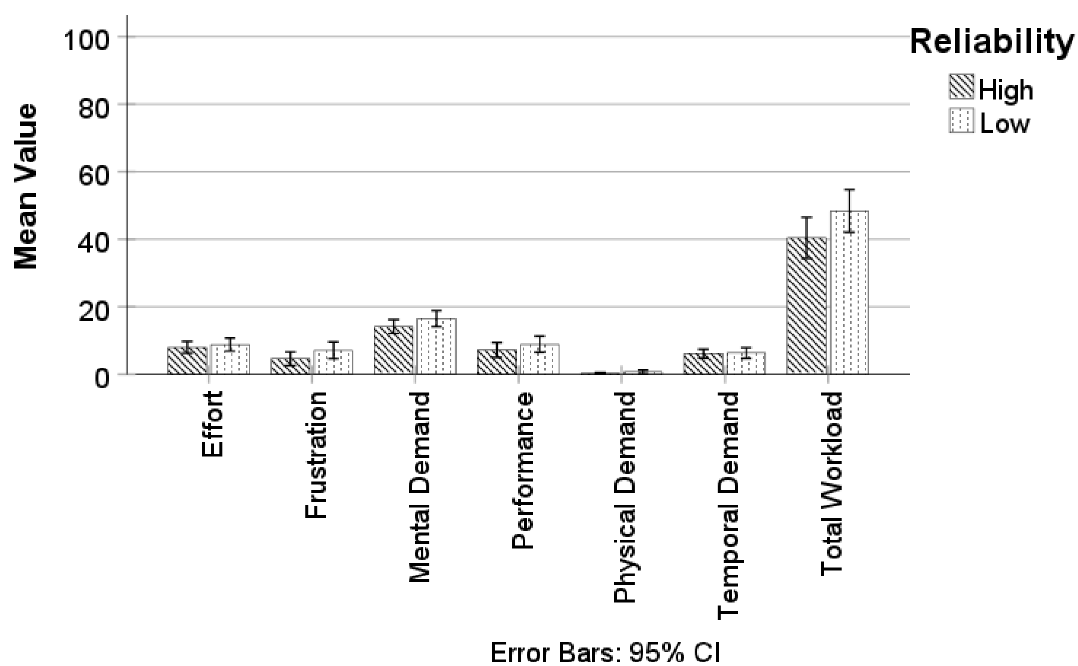


Figure 46:

Effect of Reliability on Workload



Preference

After the final task, the participants were asked to rank the AI agent they interacted with if they had a chance to work together again. Lower rank corresponds to higher preference. A Friedman's test was conducted to analyze the ranked data for the preference of agents across different levels of explanation provision. The analysis did not reveal a statistically significant difference in the ranking of agents, $\chi^2(2) = 0.07$, $p = 0.967$. The mean ranks for the continuous, when AI confidence was low, and spaced conditions were 2.03, 2.00, and 1.97, respectively, indicating very similar preferences among the participants for these conditions. The results suggest that participants did not prefer one way of providing transparency/explanation over another.

Discussion

To understand the effect of the independent variables on the overall trust, a two-way mixed ANOVA was conducted. The lack of a significant interaction effect between AI reliability and explanation provision strategy suggests that the impact of different explanation provisions on trust is consistent across different reliabilities of AI. Also, the significant main effect of reliability on trust in AI shows the importance of the reliability of the AI agent in maintaining appropriate trust, a result similar to the previous two experimental studies identified. This shows that even in the context of adaptive explanations, the reliability of AI plays an important role in trust in AI.

Supporting hypothesis one, the results reveal that there was no difference in trust between the three conditions of explanation provisions. This shows that, compared to providing transparency and explanations continuously, which enhances trust compared to not providing them (as found in study 2), the two provisions of transparency and explanation do not decrease trust. This means that even though the number of times the transparency and explanations provided decreased, the trust in AI did not decrease. This could be because it was given to the user at the right time when they needed it to maintain trust. The dynamics of situational trust are highly influenced by contextual factors; in this case, the AI provides the transparency and explanations only when the user needs it, which is similar to providing them continuously. This is an important finding that needs to be discussed further.

From the situational trust literature, it is understood that the cognitive process of the users to weigh the trustworthiness of the system during the interaction can be altered by the ways by which the tasks are presented (Hoff & Bashir, 2015). Similarly, the organizational setting affects the trust the user has in the AI in such a way that the trusting behavior is affected by when the recommendations are presented (Madhavan & Wiegmann, 2005). In this experiment, the two ways that we presented the transparency and explanation were effective in making the users trust the recommender AI in line with when the information was provided continuously. Situational trust is influenced by different internal and external factors such as the user's self-confidence, task expertise, type of automation, the complexity of automation etc. (Hoff & Bashir, 2015). The two additional provisions of presenting the information in this experiment might have helped the users maintain the appropriate trust due to the complexity of the task and the operators' decision freedom. These provision's environment might have also helped the users to assess automation performance relative to the manual performance.

One of the provisions of presenting transparency and explanation was providing them only when the recommender AI's confidence was low. This means the additional information on how and why the AI provided a particular recommendation, was presented only when the AI was unsure about its recommendation. The non-reduction in trust in AI during this condition, compared to providing the them continuously, shows that the need for transparency and explanation to maintain appropriate trust is not just continuous, but also can be achieved by providing them when the AI's confidence is low. This could be because the user analyzed the situation further and made the decision when they saw the AI was not confident about the recommendation. The users could have been more considerate of the recommender AI when they were provided with transparency and explanation even though the AI was unsure about the recommendation. Also, the users might have appreciated the recommender AI for being transparent about the fact that it is unsure about the recommendation provided. This thought is supported by the findings from a previous study that looked into the effect of providing an explanation when a decision aid made errors in a military context, which found that, this provision prevents distrust (Dzindolet et al., 2003). This approach also aligns with the concept that users value explanations more when they contribute directly to understanding AI's decision-making process in uncertain situations, fostering a deeper trust in the system's capabilities (Kizilcec, 2016). This needs to be researched further to investigate effective ways of representing AI’s confidence in providing recommendations that could maintain the appropriate trust level in AI.

Another way of providing explanation and transparency that was studied in this experiment was providing them with the first, middle and last steps. This method of provision also did not reduce the trust level of the users that they had in the recommender AI. Providing transparency and explanation about AI’s decision to provide a particular recommendation in the initial step might have helped the users understand the process that the AI goes through and the rationale of AI for providing a particular recommendation. After the first step, interacting with the AI that provides just the recommendation might change the understanding of the underlying process and the rationale of providing a recommendation, which could have been corrected by providing transparency and explanation in the middle of the task and the final step of the task. The trust did not decrease compared to providing the transparency and explanation continuously, which shows that reinforcing users with the process and rationale of providing recommendations by the AI in the initial, middle, and last steps could maintain appropriate trust in a similar manner when they were provided continuously. This approach might also help users develop a mental model of how AI makes decisions, which is an important aspect of maintaining trust. This result aligns with the previous literature that found that the AI explanation is not always desirable (Jiang et al., 2022). This needs to be investigated further for tasks that are longer and more complex since the task in this experiment did not involve a large number of steps. Similarly, the provision of providing an explanation could be explored further to see the impact of trust, such as providing users the freedom to activate the required information when they need it.

