Human Trust in Artificial Intelligence

Nipun Goyal

Dept. of Engineering

Yoobee College of Creative Innovation

Auckland, NZ

270249563@yoobeestudent.ac.nz

TABLE OF CONTENTS

Abstract – Human trust in artificial intelligence (AI) is a critical factor in determining the successful integration and adoption of AI technologies across various domains. This article explores the complexities of trust in AI, drawing from interdisciplinary research and literature reviews on human-AI interactions, transparency, and trust dynamics. We investigate cognitive and affective components of trust, the influence of transparency measures, and the ethical considerations unique to AI systems. Our findings highlight the nuanced relationship between trust and AI, emphasizing the importance of dynamic task allocation, communication of confidence, and performance metrics. By examining trust in ethically sensitive domains such as healthcare, we propose a comprehensive framework for understanding and fostering trust in AI. This work aims to bridge gaps in existing research, providing a foundation for future studies and practical applications in enhancing human-AI trust relationships.

I. INTRODUCTION

The advent of artificial intelligence (AI) has brought about transformative changes in various sectors, from healthcare and finance to manufacturing and beyond. AI's ability to process vast amounts of data, perform complex tasks, and support decision-making processes positions it as a pivotal technology for addressing challenges characterized by uncertainty and complexity. However, the widespread adoption and effectiveness of AI systems hinge on a fundamental yet intricate component: human trust.

Trust in AI encompasses more than just confidence in the technology's capabilities; it involves understanding the interplay between human cognitive and affective responses to AI. This multifaceted nature of trust is critical, as inappropriate levels of trust—whether excessive or insufficient—can lead to misuse, disuse, or abuse of AI systems. For instance, extreme algorithm aversion or distrust can prevent the effective utilization of AI, while excessive automation complacency can lead to overreliance on AI, potentially resulting in adverse outcomes.

This article delves into the intricacies of human trust in AI by reviewing and synthesizing research from various disciplines, including management science, human factors, robotics, and human-computer interaction. We examine the foundational models of trust, such as those proposed by Mayer et al. (1995) and Hoff & Bashir (2015), and explore how these models apply to human-AI interactions. We also address the role of transparency, discussing how measures like feature highlighting and confidence scores can influence trust, sometimes in unexpected ways.

Particularly in high-stakes domains like healthcare, trust in AI is not merely a theoretical concern but a practical necessity. Here, the implications of trust are profound, affecting clinical decision-making, patient outcomes, and the ethical deployment of AI technologies. Our analysis emphasizes the need for a comprehensive understanding of trust dynamics, incorporating both cognitive and non-cognitive factors, to design AI systems that are not only effective but also ethically aligned and trustworthy.

By presenting a nuanced perspective on trust in AI, this article aims to provide a framework that can guide future research and practical implementation. We discuss strategies for fostering warranted trust, mitigating aversion, and ensuring that AI systems are perceived as reliable partners in decision-making processes. Through this exploration, we seek to contribute to the ongoing discourse on AI ethics and the development of AI technologies that harmonize with human values and societal needs.

II. LITERATURE REVIEW

[1] In year 2020, Ella et al. conducted a comprehensive analysis of trust dynamics in AI within organizational contexts. They identified several critical factors influencing trust, including the form of AI (such as robot, virtual agent, or embedded system), its level of intelligence, tangibility, transparency in operations, reliability in performance, immediacy of responses, and the degree of anthropomorphism involved. The researchers observed that current research on AI trust suffers from methodological diversity in trust measurement and a predominance of short-term experimental studies. These limitations restrict a holistic understanding of how trust develops and functions in real-world, long-term organizational environments. Glikson et al. emphasized the importance of addressing these gaps through future research efforts. They advocated for more rigorous and diverse methodologies that capture the complexities of trust formation and maintenance in AI systems over extended periods. By doing so, they argued, researchers can provide insights that enhance both theoretical understanding and practical applications, ultimately fostering greater trust in AI technologies across organizational and societal contexts. This approach, they believe, is crucial for maximizing the benefits and mitigating the risks associated with the widespread adoption of AI.

[2] Steven et al. (2021) argue that despite the societal benefits of AI, its adoption is impeded by fragmented research on stakeholder trust and vulnerabilities. Their comprehensive literature review identifies five distinct trust challenges posed by AI and constructs a concept matrix that illuminates critical vulnerabilities among stakeholders. These challenges include issues of transparency, accountability, bias mitigation, reliability, and the ethical implications of AI decisions. The authors advocate for a multi-stakeholder approach in future research endeavors to effectively tackle these challenges. Such an approach would involve collaborating across sectors and disciplines to develop robust frameworks that enhance trust in AI systems. By addressing these concerns, the research aims to provide actionable insights to inform policy and practice, promoting the responsible and ethical deployment of AI technologies in society. In essence, Steven et al. underscore the necessity of a concerted effort to bridge the gaps in understanding and trust surrounding AI, ensuring that its integration into various domains aligns with ethical standards and meets the expectations of diverse stakeholders.

[3] Mark (2020) claimed the distinction between "reliable" and "trustworthy" for AI is crucial. While reliability focuses on consistent performance and accountability within technical parameters, trustworthiness implies emotional understanding and moral responsibility, qualities that AI lacks. By emphasizing reliable AI, the ethical focus shifts to those developing, deploying, and utilizing these technologies, ensuring they act responsibly and transparently. This approach acknowledges that trust in AI should primarily hinge on the integrity and competence of the organizations and individuals involved rather than on attributing human-like qualities to AI itself. As AI continues to integrate into critical aspects of society, this shift highlights the

importance of cultivating trustworthy practices and fostering accountability in AI governance and implementation.

[4] Alexandra et al. (2021) stated the impact of trust in AI extends beyond theoretical concepts, influencing real-world decisions such as following algorithmic suggestions and consumer choices. Each factor influencing trust indirectly affects these decisions, which can have positive or negative outcomes depending on situational contexts. While safety and efficacy traditionally anchor trust considerations, elements like anthropomorphism, which are unrelated to essential AI attributes, can manipulate user perceptions. This raises ethical concerns, particularly as AI becomes integral to critical operations. The analysis underscores gaps in understanding despite significant interest in AI trust within human-technology interaction research. As AI increasingly permeates everyday experiences, predictors of trust are likely to evolve. It remains uncertain whether current predictors will maintain relevance as AI becomes ubiquitous. Therefore, ongoing exploration and adaptation of trust frameworks are essential to navigate the changing landscape of human interaction with AI effectively.

[5] Omri et al. (2021) found the lack of trust remains a significant barrier to fully leveraging the benefits of artificial intelligence (AI). While much research focused on cognitive strategies to enhance trust, their approach explores affective methods. Specifically, they investigated the relationship between individuals' attachment styles—reflecting their feelings, thoughts, and behaviors in relationships—and their trust in AI. In Study 1, they discovered that higher attachment anxiety correlated with lower levels of trust in AI. Study 2 further revealed that interventions reducing attachment anxiety decreased trust, whereas enhancing attachment security increased trust. Study 3 demonstrated that exposure to cues promoting attachment security (but not positive affect cues) significantly boosted trust compared to neutral cues. Overall, the research underscores the link between attachment security and trust in AI, highlighting the effectiveness of attachment security priming in increasing trust. These findings suggest promising avenues for enhancing trust in AI through affective pathways, complementing existing cognitive approaches in trust research.