The total time taken to complete the tasks was measured to evaluate the participants' performance while completing the tasks in different explanation provisions. Supporting hypothesis two, participants required significantly more time to complete the task when the AI provided continuous explanation and transparency compared to when explanations were given when the AI's confidence was low or at spaced intervals. These findings suggest that although continuous explanation and transparency may improve comprehension and confidence in AI recommendations, it may also result in information overload, where the constant presentation of data delays the ability to make quick decisions and complete tasks. Participants might have to spend additional time processing the information provided by the AI agent, which adversely affects their speed of task completion (Helldin, 2014; Zuk & Carpendale, 2007). On the other hand, the low confidence and the spaced conditions provide a more optimal balance between providing necessary guidance and maintaining task efficiency. This allows participants to receive important information at decision points where uncertainty is likely highest, supporting faster and potentially more accurate decision-making without the constant cognitive load of processing information.

The findings from the negative binomial regression analysis focusing on the effect of reliability and explanation provision on the number of errors made found that both the reliability and the explanation provision did not significantly influence the number of errors, supporting hypothesis two. This result of the reliability variable is similar to study 2, which did not find a significant difference between the two reliable conditions. Similarly, the number of errors did not vary between the three conditions of explanation provision. This is an important finding since there was no decrease in the number of errors even when the number of times the transparency and explanation were provided. This could be because, for this particular task, the possibility of making errors was mainly when the AI's confidence was low. Providing transparency and explanation at the point where the AI was unsure might have made the participant think more about the situation rather than blindly believing the AI, thereby making the correct decision. Similarly, the condition where the transparency and explanation was provided in the initial step, middle step and final step, might have helped the participants to make a mental model of how AI comes up with the recommendations and could have made the right decision. The important takeaway from this is that providing explanations and transparency non-continuously does not increase error rates. However, this needs to be explored further since the experiment used in the study could be simple enough to make many mistakes. Different complexities of tasks should be studied before generalizing this result. Also, future research could investigate the effects of providing explanations and transparency at the decision points which are more prone to errors.

The perceived workload while completing the procedural task was measured using NASA-TLX survey. Supporting hypothesis three, the results indicate no significant interaction between the reliability of AI and the explanation provisions in relation to multiple components of the workload suggesting a uniformity in perceived workload across the three levels of explanation provision and the two levels of reliability of the recommender AI. The temporal demand subscale revealed that it was higher when the transparency and explanation was provided when the recommender AI's confidence was low compared to providing them in a spaced manner. This could be because when the users realized that the AI's confidence was low, they had to adjust their mental model of the AI's recommendations, which necessitated maintaining a higher level of alertness, potentially leading to an increased temporal demand. Whereas for the spaced condition, even though the explanation provided the confidence information, they might have mostly received it when the AI's confidence level was higher. However, the perceived performance was higher for the low confidence condition compared to the spaced condition. This could be because, in the low confidence condition, since they knew that AI's confidence was low and subsequently the recommendation could be wrong, the users might have spent more mental resources to make sure the component they selected was correct, leading to an increased perceived performance compared to the spaced condition.

The results of the analysis of the number of hits and correct rejections were consistent with study 2, such that the low reliable condition had fewer hits and more correct rejections compared to the high reliable condition. This is because the number of correct recommendations provided by the AI was lower for the low reliable condition compared to the high reliable one. From Figures 44 and 46, it can be seen that the mean hits were around 6 and 9 for the low reliable and high reliable conditions, whereas the correct rejection was around 4 and 1 for the low and high reliable conditions, respectively. This shows that the users perceived reliability of the AI was in line with the actual reliability of the AI since there were 6 and 9 correct recommendations and 4 and 1 wrong recommendations for the low and high reliable conditions respectively. This shows that the users had appropriate trust in AI and highlights the importance of calibrating human trust to an appropriate level, which is necessary for the reliability of the AI that they are interacting with; if not, that will lead to disuse and misuse of the AI system, respectively (Parasuraman & Manzey, 2010).

The ranking of the AI agents after completing the task did not find any significant difference between the three conditions of explanation provision showing that the users ranked the spaced condition and the low confident condition similar to that of providing them continuously. This result aligns with the overall trust result where there was no significant difference between the three levels of explanation provision. The users trusted the three recommender AIs in a similar manner, and thus, there was no significant difference in the ranking of these different recommender AIs.

Conclusion

This study explored the effects of the reliability of the AI agents and the provision of explainability and transparency of the AI's recommendation on users' trust in AI, performance and workload while completing a procedural task. The study used a 2*3 mixed-subject experimental design, to understand these effects.

The findings from this study revealed that providing transparency and explanation in a continuous manner is not necessary to maintain appropriate trust. Providing them at the right time the user needs is an important aspect of transparency and explanation. Specifically, when AI's confidence is low, intermittently reminding the user about how the AI makes decisions provides the users enough confidence to trust in the AI as similar as they trust when the transparency and explanation is provided continuously. In addition to that, the main advantage is that they don't have to spend as much time as they spend when transparency and explanations are provided continuously to complete the task. This study also shows that, while the AI provides transparency and explanation only when needed, users can complete tasks faster without an increase in total workload and the number of errors while maintaining the trust they have in the recommender AI, similar to when that is provided continuously.

This study comes with a few limitations that need to be considered before generalizing the result. The task selected was a pipe fitting task and was simulated as a user completing the task while wearing an AR glass where the AI provided recommendations on the AR module. This could be validated by conducting experiments with other procedural tasks and using an AR glass instead of a computer simulation. Also, the task selected was a smaller one, and the difficulty level of the task was not too high, as seen from the number of errors found in the study data. This warrants investigation of more complex tasks to see if there is any variation in the findings of this study on the effect of trust, performance, and workload due to the transparency and explanation provision.

The findings from this experimental study will help the designers of the AI systems to decide on when exactly to provide the explanation and transparency while considering factors such as the reliability of the AI, trust the humans have in the AI, the performance and workload the users have while using the AI.

CHAPTER 6

CONCLUSION

The overarching goal of this dissertation was to characterize human performance and trust in AI while completing procedural tasks in the AEC domain with the help of AI and AR technologies. This research aimed to formulate design principles that improve the interaction of humans with these technologies, focusing on explainability and transparency to maintain appropriate trust in AI. The objectives included investigating how AR technology is used in the AEC domain for remote collaboration, understanding the effect of AI agent reliability and task complexity on trust in AI, evaluating the impact of transparency and explainability on trust, and assessing the effectiveness of adaptive explanations in maintaining appropriate trust in AI.

The systematic literature review was conducted to learn about the application of AR for remote collaboration in the AEC industry, identifying the different devices being used, the important applications, considerations, and limitations of using AR technologies for remote collaboration. The development of a POMDP model to simulate trust in AI, followed by an experimental study to validate it provided valuable insights into trust dynamics during procedural task completion. The findings from the first experimental study showed that even though the complexity of the task did not affect the trust the users have in AI, the AI agent reliability significantly influences trust, workload, and performance, highlighting the importance of designing highly reliable AI systems to reduce cognitive load and to improve the efficiency of the task being completed.

The final two experimental studies revealed the important role of transparency and explainability and the importance of providing them at the right time in building trust in AI. The second experimental study showed that while users preferred AI agents that provided transparency and explanations, with an increase in trust compared to not providing them, these features increased task completion time without adding to the perceived workload. This indicates that the users had spent more time engaging with the AI and the transparency and explanation provided, which might have enhanced their trust in the AI. The final experimental study found that providing transparency and explanation at the right time, particularly when AI confidence is low and during the first middle and last steps, can maintain appropriate trust in AI without significantly increasing task completion time, errors, and workload. This shows that to maintain appropriate trust, transparency and explanations need not be provided at every step. Instead, providing them strategically at the points where the user might use them in an adaptive manner improved the performance without affecting trust.

The findings of this dissertation have significant practical implications for designers of AI systems and stakeholders in the AEC industry. The research highlights the importance of integrating transparency and explainability features along with the recommendations to build and maintain appropriate trust in AI. This is important for ensuring that users are using the AI technology as it is intended, rather than misusing and disusing, which can enhance overall task performance and efficiency. Additionally, the results suggest that designing AI systems to provide transparency and explanations in an adaptive way or only when necessary, can maintain appropriate trust without increasing the cognitive load on users. This could be helpful to maintain appropriate trust especially when the AI provides incorrect recommendations in between. Providing transparency and explanations could help understand the user why AI provided wrong recommendations and could make informed decisions.

The application of the integration of AR and AI in the AEC industry is not limited to the plumbing task used in the experiment but could be extended to activities such as site inspection, safety inspection, measurement and layout, quality control, resource management, training etc.

Scenarios where the user takes more time to figure out a problem could be handled with the advantage of time and cost if the recommender AI is well trained with the previous data of similar problems and is presented to the user in an effective manner. The data associated with each activities needs to be collected extensively to train the AI model to implement the recommender system with higher reliability. While implementing such systems in the AEC industry, it is important to make sure the privacy of the users is protected. This could be done by providing users the freedom to control the amount of data collected or by collecting only the absolutely necessary data. Also, the data collected could be anonymized and encrypted. The safety while using these technologies in the construction site is another important aspect. While wearing the AR devise, the users could be distracted and might get overloaded with information leading to decreased situation awareness. The awareness of the real world could also be affected by the vision being obstructed due to a lot of information presented on the AR glasses and could lead to safety incidents. This could be avoided my making sure the interface is not overloaded with information and is designed in a user-friendly manner with the freedom of toggling the information on and off.

This dissertation acknowledges several limitations that should be considered before generalizing the findings. The experimental tasks were completed on a computer simulation that emulated a user wearing an AR glass while completing a pipe fitting task, which may not fully capture the complexity and variability of tasks in the AEC domain. Future research could validate these findings with different tasks in the AEC industry and could use actual AR glasses instead of simulations. Additionally, the tasks used in the experiments were relatively simple, which may have influenced the observed trust dynamics and performance outcomes. Future research could look into applications of these techniques to maintain appropriate trust on tasks with higher complexities to see the effect on performance, workload and trust in AI. Further, different ways of providing adaptive explanations could be studied to understand if they can help maintain appropriate trust. This could include providing explanation and transparency when the user stops following the recommendation provided by the AI or could consider providing them when the subtask is very confusing or at a very important decision point where most users make mistakes.

In conclusion, this dissertation provides an understanding of the factors influencing trust in AI when integrated with AR for procedural tasks in the AEC industry. The insights gained from this research can guide the design and implementation of AI systems in the AEC industry to maintain appropriate trust in AI, enhance task performance, and reduce workload, ultimately contributing to more efficient and effective use of these emerging technologies.

APPENDICES

Appendix A

Adapted Propensity to Trust in Technology

(Schneider et al., 2017)

<u>Instruction:</u> For the below-listed items, please read each statement carefully. Using multiple choice answers from strongly agree to strongly disagree, select the answer that most accurately describes your feelings. An automated agent can be anything from spell check on Word to Siri or Alexa.

Q1 Generally, I trust automated agents.

o Strongly agree (1)
o Somewhat agree (2)
o Neither agree nor disagree (3)
o Somewhat disagree (4)
o Strongly disagree (5)

Q2 Automated agents help me solve many problems.

o Strongly agree (1)
o Somewhat agree (2)
o Neither agree nor disagree (3)
o Somewhat disagree (4)
o Strongly disagree (5)

Q3 I think it's a good idea to rely on automated agents for help.

o Strongly agree (1)
o Somewhat agree (2)
o Neither agree nor disagree (3)
o Somewhat disagree (4)
o Strongly disagree (5)

Q4 I don't trust the information I get from automated agents.

o Strongly DISAGREE (1)
o Somewhat DISAGREE (2)
o Neither agree nor disagree (3)
o Somewhat AGREE (4)
o Strongly AGREE (5)

Q5 Automated agents are reliable.

o Strongly agree (1)
o Somewhat agree (2)
o Neither agree nor disagree (3)
o Somewhat disagree (4)
o Strongly disagree (5)

Q6 I rely on automated agents.

o Strongly agree (1)
o Somewhat agree (2)
o Neither agree nor disagree (3)
o Somewhat disagree (4)
o Strongly disagree (5)

Appendix B

Trust in Automated Systems
(Jian et al., 2000)

Instructions: Below is a list of statements for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please select the option which best describes your feeling or your impression using the 7-point scale ranging from 1 (not at all) to 7 (extremely).