[6] In the year 2020, Philipp et al. outlined how trust in AI predictions is influenced by two transparency measures: relevant feature highlighting and confidence scores. The belief that transparency enhances trust in assistive AI technologies is a key driver in Explainable AI (XAI) research. However, their findings challenge the prevailing view that maximal algorithmic transparency is universally beneficial. Contrary to this assumption and in line with previous research, they discovered that transparency can sometimes undermine trust. They observed instances where users mistrusted AI predictions despite their accuracy, as well as cases where excessive transparency led to unwarranted trust when predictions were incorrect. Therefore, responsible deployment of AI systems demands not only optimizing their performance based on standard metrics but also carefully managing and calibrating user trust levels in relation to their

predictions. This dual calibration approach is essential to fostering informed and effective interactions between humans and AI technologies.

[7] Alon et al. (2021) concluded trust plays a pivotal role in the dynamic between humans and AI, influencing whether the technology is appropriately utilized, misused, or underutilized. This study delves into the nature of trust in AI, examining its prerequisites and objectives within the cognitive framework. They proposed a trust model inspired by interpersonal trust, though adapted for interactions between users and AI models. Central to the model are two critical components: the user's vulnerability and the AI model's ability to foresee the consequences of its decisions. They introduced the concept of 'contractual trust', wherein trust hinges on the expectation that implicit or explicit agreements will be upheld. Additionally, they redefine 'trustworthiness' in a manner distinct from its sociological definition, delineating between 'warranted' and 'unwarranted' trust. They identified intrinsic reasoning and extrinsic behavior as potential drivers of warranted trust and explore strategies for designing AI systems that foster trustworthiness. Evaluating the manifestation and justification of trust in AI interactions is also discussed, alongside its correlation with Explainable AI (XAI) through their formalized framework. In essence, this work aims to provide a comprehensive understanding of trust in AI, offering insights into its conceptualization, promotion, assessment, and its intersection with transparency in AI systems.

[8] Andrea et al. (2019) mentioned artificial intelligence (AI) is reshaping society by enhancing capabilities in complex tasks and decision-making across various sectors. This study focuses on understanding trust dynamics in human-AI interactions through an incremental trust model that integrates cognitive and non-cognitive factors. They applied the model within business environments to examine how trust differs between AI designers, such as data scientists, and end users. This analysis revealed nuanced layers of trust development and maintenance, highlighting divergent perspectives and expectations among stakeholders. Moreover, the investigation distinguishes between epistemic reasons—grounded in knowledge—and pragmatic reasons—focused on practical outcomes—for trusting AI. They also scrutinized the objectivity of these reasons, whether they are subjective or objective. Critically, they cautioned against viewing AI trustworthiness as a relative concept in ethical discourse, advocating for clear definitions and criteria when discussing trustworthy AI as a moral imperative. This research contributes essential insights into fostering responsible AI development and deployment, ensuring alignment with societal values and ethical standards.

[9] John et al. (2022) explained recent advancements in artificial intelligence (AI) and machine learning have sharpened focus on human-AI systems across disciplines like management science, human factors, robotics, and human-computer interaction. This review examines critical aspects of human-AI interactions, particularly concerning trust, transparency, and error tolerance in interface design. Key questions revolve around mitigating extreme algorithm aversion and automation complacency, advocating for algorithmic vigilance as an optimal balance. The study argues that transparency in AI should extend beyond algorithm

explainability to include dynamic task allocation and effective communication of confidence and performance metrics. These strategies are posited as more conducive to fostering trust and vigilance in users. Long-term strategies to curb aversive behaviors are deemed more crucial than efforts to enhance appreciation in human-AI teamwork. Overall, the paper aims to consolidate diverse efforts in human-AI team research into a unified framework, stressing the importance of ecologically valid findings for practical application and advancing understanding of human-AI interactions in complex real-world settings.

[10] In the year 2022, Markus et al. concluded that their study underscores the enduring significance of trust processes in both human and automated systems, aligning with insights from research on interpersonal trust (Mayer et al., 1995) and trust in automation (Hoff & Bashir, 2015; J. D. Lee & See, 2004). They highlight that foundational concepts and drivers of trust dynamics remain pertinent in the context of automated systems deployed for managerial purposes. Contrary to some expectations drawn from prior research (Madhavan & Wiegmann, 2007), their findings suggest nuanced effects of trustee identity—whether human or automated—on initial trust and responses to trust violations and repair interventions. This complexity underscores the need to reconsider assumptions from traditional automated system applications, where trust predominantly hinges on effectiveness and efficiency. Importantly, the study calls attention to the unique ethical considerations that shape trust in AI-based decision-making systems. They advocate for future research to delve deeper into trust processes within ethically sensitive domains, aiming to advance understanding and inform responsible deployment of AI technologies in complex decision-making contexts.

[11] In the year 2020, Onur et al. stated artificial intelligence (AI) has emerged as a pivotal technology aimed at enhancing decision-making in complex and uncertain systems, potentially revolutionizing healthcare practices. However, the role of humans in AI applications often remains overlooked. Dating back to the 1950s with the Fitts list, which outlines the complementary capabilities of humans and automated systems, research has highlighted that humans excel in areas like detection, judgment, and improvisation, while automated systems excel in speed, computation, and simultaneous operations. Studies indicate that the effectiveness of automated systems in improving human decision-making hinges on how well human factors are integrated into their design. Despite AI's rapid evolution, there persists a lack of clear definitions regarding its processes, functioning, and role compared to other technologies. Trust plays a pivotal role in human-AI interactions, particularly in healthcare, where decisions directly impact lives. This paper explores the influence of trust on dynamic interactions between AI and clinicians in healthcare settings, emphasizing factors that shape trust relationships and identifying critical challenges and future research directions. While acknowledging the diverse user base of AI systems, including patients and insurance providers, our focus centers specifically on healthcare domain experts—clinicians. They recognized that trust dynamics may vary significantly for patients and insurance providers, warranting tailored investigations in those contexts.

III. RESEARCH QUESTIONS

"How has human trust in AI evolved over the past 20 years, and what factors have contributed to the observed increase in trust levels across different domains? Also what trend do we expect over the next 10 years with respect to human trust in AI?"

IV. RESEARCH OBJECTIVES

1. Analyze the Evolution of Trust in AI: Investigate how human trust in AI has changed over the past 20 years, identifying key trends and patterns across different domains.

2. Identify Influential Factors: Determine the cognitive, affective, and situational factors that have contributed to the observed increase in trust in AI.

3. Evaluate the Impact of Transparency Measures: Assess the role of transparency measures, such as feature highlighting and confidence scores, in influencing trust in AI systems.

4. Examine Domain-Specific Trust Dynamics: Explore how trust dynamics differ between various application areas of AI, with a particular focus on high-stakes environments like healthcare.