1. The system is deceptive.

1	2	3	4	5	6	7

2. The system behaves in an underhanded manner.

1	2	3	4	5	6	7

3. I am suspicious of the system's intent, action, and outputs.

1	2	3	4	5	6	7

4. I am wary of the system.

1	2	3	4	5	6	7

5. The system's actions will have a harmful or injurious outcome.

1	2	3	4	5	6	7

6. I am confident in the system.

1	2	3	4	5	6	7

7. The system provides security.

1	2	3	4	5	6	7

8. The system has integrity.

1	2	3	4	5	6	7

9. The system is dependable.

1	2	3	4	5	6	7

10. The system is reliable.

1	2	3	4	5	6	7

11. I can trust the system.

1	2	3	4	5	6	7

12. I am familiar with the system.

1	2	3	4	5	6	7

Appendix C

Trust scale for XAI

(Hoffman et al., 2018)

1. I am confident in the recommender AI. I feel that it works well.

5	4	3	2	1
I agree strongly	I agree somewhat	I'm neutral about it	I disagree somewhat	I disagree strongly

2. The outputs of the recommender AI are very predictable.

5	4	3	2	1
I agree strongly	I agree somewhat	I'm neutral about it	I disagree somewhat	I disagree strongly

3. The tool is very reliable. I can count on it to be correct all the time.

5	4	3	2	1
I agree strongly	I agree somewhat	I'm neutral about it	I disagree somewhat	I disagree strongly

4. I feel safe that when I rely on the recommender AI I will get the right answers.

5	4	3	2	1
I agree strongly	I agree somewhat	I'm neutral about it	I disagree somewhat	I disagree strongly

5. The recommender AI is efficient in that it works very quickly.

5	4	3	2	1
I agree strongly	I agree somewhat	I'm neutral about it	I disagree somewhat	I disagree strongly

6. I am wary of the recommender AI.

5	4	3	2	1
I agree strongly	I agree somewhat	I’m neutral about it	I disagree somewhat	I disagree strongly

7. The recommender AI can perform the task better than a novice human user.

5	4	3	2	1
I agree strongly	I agree somewhat	I’m neutral about it	I disagree somewhat	I disagree strongly

8. I like using the system for decision making.

5	4	3	2	1
I agree strongly	I agree somewhat	I’m neutral about it	I disagree somewhat	I disagree strongly

Appendix D

NASA Task Load Index

(Hart & Staveland, 1988)

Task Questionnaire – Part 1

Click on each scale at the point that best indicates your experience of the task

Mental Demand — How mentally demanding was the task?

Low — High

Physical Demand — How physically demanding was the task?

Low — High

Temporal Demand — How hurried or rushed was the pace of the task?

Low — High

Performance — How successful were you in accomplishing what you were asked to do?

Good — Poor

Effort — How hard did you have to work to accomplish your level of performance?

Low — High

Frustration

How insecure, discouraged, irritated, stressed and annoyed were you?

Low High

Task Questionnaire – Part 2

On each of the following 15 screens, click on the scale title that represents the more important contributor to workload for the task.

Click on the factor that represents the more important contribution to workload for the task.

Frustration or Mental Demand

Click on the factor that represents the more important contribution to workload for the task.

Temporal Demand or Frustration

Click on the factor that represents the more important contribution to workload for the task.

Effort or Performance

Click on the factor that represents the more important contribution to workload for the task.

Temporal Demand or Mental demand

Click on the factor that represents the more important contribution to workload for the task.

Temporal Demand or Effort

Click on the factor that represents the more important contribution to workload for the task.

Mental or Physical Demand

Click on the factor that represents the more important contribution to workload for the task.

Effort or Physical Demand

Click on the factor that represents the more important contribution to workload for the task.

Mental Demand or Effort

Click on the factor that represents the more important contribution to workload for the task.

Physical Demand	or	Frustration

Click on the factor that represents the more important contribution to workload for the task.

Performance	or	Frustration

Click on the factor that represents the more important contribution to workload for the task.

Physical Demand	or	Performance

Click on the factor that represents the more important contribution to workload for the task.

Performance	or	Temporal Demand

Click on the factor that represents the more important contribution to workload for the task.

Performance	or	Mental Demand

Click on the factor that represents the more important contribution to workload for the task.

Physical Demand	or	Temporal Demand

Click on the factor that represents the more important contribution to workload for the task.

Frustration	or	Effort

REFERENCES

Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*. https://doi.org/10.1016/J.JOBE.2021.103299

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Akanmu, A. A., Anumba, C. J., & Ogunseiju, O. O. (2021). Towards next generation cyber-physical systems and digital twins for construction. Journal of Information Technology in Construction, 26, 505–525. https://doi.org/10.36680/j.itcon.2021.027

Al, P. (2023). (E)-Trust and its function: Why we shouldn't apply trust and trustworthiness to human–AI relations. *Journal of Applied Philosophy*, *40*(1), 95–108. https://doi.org/10.1111/JAPP.12613

Alam, L., & Mueller, S. (2021). Examining the effect of explanation on satisfaction and trust in AI diagnostic systems. *BMC Medical Informatics and Decision Making*, *21*(1), 1–15. https://doi.org/10.1186/S12911-021-01542-6

Alem, L., & Li, J. (2011). A study of gestures in a video-mediated collaborative assembly task. *Advances in Human-Computer Interaction*, *2011,* 1–7. https://doi.org/10.1155/2011/987830

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, *99*, 101805. https://doi.org/10.1016/J.INFFUS.2023.101805

Alizadehsalehi, S., Hadavi, A., & Huang, J. C. (2020). From BIM to extended reality in AEC industry. *Automation in Construction*, *116*, 103254. https://doi.org/10.1016/J.AUTCON.2020.103254

Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2021). Does Explainable Artificial Intelligence Improve Human Decision-Making? *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6618–6626. https://doi.org/10.1609/AAAI.V35I8.16819

Ammari, K. E. , & Hammad, A. (2019). Remote interactive collaboration in facilities management using BIM-based mixed reality. *Automation in Construction (Vol. 107, p. 102940)*. https://doi.org/10.1016/j.autcon.2019.102940.

Andrada, G., Clowes, R. W., & Smart, P. R. (2023). Varieties of transparency: Exploring agency within AI systems. *AI and Society*, *38*(4), 1321–1331. https://doi.org/10.1007/S00146-021-01326-6

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., & Zhang, J. (2020). Artificial intelligence in advanced manufacturing: Current status and future outlook. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, *142*(11). https://doi.org/10.1115/1.4047855/1085487

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/J.INFFUS.2019.12.012

Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: Focus on clinicians. *Journal of Medical Internet Research, 22(*6*):E15154*. https://doi.org/10.2196/15154

Aschenbrenner, D., Rojkov, M., Leutert, F., Verlinden, J., Lukosch, S., Latoschik, M. E., & Schilling, K. (2018). Comparing different augmented reality support applications for cooperative repair of an industrial robot. *Adjunct Proceedings - 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2018*, 69–74. https://doi.org/10.1109/ISMAR-ADJUNCT.2018.00036

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2022). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction,* 1–16. https://doi.org/10.1080/10447318.2022.2138826

Barlish, K., & Sullivan, K. (2012). How to measure the benefits of BIM - A case study approach. *Automation in Construction*, *24*, 149–159. https://doi.org/10.1016/J.AUTCON.2012.02.008

Beauxis-Aussalet, E., Behrisch, M., Borgo, R., Chau, D. H., Collins, C., Ebert, D., El-Assady, M., Endert, A., Keim, D. A., Kohlhammer, J., Oelke, D., Peltonen, J., Riveiro, M., Schreck, T., Strobelt, H., Van Wijk, J. J., & Rhyne, T. M. (2021). The role of interactive visualization in fostering trust in AI. *IEEE Computer Graphics and Applications*, *41*(6), 7–12. https://doi.org/10.1109/MCG.2021.3107875

Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, *35*(2), 530–549. https://doi.org/10.1108/JEIM-06-2020-0233