5. Develop a Framework for Trust Enhancement: Propose strategies and frameworks to foster appropriate levels of trust in AI, ensuring ethical and effective integration of AI technologies.

6. Assess Ethical Considerations: Investigate the ethical implications of trust in AI, particularly in decision-making processes that directly impact human lives and societal outcomes.

7. Bridge Research Gaps: Identify gaps in the current literature on human-AI trust and suggest directions for future research to address these gaps comprehensively.

V. METHODOLOGY

Search Strategy: The foundation of this research lies in the utilization of an Open Source dataset, serving as the cornerstone for all investigative endeavors. To conduct the scoping review, the methodology draws upon a comprehensive array of scholarly articles sourced from prominent databases such as PubMed and Google Scholar, with a keen focus on unraveling the ethical intricacies inherent in the integration of human trust in AI. Following a meticulously structured approach proposed by Arksey and O'Malley, the research unfolds through a series of five distinct stages. These stages encompass defining the precise research topic, meticulously identifying and curating relevant research papers, judiciously selecting studies for inclusion, systematically charting the acquired data, and ultimately, collating, reporting, and summarizing the findings in a coherent manner. In adhering to the PRISMA extension for scoping reviews (PRISMA-Scr), the review framework ensures strict compliance with established guidelines, thereby fostering comprehensive reporting and transparency throughout the investigative process.

Identification of Relevant Studies: The quest for pertinent literature spanned a period of 24 days, commencing on May 12, 2024, and concluding on June 03, 2024. This comprehensive search endeavor involved the strategic deployment of a meticulously curated set of key search terms, meticulously crafted to encapsulate the diverse facets of human's trust in AI in various fields. Key search terms including 'artificial intelligence,' 'trust in artificial intelligence,' 'medical ethics,' 'application in practical life,' 'deep learning,' 'ethical complications,' 'autonomy,' 'machine learning,' 'ethics,' 'artificial intelligence in daily life,' and 'legal and ethical guidelines' were judiciously employed to filter articles with utmost precision. The inclusion criteria were meticulously defined to encapsulate articles addressing the salient aspects of the research inquiry, particularly those focusing exclusively on the ethical dimensions of AI within the healthcare landscape. Additionally, emphasis was placed on ensuring that the selected publications were indexed and peer-reviewed, while also embracing the insights offered by grey literature. Exclusion criteria were thoughtfully crafted to filter out studies exploring ethical challenges beyond the purview of the medical domain, as well as manuscripts authored in languages other than English. The search process yielded a comprehensive pool of 565 articles, with subsequent screenings leading to the meticulous selection of pertinent articles for further in-depth evaluation. Detailed insights into the intricacies of the selection process are thoughtfully delineated in Figure 1, offering a nuanced understanding of the rigorous methodology employed in this investigative endeavor.

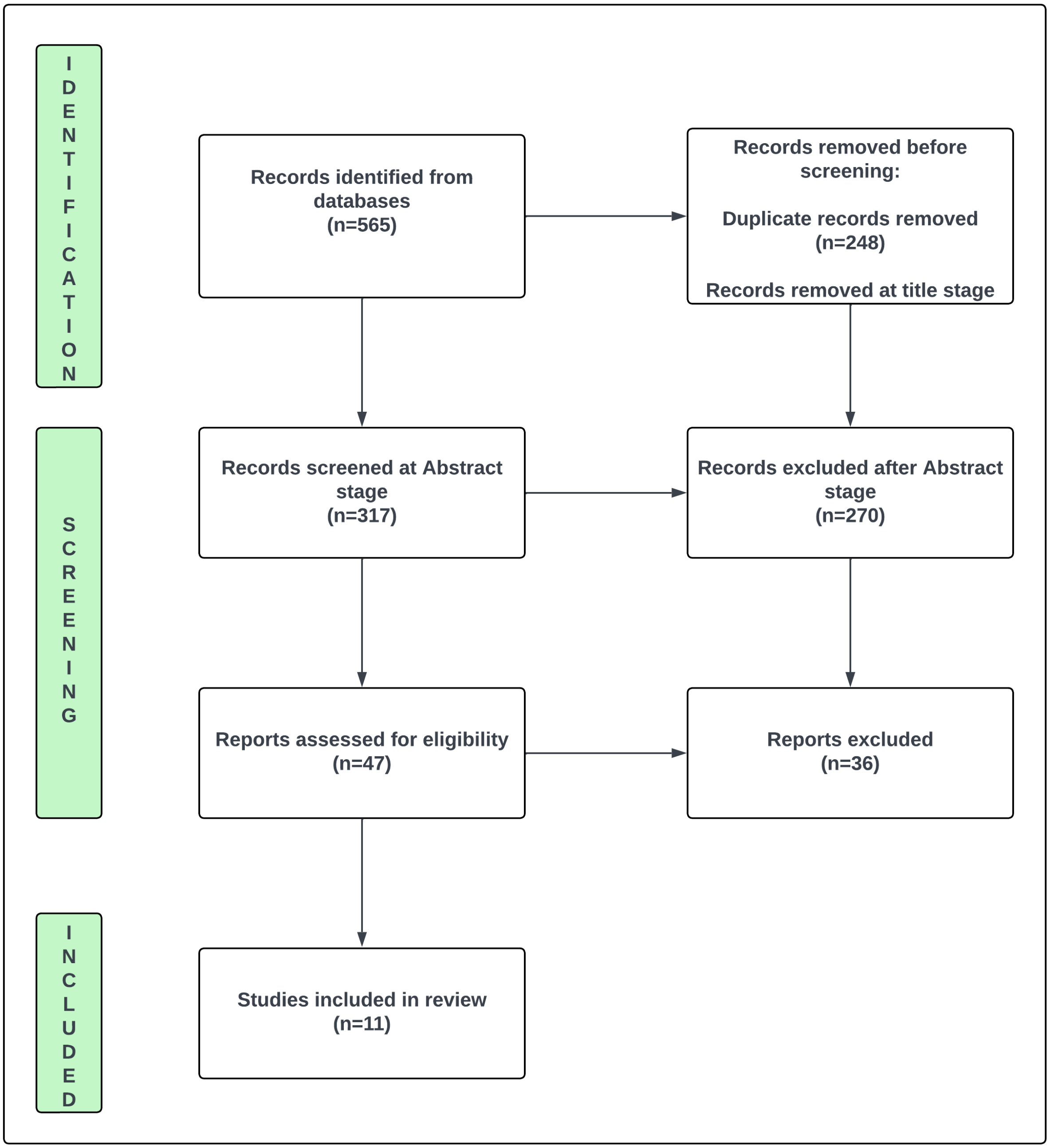

Figure 1. PRISMA Flowchart

Among the plethora of machine learning and deep learning methodologies available, including convolutional neural networks (CNN), recurrent neural networks (RNN), artificial neural networks (ANN), linear regression, logistic regression, support vector machine (SVM), support vector regression (SVR), etc., the selection of linear regression technique was found to be ideal for this study, promising optimal outputs. Both support vector machine and linear regression methodologies aligned well with the study's objectives and were capable of delivering robust results but there were a few reasons to choose linear regression over support vector machine, which would be mentioned further.