Bhanu, A., Sharma, H., Piratla, K. &, & Chalil Madathil, K. (2022). Application of augmented reality for remote collaborative work in architecture, engineering, and construction – A systematic review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1829–1833. https://doi.org/10.1177/1071181322661167

Bowman, D. A., Gabbard, J., Auerbach, D., Roofigari-Esfahan, N., Britt, K., Ilo, C. I., & Adapa, K. (2022). BuildAR: A proof-of-concept prototype of intelligent augmented reality in construction. *Proceedings - 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW 2022*, 508–512. https://doi.org/10.1109/VRW55335.2022.00113

Bronack, S. C. (2011). The role of immersive media in online education. *Journal of Continuing Higher Education*, *59*(2), 113–117. https://doi.org/10.1080/07377363.2011.583186

Cahour, B., & Forzy, J. F. (2009). Does projection into use improve trust and exploration? An example with a cruise control system. *Safety Science*, *47*(9), 1260–1270. https://doi.org/10.1016/J.SSCI.2009.03.015

Cain, B. (2007). *A review of the mental workload literature*. Toronto, Canada: Defence Research And Development.

Capobianco, E. (2022). High-dimensional role of AI and machine learning in cancer research. *British Journal of Cancer*, *126*(4), 523–532. https://doi.org/10.1038/s41416-021-01689-z

Chantziaras, G., Triantafyllidis, A., Papaprodromou, A., Chatzikonstantinou, I., Giakoumis, D., Tsakiris, A., Votis, K., & Tzovaras, D. (2021). An augmented reality-based remote collaboration platform for worker assistance. *Pattern Recognition. ICPR International*

Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII (pp. 404-416). Springer International Publishing. https://doi.org/10.1007/978-3-030-68787-8_30

Chavaillaz, A., Wastell, D., & Sauer, J. (2016). System reliability, performance and trust in adaptable automation. *Applied Ergonomics*, *52*, 333–342. https://doi.org/10.1016/J.APERGO.2015.07.012

Chen, K., & Xue, F. (2022). The renaissance of augmented reality in construction: History, present status and future directions. *Smart and Sustainable Built Environment*, 11(3), 575–592. https://doi.org/10.1108/SASBE-08-2020-0124

Chen, M., Nikolaidis, S., Soh, H., Hsu, D., & Srinivasa, S. (2018). Planning with trust for human-robot collaboration. *ACM/IEEE International Conference on Human-Robot Interaction*, 307–315. https://doi.org/10.1145/3171221.3171264

Chen, P. H. C., Gadepalli, K., MacDonald, R., Liu, Y., Kadowaki, S., Nagpal, K., Kohlberger, T., Dean, J., Corrado, G. S., Hipp, J. D., Mermel, C. H., & Stumpe, M. C. (2019). An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nature Medicine*, *25*(9), 1453–1457. https://doi.org/10.1038/s41591-019-0539-7

Chin, C. H., Wong, W. P. M., Cham, T. H., Thong, J. Z., & Ling, J. P. W. (2024). Exploring the usage intention of AI-powered devices in smart homes among millennials and zillennials: the moderating role of trust. *Young Consumers*, *25*(1), 1–27. https://doi.org/10.1108/YC-05-2023-1752

Cisterna, D., Von Heyl, J., Alarcón, D. M., Herrera, R. F., & Alarcón, L. F. (2018). Application of Social Network Analysis in Lean and Infrastructure Projects. *Proceedings of the 26th Annual Conference of the International Group for Lean Construction: Evolving Lean*

Construction Towards Mature Production Management Across Cultures and Frontiers, *1*, 412–421. https://doi.org/10.24928/2018/0483

Danker, F. , & Jones, O. (2014). Combining Augmented Reality and Building Information Modelling - An industry perspective on applications and future directions. *IOP Conference Series: Earth and Environmental Science, Volume 397, Simulation for Sustainable Built Environment 28–30 November 2019, New Cairo, Egypt.* https://doi.org/10.1088/1755-1315/397/1/012012

Daronnat, S., Azzopardi, L., Halvey, M., & Dubiel, M. (2020). Impact of agent reliability and predictability on trust in real time human-agent collaboration. *Proceedings of the 8th International Conference on Human-Agent Interaction*, 131–139. https://doi.org/10.1145/3406499.3415063

De Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231. https://doi.org/10.1177/1555343411410160

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*, *58*(6), 719–735. https://doi.org/10.1016/S1071-5819(03)00039-9

Degas, A., Islam, M. R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M. U., Begum, S., Rahman, M. A., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G., Babiloni, F., & Aricó, P. (2022). A survey on artificial intelligence (AI) and explainable AI in air traffic management: Current trends and development with future research trajectory. *Applied Sciences*, *12*(3), 1295. https://doi.org/10.3390/APP12031295

Demian, P. , & Walters, D. (2014). The advantages of information management through building information modelling. *Construction Management and Economics*, *32(12),* 1153–1165.

Devagiri, J. S., Paheding, S., Niyaz, Q., Yang, X., & Smith, S. (2022). Augmented reality and artificial intelligence in industry: trends, tools, and future challenges. *Expert Systems with Applications*, *207*, 118002. https://doi.org/10.1016/J.ESWA.2022.118002

Dong, S. , Behzadan, A. H. , Chen, F. , & Kamat, V. R. (2013). Collaborative visualization of engineering processes using tabletop augmented reality. *Advances in Engineering Software*. *55,* 45-55. https://doi.org/10.1016/j.advengsoft.2012.09.001

Dosilovic, F. K., Brcic, M., & Hlupic, N. (2018). Explainable artificial intelligence: A survey. *41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, 210–215. https://doi.org/10.23919/MIPRO.2018.8400040

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Emaminejad, N., Maria North, A., & Akhavian, R. (2021). Trust in AI and implications for AEC research: A literature analysis. *Computing in Civil Engineering 2021*, 295–303. https://doi.org/10.1061/9780784483893.037

Fernández del Amo, I., Erkoyuncu, J., Vrabič, R., Frayssinet, R., Vazquez Reynel, C., & Roy, R. (2020). Structured authoring for AR-based communication to enhance efficiency in remote diagnosis for complex equipment. *Advanced Engineering Informatics*, *45,* 101096. https://doi.org/10.1016/J.AEI.2020.101096

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. *Proceedings of the 2007 International Symposium on Collaborative Technologies and Systems*, 106–114. https://doi.org/10.1109/CTS.2007.4621745

Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607. https://doi.org/10.1016/j.chb.2020.106607

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2011). Integrated sequential as-built and as-planned representation with D4AR tools in support of decision-making tasks in the AEC/FM industry. *Journal of Construction Engineering and Management*, *137*(12), 1099–1116. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000371

Grassmann, W. K. (1983). Markov Modelling. *Proceedings of the 15th conference on Winter Simulation - Volume 2* (pp. 613–619).

Hala, N., Christian, S., Makram, B. H., & Harald, U. (2020). A comprehensive map for integrating augmented reality during the construction phase. *Proceedings of the Creative Construction E-Conference 2020*, 56–64. https://doi.org/10.3311/CCC2020-069

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Havard, V., Baudry, D., Louis, A., & Mazari, B. (2015). Augmented reality maintenance demonstrator and associated modelling. *2015 IEEE Virtual Reality Conference, VR 2015 - Proceedings*, 329–330. https://doi.org/10.1109/VR.2015.7223429

Henderson, S. J., & Feiner, S. K. (2011). Augmented reality in the psychomotor phase of a procedural task. *10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011*, 191–200. https://doi.org/10.1109/ISMAR.2011.6092386

Herm, L.-V. (2023). Impact of explainable ai on cognitive load: Insights from an empirical study. *ArXiv Preprint*. https://arxiv.org/abs/2304.08861v1

Hincapie, M., Caponio, A., Rios, H., & Gonzalez Mendivil, E. (2011). An introduction to Augmented Reality with applications in aeronautical maintenance. *13th International Conference on Transparent Optical Networks*, 1–4. https://doi.org/10.1109/ICTON.2011.5970856

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *ArXiv Preprint*. https://arxiv.org/abs/1812.04608v2

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). Explainable AI methods - A brief overview. *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, *13200 LNAI*, 13–38. https://doi.org/10.1007/978-3-031-04083-2_2

Jakubik, J., Schöffer, J., Hoge, V., Vössing, M., & Kühl, N. (2022). An empirical evaluation of predicted outcomes as explanations in human-AI decision-making. *Communications in Computer and Information Science,* 353–368. https://doi.org/10.1007/978-3-031-23618-1_24

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Jiang, J., Kahai, S., & Yang, M. (2022). Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies*, *165*, 102839. https://doi.org/10.1016/J.IJHCS.2022.102839

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, *65*(2), 337–359. https://doi.org/10.1177/00187208211013988

Kapogiannis, G. (2013). A conceptual framework for project managers to improve projects performance. *The University of Salford*. United Kingdom, 2013

Kim, S., Chang, S., & Castro-Lacouture, D. (2020). Dynamic modeling for analyzing impacts of skilled labor shortage on construction project management. *Journal of Management in Engineering*, *36*(1), 04019035. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000720

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *Conference on Human Factors in Computing Systems - Proceedings*, 2390–2395. https://doi.org/10.1145/2858036.2858402

Klopfer, E., & Squire, K. (2008). Environmental detectives-the development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development*, *56*(2), 203–228. https://doi.org/10.1007/S11423-007-9037-6

Kok, I., Okay, F. Y., Muyanli, O., & Ozdemir, S. (2023). Explainable artificial intelligence (XAI) for internet of things: A Survey. *IEEE Internet of Things Journal*, *10*(16), 14764–14779. https://doi.org/10.1109/JIOT.2023.3287678

Laato, S., Tiainen, M., Najmul Islam, A. K. M., & Mäntymäki, M. (2021). How to explain AI systems to end users: a systematic literature review and research agenda. *Internet Research*, *32*(7), 1–31. https://doi.org/10.1108/INTR-08-2021-0600/FULL/PDF

Lagaros, N. D., & Plevris, V. (2022). Artificial intelligence (AI) applied in civil engineering. *Applied Sciences* , *12*(15), 7595. https://doi.org/10.3390/APP12157595

Le Chenechal, M., Duval, T., Gouranton, V., Royan, J., & Arnaldi, B. (2016). Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, 9–12. https://doi.org/10.1109/3DCVE.2016.7563559

Lee, G., Kang, H., Lee, J., & Han, J. (2020). A user study on view-sharing techniques for one-to-many mixed reality collaborations. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 343–352. https://doi.org/10.1109/VR46266.2020.00054

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/HFES.46.1.50_30392

Livingston, M. A., Rosenblum, L., & Macedonia, M. (2005). Evaluating human factors in augmented reality systems. *IEEE Computer Graphics and Applications*, *25*(6), 6–9. https://doi.org/10.1109/MCG.2005.130

Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 5463–5472. https://hdl.handle.net/10125/71284

Loft, S., Bhaskara, A., Lock, B. A., Skinner, M., Brooks, J., Li, R., & Bell, J. (2023). The Impact of Transparency and Decision Risk on Human–Automation Teaming Outcomes. *Human Factors*, *65*(5), 846–861. https://doi.org/10.1177/00187208211033445

Lukosch, S., Billinghurst, M., Alem, L., & Kiyokawa, K. (2015). Collaboration in augmented reality. *Computer Supported Cooperative Work: CSCW: An International Journal*, *24*(6), 515–525. https://doi.org/10.1007/S10606-015-9239-0/METRICS

Madhavan, P., & Wiegmann, D. A. (2005). Cognitive anchoring on self-generated decisions reduces operator reliance on automated diagnostic AIDS. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(2), 332–341. https://doi.org/10.1518/0018720054679489

Martín-Gutiérrez, J. , Fabiani, P. , Benesova, W. , Meneses, M. D. , & Mora, C. E. (2015). Augmented reality to promote collaborative and autonomous learning in higher education. *Computers in Human Behavior*, *51, 752–761*. https://doi.org/10.1016/j.chb.2014.11.093

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management. *Human Factors*, *58*(3), 401–415. https://doi.org/10.1177/0018720815621206

Merritt, S. M. (2011). Affective processes in human-automation interactions. *Human Factors*, *53*(4), 356–370. https://doi.org/10.1177/0018720811411912

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, *39*(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: a class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies*, *2351*, 282–292. https://doi.org/10.1117/12.197321

Millecamp, M., Conati, C., Htun, N. N., & Verbert, K. (2019). To explain or not to explain: The effects of personal characteristics when explaining music recommendations. *International Conference on Intelligent User Interfaces, Proceedings IUI*, *Part F147615*, 397–407. https://doi.org/10.1145/3301275.3302313

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/J.ARTINT.2018.07.007

Mueller, S. T., Veinott, E. S., Hoffman, R. R., Klein, G., Alam, L., Mamun, T., & Clancey, W. J. (2021). Principles of Explanation in Human-AI Systems. *ArXiv Preprint*. https://arxiv.org/abs/2102.04972v1

Nikas, A. , Poulymenakou, A. , Balali, V. , & Han, K. (2007). Investigating antecedents and drivers affecting the adoption of collaboration technologies in the construction industry. *Automation in Construction*, *16 (5)-632–641*. https://doi.org/10.1016/j.autcon.2006.10.003

Noghabaei, M., Heydarian, A., Balali, V., & Han, K. (2020). Trend analysis on adoption of virtual and augmented reality in the architecture, engineering, and construction industry. *Data, 5*(1), 26. https://doi.org/10.3390/DATA5010026

Obermair, F., Althaler, J., Seiler, U., Zeilinger, P., Lechner, A., Pfaffeneder, L., Richter, M., & Wolfartsberger, J. (2020). Maintenance with augmented reality remote support in comparison to paper-based instructions: Experiment and analysis. *IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, 942–947. https://doi.org/10.1109/ICIEA49774.2020.9102078