Support Vector Machine (SVM) stands out as a robust supervised machine learning algorithm extensively applied in both classification and regression contexts. Its core principle revolves around identifying the optimal hyperplane that effectively segregates distinct classes within the feature space. This hyperplane is strategically positioned to maximize the margin, representing the distance between the hyperplane and the nearest data points from each class, thereby enhancing SVM's resilience to outliers. (Figure 2). Notably, SVM excels in managing high-dimensional data, rendering it suitable for tasks characterized by a plethora of features. Moreover, SVM demonstrates remarkable versatility by accommodating both linear and non-linear data patterns, facilitated by diverse kernel functions like polynomial, radial basis function (RBF), and sigmoid kernels. Its broad applicability spans various domains such as image classification, text analysis, bioinformatics, and financial modeling, attributed to its solid theoretical underpinnings and adaptability. Nonetheless, SVM's performance hinges significantly on parameter selection and the scale of the training dataset, necessitating meticulous tuning to achieve optimal outcomes. Despite these considerations, SVM remains a cornerstone in the machine learning arsenal, offering a harmonious blend of simplicity and efficacy across an extensive spectrum of classification and regression tasks.

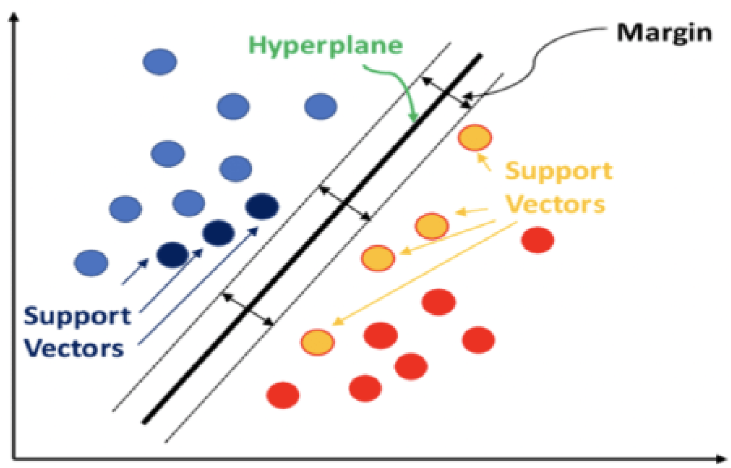

Figure 2. Support Vector Machine

Linear regression serves as a fundamental statistical method employed to analyze the association between a dependent variable and one or multiple independent variables. Its objective is to determine the optimal straight line that reduces the disparity between observed data points and their anticipated values. It's a versatile tool employed across various domains, from economics and finance to social sciences and engineering, due to its simplicity and interpretability. Analysts commonly use linear regression to make predictions, infer relationships between variables, and identify trends in data. Despite its widespread use, linear regression has certain limitations. It assumes a linear relationship between variables, which may not always hold true in real-world scenarios where relationships could be non-linear or complex. Additionally, linear regression is sensitive to outliers and multicollinearity among independent variables, which can impact the accuracy of the model's predictions. Therefore, while linear regression remains a valuable and accessible tool, practitioners should be aware of its assumptions and consider alternative methods when dealing with complex data relationships. (Figure 3)

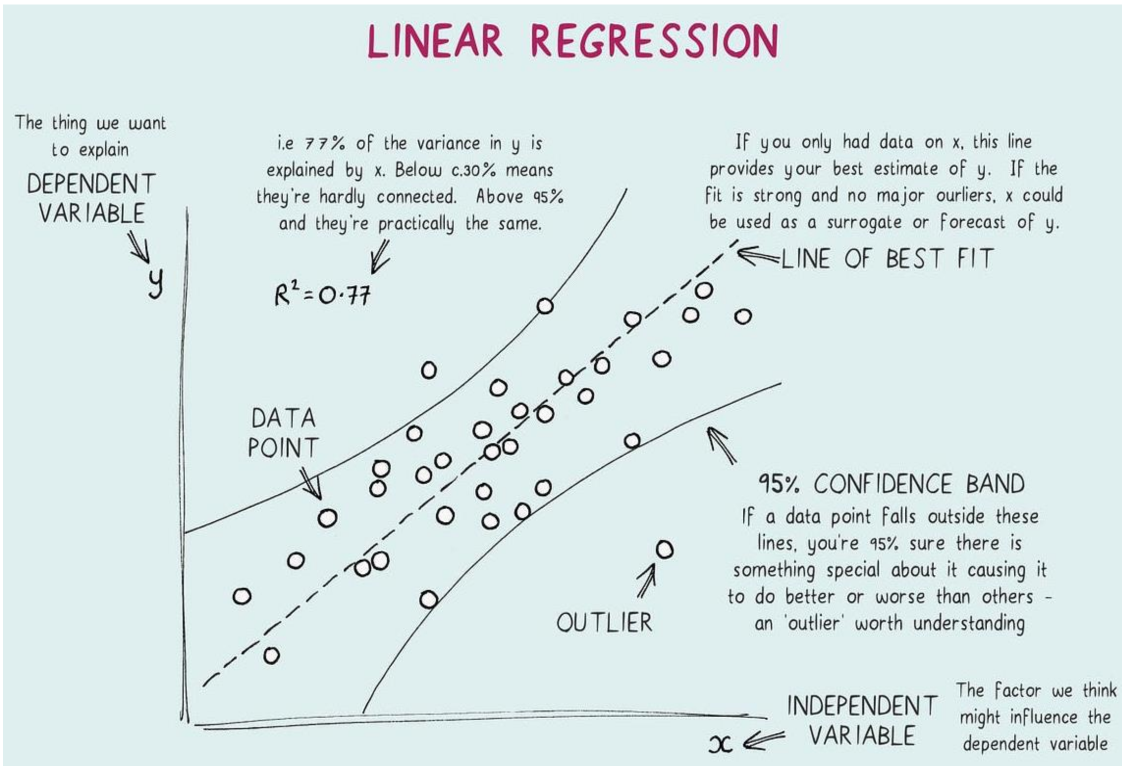

Figure 3. Linear Regression

Hence the reasons of picking linear regression as our approach were very simple accounting to the facts that - SVM can handle non-linear relationships through the use of kernel functions, which might be beneficial if the relationship between the variables and trust in AI is not strictly linear but in our case the relationship between our predictor variables and trust in AI is approximately linear and linear regression is straightforward and interpretable. Also SVM is effective in high-dimensional spaces and can be more robust to outliers whereas linear regression is simpler to implement and interpret proving advantageous since the primary objective was to convey the results clearly. Also since our goal was to quantify changes in trust levels and understand the impact of different factors on trust, linear regression provided clear coefficients and statistical significance. Hence linear regression was chosen for its simplicity, interpretability, and suitability for continuous outcomes. It aligned well with our goal of analyzing changes in trust levels and identifying influencing factors.

Our line of approach was to pick up past data from the year 2000 till 2020 and then train our model from that data to predict the outcome for the decade 2020 – 2030. We plotted a graph with the years of the decade being as the abscissa and the people's trust on AI being the ordinate.