Olorunfemi, A., Dai, F., Tang, L., & Yoon, Y. (2018). Three-dimensional visual and collaborative environment for jobsite risk communication. *Construction Research Congress 2018: Safety and Disaster*, 345–355. https://doi.org/10.1061/9780784481288.034

O'Neill, O. (2002). *Autonomy and Trust in Bioethics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511606250

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International journal of surgery, 88,* 105906.. https://doi.org/10.1371/JOURNAL.PMED.1003583

Panganiban, A. R., Matthews, G., & Long, M. D. (2020). Transparency in autonomous teammates: Intention to support as teaming information. *Journal of Cognitive Engineering and Decision Making*, *14*(2), 174–190. https://doi.org/10.1177/1555343419881563

Papenmeier, A., Kern, D., Englebienne, G., & Seifert, C. (2022). It's complicated: The relationship between user trust, model accuracy and explanations in AI. *ACM Transactions on Computer-Human Interaction*, *29*(4), 1–33. https://doi.org/10.1145/3495013

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Park, E., Jenkins, Q., & Jiang, X. (2008). measuring trust of human operators in new generation rescue robots. *Proceedings of the JFPS International Symposium on Fluid Power*, *7*(2), 489–492. https://doi.org/10.5739/ISFP.2008.489

Patel, H., Pettitt, M., & Wilson, J. R. (2012). Factors of collaborative working: A framework for a collaboration model. *Applied Ergonomics*, *43*(1), 1–26. https://doi.org/https://doi.org/10.1016/j.apergo.2011.04.009

Phillips, R. R., & Madhavan, P. (2011). The effect of distractor modality and processing code on human-automation interaction. *Cognition, Technology and Work*, *13*(4), 233–244. https://doi.org/10.1007/S10111-010-0163-2

Pynadath, D. V., Barnes, M. J., Wang, N., & Chen, J. Y. C. (2018). Transparency communication for machine learning in human-automation interaction. *Human and Machine Learning: Human–Computer Interaction Series*, 75–90. https://doi.org/10.1007/978-3-319-90403-0_5

Rafsanjani, H. N., & Nabizadeh, A. H. (2023). Towards human-centered artificial intelligence (AI) in architecture, engineering, and construction (AEC) industry. *Computers in Human Behavior Reports*, *11*, 100319. https://doi.org/10.1016/J.CHBR.2023.100319

Rankohi, S., & Waugh, L. (2013). Review and analysis of augmented reality literature for construction industry. *Visualization in Engineering*, *1*(1), 1–18. https://doi.org/10.1186/2213-7459-1-9

Ravi, M., Negi, A., & Chitnis, S. (2022). A comparative review of expert systems, recommender systems, and explainable AI. *IEEE 7th International Conference for Convergence in Technology, I2CT,* 1–8. https://doi.org/10.1109/I2CT54291.2022.9824265

Rebolledo Font de la Vall, R., & González Araya, F. (2023). Exploring the benefits and challenges of AI-Language Learning Tools. International Journal of Social Sciences and Humanities Invention, 10(01), 7569–7576. https://doi.org/10.18535/ijsshi/v10i01.02

Riveiro, M., & Thill, S. (2022). The challenges of providing explanations of AI systems when they do not behave like users expect. *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 110–120. https://doi.org/10.1145/3503252.3531306

Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, *26*(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Sabeti, S., Shoghli, O., Baharani, M., & Tabkhi, H. (2021). Toward AI-enabled augmented reality to enhance the safety of highway work zones: Feasibility, requirements, and challenges. *Advanced Engineering Informatics*, *50*, 101429. https://doi.org/10.1016/J.AEI.2021.101429

Sadhu, A., Peplinski, J. E., Mohammadkhorasani, A., & Moreu, F. (2023). A review of data management and visualization techniques for structural health monitoring using bim and virtual or augmented reality. *Journal of Structural Engineering*, *149*(1), 03122006. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003498

Sahu, C. K., Young, C., & Rai, R. (2021). Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review. *International Journal of Production Research*, *59*(16), 4903–4959. https://doi.org/10.1080/00207543.2020.1859636

Sasikumar, P., Gao, L., Bai, H., & Billinghurst, M. (2019). Wearable remotefusion: A mixed reality remote collaboration system with local eye gaze and remote hand gesture sharing. *Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR,* 393–394. https://doi.org/10.1109/ISMAR-ADJUNCT.2019.000-3

Schmalstieg, D. , Fuhrmann, A. , Hesina, G., Szalavári, Z., Encarnaçao, L. M., Gervautz, M., & Purgathofer, W. (2002). The Studierstube Augmented Reality Project. *Presence: Teleoperators & Virtual Environments, 11*(1), 33-54. https://doi.org/10.1162/105474602317343640

Schneider, T. R., Jessup, S. A., Stokes, C., Rivers, S., Lohani, M., & McCoy, M. (2017, May). The influence of trust propensity on behavioral trust. *Poster Session Presented at the Meeting of Association for Psychological Society, Boston.*

Schraagen, J. M., Lopez, S. K., Schneider, C., Schneider, V., Tönjes, S., & Wiechmann, E. (2021). The Role of Transparency and Explainability in Automated Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 27–31. https://doi.org/10.1177/1071181321651063

Shafei, H., Radzi, A. R., Algahtany, M., & Rahman, R. A. (2022). Construction 4.0 Technologies and Decision-Making: A Systematic Review and Gap Analysis. *Buildings 2022, Vol. 12, Page 2206*, *12*(12), 2206. https://doi.org/10.3390/BUILDINGS12122206

Shah, S. J., & Bliss, J. P. (2017). Does accountability and an automation decision aid's reliability affect human performance in a visual search task? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 183–187. https://doi.org/10.1177/1541931213601530

Shen, Y. , Ong, S. K. , & Nee, A. Y. C. (2010). Augmented reality for collaborative product design and development. *Design Studies*, *31(2), 118–145*. https://doi.org/10.1016/j.destud.2009.11.001

Shin, D. H., & Dunston, P. S. (2008). Identification of application areas for augmented reality in industrial construction based on technology suitability. *Automation in Construction*, *17*(7), 882–894. https://doi.org/10.1016/j.autcon.2008.02.012

Siemionow, K. B., Katchko, K. M., Lewicki, P., & Luciano, C. J. (2020). Augmented reality and artificial intelligence-assisted surgical navigation: Technique and cadaveric feasibility study. *Journal of Craniovertebral Junction and Spine*, *11*(2), 81–85. https://doi.org/10.4103/JCVJS.JCVJS_48_20

Solberg, E., Kaarstad, M., Eitrheim, M. H., Bisio, R., Reegård, K., & Bloch, M. (2022). A conceptual model of trust, perceived risk, and reliance on AI decision aids. Group & Organization Management, 47(2), 187–222. https://doi.org/10.1177/10596011221081238