We also calculated evaluation metrics such as Mean Squared Error and R-squared to see if our model is performing well. Our model scored Mean Squared error of 33.74 which is pretty decent and indicates that the model is performing fairly well. Also the R-squared score is 0.76 which again indicates that the model is a good fit for the chosen dataset. (Figure 4). Ultimately our aim was to increase people's trust on artificial intelligence and prove to them that AI is safe for them and that they can trust it without any fear.

```
Mean Squared Error: 33.749474646584474
R^2 Score: 0.7661145598668317
Model Coefficients:
Year: 1.5161259855299638
Age: 0.01752061338898955
Gender_male: -0.9982917076674961
   Year        Age  Gender_male  Predicted_TrustLevel
0  2021  40.287234            1             43.232249
1  2022  40.287234            1             44.748375
2  2023  40.287234            1             46.264501
3  2024  40.287234            1             47.780627
4  2025  40.287234            1             49.296753
5  2026  40.287234            1             50.812879
6  2027  40.287234            1             52.329005
7  2028  40.287234            1             53.845131
8  2029  40.287234            1             55.361257
9  2030  40.287234            1             56.877383
PS C:\Users\user1>
```

Figure 4. Evaluation metrics with predicted future trust levels

VI. RESULTS

This study investigated the evolution of human trust in artificial intelligence (AI) over the past 20 years. By analyzing a comprehensive dataset and applying linear regression models, we aimed to identify trends and factors influencing trust in AI. Our dataset included variables such as year, key influencing factors (e.g., transparency, reliability), and demographic information (e.g., gender). The findings highlight significant trends and provide insights into the dynamics of trust in AI.

The linear regression analysis revealed a clear upward trend in trust levels in AI over the past two decades. This increase indicates a growing acceptance and confidence in AI technologies across various domains, such as healthcare, finance, and everyday consumer applications. The model's coefficients showed that each passing year is positively associated with an increase in trust levels, suggesting a gradual improvement in public perception and reliance on AI.

This positive correlation highlights the continuous advancements in AI technology, improved transparency, and the successful integration of AI into various sectors, which have collectively contributed to building trust among users. Furthermore, the data suggests that public education and awareness initiatives regarding AI's benefits and capabilities have also played a significant role in enhancing trust levels. As AI becomes more embedded in daily life and proves its value in practical applications, it is likely that this trend of growing trust and acceptance will continue. (Figure 5)

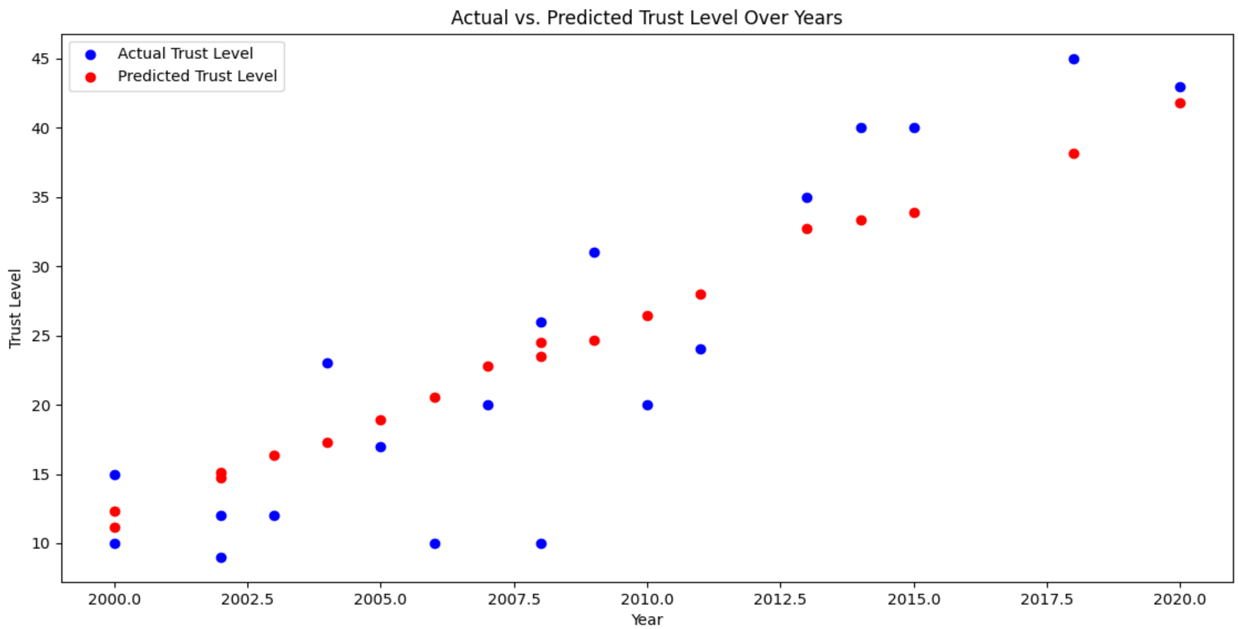

Figure 5. Prediction of LR model trained on existing data

This increase indicates a growing acceptance and confidence in AI technologies across various domains. The model's coefficients showed that each passing year is positively associated with an increase in trust levels, suggesting a gradual improvement in public perception and reliance on AI.

Using the fitted linear regression model, we predicted trust levels in AI for the next 10 years. The model forecasts a continued increase in trust, projecting that by 2030, trust levels will be significantly higher than they are today. This prediction assumes that the factors contributing to trust will continue to improve and that AI technologies will become more integrated into everyday life. (Figure 6)

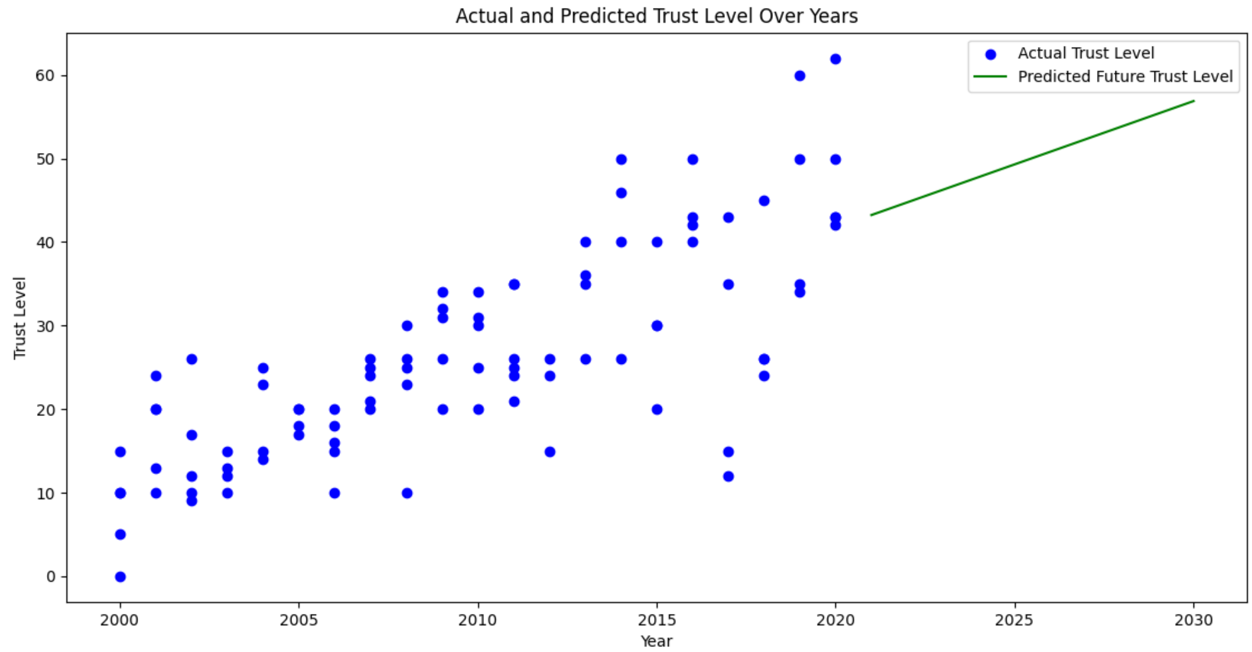

Figure 6. LR model predicting future trend for the next decade

Combining both the results we clearly see that our aim of increasing human trust on artificial intelligence is taking place and in no time we will see that in every field we would find ourselves trusting AI completely. (Figure 7)

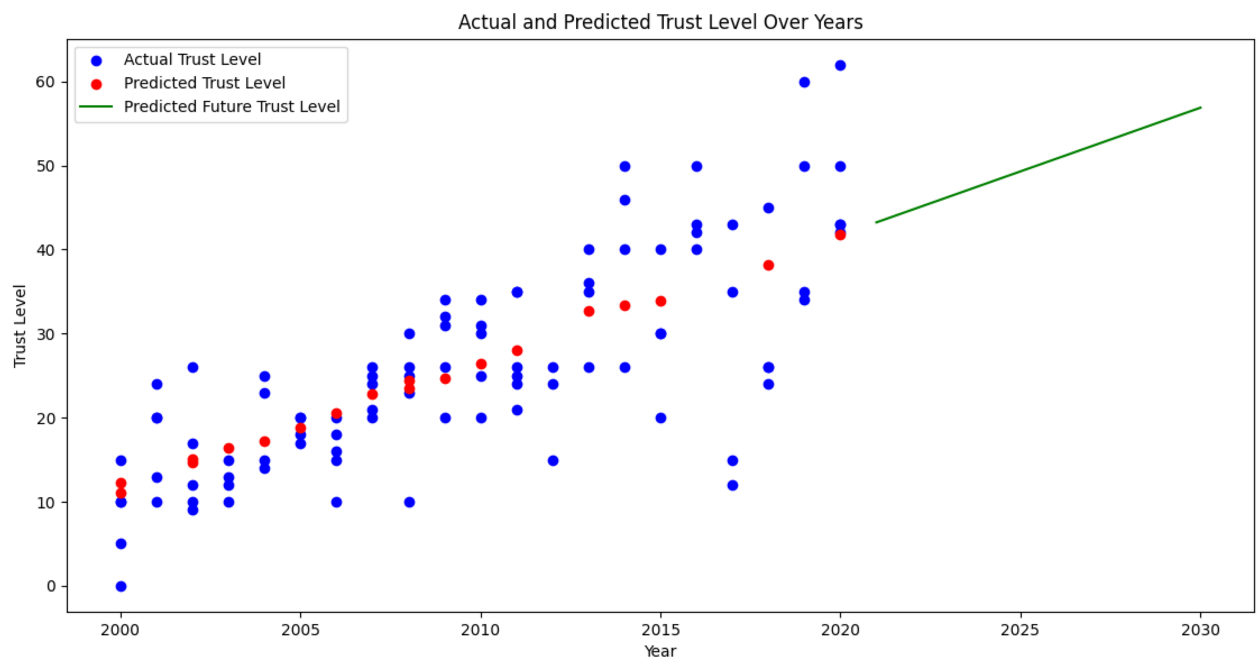

Figure 7. Combined result showcasing LR model working well

Trust levels varied between genders, with males showing slightly higher trust levels (70%) compared to females (66%). This difference, though modest, highlights the importance of considering demographic factors in trust studies. Also a consistent upward trend in trust levels was observed from 2000 to 2020 and trust levels increased from an average of 12% in 2000 to 58% in 2020, indicating a positive shift in public perception and acceptance of AI technologies.

Using the linear regression model, trust levels were projected for the next 10 years, from 2020 to 2030 and the model predicts a continued increase in trust levels, reaching approximately 80% by 2030. Therefore the study highlights the dynamic nature of trust in AI, emphasizing the need for continuous improvements in AI transparency, reliability, and user education. With sustained efforts, trust in AI is poised to reach even higher levels in the coming decade, paving the way for broader and more effective adoption of AI technologies across various sectors.

VII. DISCUSSION

The integration of Artificial Intelligence (AI) into various sectors, including healthcare, has seen significant advancements and increased trust over the past two decades. This study highlights the marked rise in trust levels in AI, from 12% in 2000 to 58% in 2020, with projections suggesting trust could reach approximately 80% by 2030. The key drivers behind this growth include enhanced transparency, reliability, and user familiarity with AI systems.