Szymanski, M., Millecamp, M., & Verbert, K. (2021). Visual, textual or hybrid: The effect of user expertise on different explanations. *International Conference on Intelligent User Interfaces, Proceedings of 26th International Conference on Intelligent User Interfaces*, 109–119. https://doi.org/10.1145/3397481.3450662

Takim, R., Harris, M., & Nawawi, A. H. (2013). building information modeling (BIM): A new paradigm for quality of life within architectural, engineering and construction (AEC) Industry. *Procedia - Social and Behavioral Sciences*, *101*, 23–32. https://doi.org/10.1016/J.SBSPRO.2013.07.175

Tecuci, G. (2012). Artificial intelligence. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 168–180. https://doi.org/10.1002/WICS.200

Tilloo, P., Parron, J., Obidat, O., Zhu, M., & Wang, W. (2022). A POMDP-based robot-human trust model for human-robot collaboration. *12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems*, 1009–1014. https://doi.org/10.1109/CYBER55403.2022.9907660

Um, T.-W., Kim, J., Lim, S., & Lee, G. M. (2022). Trust management for artificial intelligence: a standardization perspective. *Applied Sciences*, *12*(12), 6022. https://doi.org/10.3390/app12126022

Utzig, S., Kaps, R., Azeem, S. M., & Gerndt, A. (2019). Augmented reality for remote collaboration in aircraft maintenance tasks. *IEEE Aerospace Conference*, 1–10. https://doi.org/10.1109/AERO.2019.8742228

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy and Technology*, *34*(4), 1607–1622. https://doi.org/10.1007/S13347-021-00477-0

Wang, K., Guo, F., Zhang, C., Hao, J., & Schaefer, D. (2022). Digital technology in architecture, engineering, and construction (AEC) industry: Research trends and practical status toward construction 4.0. *Construction Research Congress 2022*, 983–992. https://doi.org/10.1061/9780784483978.100

Wang, N., Pynadath, D. V., & Hill, S. G. (2016). Trust calibration within a human-robot team: Comparing automatically generated explanations. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016-April*, 109–116. https://doi.org/10.1109/HRI.2016.7451741

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Han, D., Lv, H., He, W., Yan, Y., Zhang, X., & Min, H. (2019). An MR remote collaborative platform based on 3D CAD models for training in industry. In *Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 91-92). IEEE https://doi.org/10.1109/ismar-adjunct.2019.00038

Wang, X., & Dunston, P. S. (2009). Comparative effectiveness of mixed reality based virtual environments in collaborative design. *2009 IEEE International Conference on Systems, Man and Cybernetics*, *41*(3), 284–296. https://doi.org/10.1109/ICSMC.2009.5346691

Wang, X., & Dunston, P. S. (2013). Tangible mixed reality for remote design review: a study understanding user perception and acceptance. *Visualization in Engineering*, *1*(1). https://doi.org/10.1186/2213-7459-1-8

Wang, Z., He, B., Yang, Y., Shen, C., & Peña-Mora, F. (2020). Building a next generation AI platform for AEC: A review and research challenges. *Proceedings of the 37th International Conference of CIB*, (pp. 27–45). http://dx.doi.org/10.46421/27066568.37.2020.paper003

Williams, K. J., Yuh, M. S., & Jain, N. (2023). a computational model of coupled human trust and self-confidence dynamics. *ACM Transactions on Human-Robot Interaction*, *12*(3). https://doi.org/10.1145/3594715

Xie, B., & Salvendy, G. (2010). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & stress*, *14*(1), 74–99. https://doi.org/10.1080/026783700417249

Zenati-Henda, N., Bellarbi, A., Benbelkacem, S., & Belhocine, M. (2014). Augmented reality system based on hand gestures for remote maintenance. In *2014 International Conference on Multimedia Computing and Systems*, (pp.5–8). IEEE. https://doi.org/10.1109/ICMCS.2014.6911258

Zeng, Z., Miao, C., Leung, C., & Jih, C. J. (2018). Building more explainable artificial intelligence with argumentation. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), 8044–8045. https://doi.org/10.1609/AAAI.V32I1.11353

Zhang, Z., Wen, F., Sun, Z., Guo, X., He, T., & Lee, C. (2022). Artificial intelligence-enabled sensing technologies in the 5g/internet of things era: From virtual reality/augmented reality to the digital twin. Advanced Intelligent Systems, *4*(7), 2100228. https://doi.org/10.1002/AISY.202100228

Zhong, X. W. , Boulanger, P. , & Georganas, N. D. (2002). Collaborative augmented reality: A prototype for industrial training. 21th Biennial Symposium on Communication, Canada. 387–391

Zuk, T., & Carpendale, S. (2007). Visualization of uncertainty and reasoning. In A. Butz, B. Fisher, A. Krüger, P. Olivier, & S. Owada (Eds.), *Smart graphics* (pp. 164–177). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-540-73214-3_15

Below are the publications and conference papers during my tenure at Clemson University:

Bhanu, A., Sharma, H., Pathy, S. R., Ponathil, A., Rahimian, H., & Madathil, K. C. (2023, October). Trust in Artificial Intelligent Agent while Completing a Procedural Construction Task. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Washington, DC.

Bhanu, A., Sharma, H., Piratla, K., & Chalil Madathil, K. (2022, September).Application of Augmented Reality for Remote Collaborative Work in Architecture, Engineering, and Construction–A Systematic Review. In Proceedings of Human Factors and Ergonomics Society's International Annual Meeting, Atlanta, GA.

Ponathil, A., Bhanu, A., & Madathil, K. C. (2023, October). Restoring Consumer Trust in a Home Improvement Service Provider with Negative Reviews. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Washington, DC.

Ponathil, A., Bhanu, A., Piratla, K., Sharma, V., & Chalil Madathil, K. (2020, December). A Preliminary Study of Elements Impacting the Choice of an Online Contractor. In Proceedings of Human Factors and Ergonomics Society's International Annual Meeting, Seattle, WA

Ponathil, A., Bhanu, A., Piratla, K., Sharma, V., & Chalil Madathil, K. (2022). Investigation of the factors influencing the online consumer's choice of a service provider for home improvement. Electronic Commerce Research, 1-28.

Ponathil, A., Bhanu, A., Rogers, H., Wilson, M., & Chalil Madathil, K. (2020, December). An Investigation of the Usability of a Teleconsent System. In Proceedings of Human Factors and Ergonomics Society's International Annual Meeting, Seattle, WA

Torrence, C., Bhanu, A., Bertrand, J., Dye, C., Truong, K., & Madathil, K. C. (2023). Preparing future health care workers for interactions with people with dementia: A mixed methods study. Gerontology & Geriatrics Education, 44(2), 223-242.

Torrence, C., Madathil, K., Bertrand, J., & Bhanu, A. (2020, October). Preparing Alzheimer's Disease Carers for Caregiving Through Virtual Reality. In American Public Health Association's 2020 Virtual Annual Meeting and Expo, Virtual Event