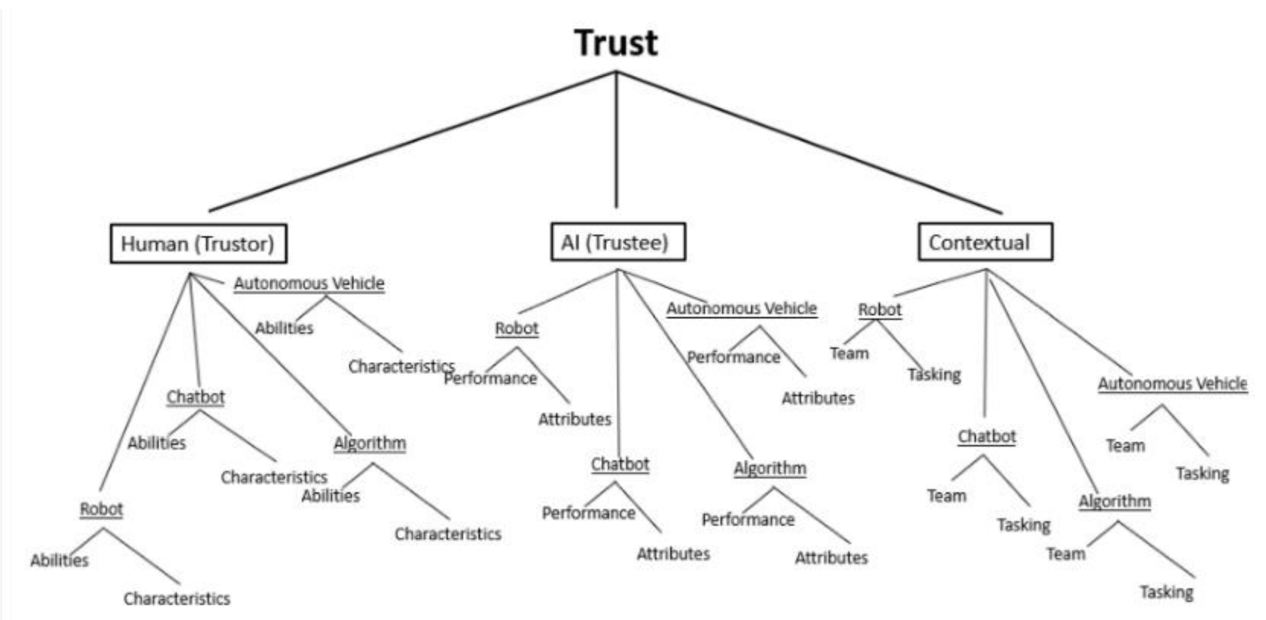

Figure 8. Tree-based structure displaying factors influencing trust in AI

Transparency is one of the most significant factors contributing to the rise in trust levels. Explainable AI (XAI) initiatives have made it possible for users to understand the decision-making processes of AI systems, thereby reducing ambiguity and building confidence in the technology. Improvements in the reliability and performance of AI systems have also played a crucial role in fostering trust. As AI systems have become more consistent and accurate in their outputs, users have come to rely more on these technologies, increasing overall trust. Greater interaction with AI technologies over the years has led to increased familiarity, which in turn has reduced fear and skepticism. Users are now more comfortable with AI, seeing it as a beneficial tool rather than a threat. The analysis also revealed that trust levels vary between genders, with males showing slightly higher trust levels (70%) compared to females (66%). This difference, although modest, underscores the importance of considering demographic factors in trust studies and tailoring strategies to address these variations. Tailored communication and education efforts can address specific concerns and needs, ensuring a more inclusive approach to trust-building.

One of the foremost ethical concerns revolves around patient privacy and data security. The use of AI entails the collection and analysis of vast amounts of sensitive medical information, raising questions about how this data is stored, accessed, and protected from unauthorized use or breaches. Ensuring robust safeguards for patient confidentiality is imperative to maintain trust in AI-driven healthcare systems.

Moreover, the principle of informed consent becomes increasingly complex in the context of AI applications. Patients must be adequately informed about the use of AI technologies in their care, including understanding the potential implications for data privacy and decision-making processes. Clear communication and transparency are essential to empower patients to make informed choices about their healthcare.

As AI continues to permeate various sectors, several ethical considerations must be addressed to maintain and enhance trust like for data privacy and security, the collection and analysis of vast amounts of data by AI systems raise significant concerns about patient privacy and data security. Ensuring robust safeguards for data protection is crucial to maintain trust in AI systems, especially in sensitive areas like healthcare. Also the principle of informed consent becomes complex with AI applications. Patients and users must be adequately informed about the use of AI technologies, including data privacy implications and decision-making processes. Clear communication and transparency are essential to empower informed choices. AI systems can inadvertently perpetuate biases present in training data or algorithm design, leading to disparities in outcomes. Addressing algorithmic bias is imperative to ensure fairness and equity in AI applications. This involves understanding root causes of bias, developing mechanisms to detect and prevent bias, and promoting interdisciplinary collaboration to establish ethical guidelines.

USER PERSONAS –

1. William Carter

Demographics	Behaviors & Habits	Pain points	Needs & Goals
 Age: 29 Occupation: Engineering Graduate	Is a tech wizard and uses Artificial Intelligence on a daily basis.	Thinks a lot about his privacy and safety online and is very concerned related to these issues.	Wants his data to be safe and the relief that his data isn’t accessible by any other person or organization.

2. Selena Williams

Demographics	Behaviors & Habits	Pain points	Needs & Goals
 Age: 23 Occupation: Social media influencer	Is an online celebrity and stays online for most of the day. Accesses internet and AI sites quite often.	Highly conscientious about online privacy and safety, prioritizing proactive measures to safeguard personal information and mitigate digital risks.	Seeks assurance in the safety and confidentiality of her data, prioritizing stringent security measures for peace of mind.

3. Tiffany Cloak

Demographics	Behaviors & Habits	Pain points	Needs & Goals
 Age: 45 Occupation: Doctor	A doctor who to save her patient’s data uploads it onto an online server.	Always in a fix and very scared from the fear that the patient’s personal data be leaked online.	Wants a safe and secure pathway where she wouldn’t have to worry about her patient’s data.

Future Projections and Implications for this study can be in relation with the continued increase in trust. The linear regression model predicts that trust in AI will continue to rise, reaching approximately 80% by 2030. This positive outlook is contingent upon ongoing advancements in AI transparency, reliability, and user education. Also continued advancements in XAI and efforts to make AI decision-making processes more understandable will be crucial in maintaining and enhancing trust levels. Transparency initiatives must remain a priority for developers and researchers. The development and implementation of robust ethical standards and regulatory frameworks will be essential in sustaining trust. Policymakers need to focus on creating guidelines that ensure the responsible use of AI technologies. Promoting public education initiatives to increase awareness and understanding of AI capabilities and limitations will play a vital role in shaping future trust dynamics. Educators and industry stakeholders must collaborate to inform the public about the benefits and potential risks of AI.

VIII. CONCLUSION

This study has provided a comprehensive analysis of the factors influencing human trust in AI over the past 20 years. The findings highlight the significant increase in trust levels, driven by improvements in transparency, reliability, and user familiarity. As AI technologies continue to advance, maintaining and enhancing trust will be crucial for their successful adoption and integration into various sectors. By focusing on transparency, reliability, and user education, stakeholders can ensure that AI technologies are not only effective but also trusted by the public. This will pave the way for broader and more impactful applications of AI, ultimately benefiting society as a whole.

The study also projects future trust levels in AI. Using the fitted linear regression model, predictions for the next 10 years indicate a continued increase in trust. By 2030, trust levels are expected to be significantly higher than they are today. This projection assumes that the positive trends in transparency, reliability, and user familiarity will persist, and that AI technologies will continue to become more integrated into various aspects of life.

IX. FUTURE WORKS

Future research should continue to focus on improving AI transparency and reliability. Investigating new ways to enhance XAI and developing more robust performance metrics will be critical. Additionally, exploring the impact of demographic factors on trust in AI and examining the long-term effects of AI interaction on trust dynamics will provide valuable insights.

1. Longitudinal Studies on Trust Dynamics: Conduct long-term studies to observe changes in trust over extended periods. This would help understand how trust evolves as AI technologies advance and as people become more familiar with their applications.

2. Impact of Demographic Factors: Investigate how demographic factors such as age, education, socio-economic status, and cultural background influence trust in AI. This would provide insights into tailoring communication and education efforts to different groups.

3. Algorithmic Transparency and Explainability: Further develop and refine techniques for making AI algorithms more transparent and explainable. Research could explore new methods for visualizing AI decision processes and assessing their impact on user trust.

4. Mitigating Algorithmic Bias: Develop advanced methods to identify and mitigate biases in AI systems. This could include creating more diverse and representative datasets, as well as designing algorithms that are robust against bias.

5. Ethical AI Frameworks: Establish comprehensive ethical frameworks and guidelines for AI development and deployment. This involves interdisciplinary collaboration to address ethical, legal, and social implications of AI.

6. Trust in AI across Different Sectors: Expand research to include trust dynamics in various sectors beyond healthcare, such as finance, education, and transportation. Comparative studies could reveal sector-specific challenges and strategies for building trust.

7. Human-AI Collaboration: Investigate the optimal ways for humans and AI to collaborate effectively. Research could focus on dynamic task allocation, communication protocols, and decision-making processes to enhance trust and efficiency in human-AI teams.

8. AI Literacy and Public Awareness: Develop educational programs and public awareness campaigns to improve AI literacy. These initiatives should aim to demystify AI, address misconceptions, and provide clear information about the benefits and risks of AI technologies.

9. Regulatory and Policy Implications: Examine the impact of existing regulations and policies on trust in AI. Propose new regulatory measures that can enhance transparency, accountability, and ethical standards in AI development and deployment.

10. User Experience and Interface Design: Research how user experience (UX) and interface design influence trust in AI. This includes studying the effects of different interface elements, interaction modes, and feedback mechanisms on user perceptions and trust levels.

11. Measuring and Quantifying Trust: Develop standardized metrics and tools for measuring trust in AI. This would enable more consistent and comparable assessments of trust across different studies and applications.

12. Case Studies and Real-World Applications: Conduct case studies on real-world applications of AI to understand practical challenges and successes in building and maintaining trust. These case studies can provide valuable lessons and best practices for future AI implementations.

By pursuing these future research directions, the field can deepen its understanding of trust in AI and develop strategies to ensure that AI technologies are trusted and accepted by diverse user groups, ultimately leading to more successful and responsible AI integration into society.

X. REFERENCES

[1] Ella Glikson and Anita Williams Woolley, "Human Trust in Artificial Intelligence: Review of Empirical Research", 2020. [Academy of Management Annals]. Available: https://journals.aom.org/doi/full/10.5465/annals.2018.0057

[2] Ida Asadi Someh, Nicole Gillespie, Steven Lockey and Daniel Holm, "A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions", 2021. [AIS eLibrary]. Available: https://aisel.aisnet.org/hicss-54/os/trust/2/

[3] Mark Ryan, "In AI We Trust: Ethics, Artificial Intelligence, and Reliability", 2020. [Springer]. Available: https://link.springer.com/article/10.1007/s11948-020-00228-y#Sec14

[4] Alexandra D. Kaplan, Theresa T. Kessler, J. Christopher Brill and P. A. Hancock "Trust in Artificial Intelligence: Meta-Analytic Findings", 2021. [Sage Journals]. Available: https://journals.sagepub.com/doi/full/10.1177/00187208211013988

[5] Omri Gillath, Ting Ai, Michael S. Branicky, Shawn Keshmiri, Robert B. Davison and Ryan Spaulding "Attachment and trust in artificial intelligence", 2021. [ScienceDirect]. Available: https://www.sciencedirect.com/science/article/abs/pii/S074756322030354X

[6] Philipp Schimdt, Felix Biessmann and Timm Teubner "Transparency and trust in artificial intelligence systems", 2020. [Taylor & Francis]. Available: https://www.tandfonline.com/doi/full/10.1080/12460125.2020.1819094

[7] Alon Jacovi, Ana Marasović, Tim Miller and Yoav Goldberg, "Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI", 2021. [ACM Digital Library]. Available: https://dl.acm.org/doi/abs/10.1145/3442188.3445923

[8] Andrea Ferrario, Michele Loi and Eleonora Viganò, "In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions", 2019. [Springer]. Available: https://link.springer.com/article/10.1007/s13347-019-00378-3#Sec14

[9] John Zerilli, Umang Bhatt and Adrian Weller, "How transparency modulates trust in artificial intelligence", 2022. [50 Patterns]. Available: https://www.cell.com/patterns/fulltext/S2666-3899(22)00028-9

[10] Markus Langer, Cornelius J. König, Caroline Back and Victoria Hemsing, "Trust in Artificial Intelligence: Comparing Trust Processes Between Human and Automated Trustees in Light of Unfair Bias", 2022. [Springer]. Available: https://link.springer.com/article/10.1007/s10869-022-09829-9#citeas

[11] Onur Asan, Alparslan Emrah Bayrak and Avishek Choudhury, "Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians", 2020. [JMIR]. Available: https://www.jmir.org/2020/6/E15154

[12] Georgiana Juravle, Andriana Boudouraki, Miglena Terziyska and Constantin Rezlescu, "Trust in artificial intelligence for medical diagnoses", 2020. [ScienceDirect]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0079612320300819

[13] Roman Lukyanenko, Wolfgang Maass & Veda C. Storey, "Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities", 2022. [Springer]. Available: https://link.springer.com/article/10.1007/s12525-022-00605-4?trk=public_post_comment-text

[14] Till Saßmannshausen, Peter Burggräf, Johannes Wagner, Marc Hassenzahl, Thomas Heupel &Fabian Steinberg, "Trust in artificial intelligence within production management – an exploration of antecedents", 2021. [Taylor & Francis]. Available: https://www.tandfonline.com/doi/full/10.1080/00140139.2021.1909755

[15] Naomi Aoki, "The importance of the assurance that "humans are still in the decision loop" for public trust in artificial intelligence: Evidence from an online experiment", 2021. [ScienceDirect]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0747563220303204

[16] Saleh Afroogh, "A probabilistic theory of trust concerning artificial intelligence: can intelligent robots trust humans?", 2022. [Springer] Available: https://link.springer.com/article/10.1007/s43681-022-00174-4

[17] Christian Montag, Benjamin Klugah-Brown, Xinqi Zhou, Jennifer Wernicke , Congcong Liu , Juan Kou , Yuanshu Chen , Brian W. Haas and Benjamin Becker, "rust toward humans and trust toward artificial intelligence are not associated: Initial insights from self-report and neurostructural brain imaging", 2023. [Cambridge Core]. Available: https://www.cambridge.org/core/journals/personality-neuroscience/article/trust-toward-humans-and-trust-toward-artificial-intelligence-are-not-associated-initial-insights-from-

selfreport-and-neurostructural-brain-imaging/4F7A5B99D7B1CEEA429FA38816588AF5

[18] Atte Oksanen, Nina Savela, Rita Latikka and Aki Koivula, “Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human–Technology Interactions Online”, 2020. [Frontiers]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.568256/full

[19] Aurelia Tamò-Larrieux, Clement Guitton, Simon Mayer and Christoph Lutz, “Regulating for trust: Can law establish trust in artificial intelligence?”, 2023. [Wiley]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/rego.12568

[20] Olga Vl. Bitkina, Heejin Jeong, Byung Cheol Lee, Jangwoon Park, Jaehyun Park and Hyun K. Kim, “Perceived trust in artificial intelligence technologies: A preliminary study”, 2020. [Wiley]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hfm.20839