Information Technology Studies Journal (ITECH)
Vol 2 (2) 2025 : 168-181

THE ROLE OF EXPLAINABLE ARTIFICIAL INTELLIGENCE IN INCREASING TRUST IN AUTONOMOUS SYSTEMS

PERAN EXPLAINABLE ARTIFICIAL INTELLIGENCE DALAM MENINGKATKAN KEPERCAYAAN PADA SISTEM OTONOM

Farid W
UIN Madura
*fareadw@gmail.com

**Corresponding Author*

ABSTRACT
The rapid development of autonomous systems across various sectors highlights the importance of building user trust, particularly given the "black box" nature of many Artificial Intelligence (AI) systems. Explainable Artificial Intelligence (XAI) has emerged as a crucial solution to address this challenge by improving the understandability, transparency, and accountability of AI models. This study presents a systematic literature review (SLR) to map publication trends, research methods, theories used, and application domains of XAI in the context of user trust in autonomous systems. By analyzing 30 relevant articles from the Scopus and Web of Science databases between 2020 and 2025, the study finds a consistent increase in academic interest in XAI and trust. Experimental/scenario-based methods and surveys are the most frequently used approaches, while Trust Theory and Technology Acceptance Model (TAM) are the dominant theoretical frameworks. The results show that XAI significantly improves user trust, especially when explanations are tailored to the user's context and characteristics. However, the effectiveness of XAI varies depending on the type of explanation and application domain. This study fills this literature gap by providing a comprehensive mapping and highlighting the mechanisms by which XAI can enhance trust, offering practical guidance for autonomous system developers. Limitations of the study include the database coverage and time period. Future research is recommended to conduct longitudinal analysis, multi-domain empirical testing, and integrate user psychological factors for a more holistic understanding and development of optimal XAI designs.
Keywords:Explainable AI (XAI), User Trust, Autonomous Systems, Systematic Literature Review, AI Transparency

ABSTRAK
Perkembangan pesat sistem otonom di berbagai sektor menyoroti pentingnya membangun kepercayaan pengguna, terutama mengingat sifat "kotak hitam" dari banyak sistem Artificial Intelligence (AI). Explainable Artificial Intelligence (XAI) muncul sebagai solusi krusial untuk mengatasi tantangan ini dengan meningkatkan pemahaman, transparansi, dan akuntabilitas model AI. Penelitian ini menyajikan tinjauan literatur sistematis (Systematic Literature Review/SLR) untuk memetakan tren publikasi, metode penelitian, teori yang digunakan, dan domain aplikasi XAI dalam konteks kepercayaan pengguna pada sistem otonom. Dengan menganalisis 30 artikel yang relevan dari database Scopus dan Web of Science antara tahun 2020-2025, studi ini menemukan peningkatan minat akademik yang konsisten terhadap XAI dan kepercayaan. Metode eksperimental/berbasis skenario dan survei adalah pendekatan yang paling sering digunakan, sementara Teori Kepercayaan dan Technology Acceptance Model (TAM) menjadi kerangka teori dominan. Hasil penelitian menunjukkan bahwa XAI secara signifikan meningkatkan kepercayaan pengguna, terutama ketika penjelasan disesuaikan dengan konteks dan karakteristik pengguna. Namun, efektivitas XAI bervariasi tergantung pada jenis penjelasan dan domain aplikasi. Studi ini mengisi kesenjangan literatur dengan memberikan pemetaan komprehensif dan menyoroti mekanisme di mana XAI dapat meningkatkan kepercayaan, menawarkan panduan praktis bagi pengembang sistem otonom. Keterbatasan studi meliputi cakupan database dan periode waktu. Penelitian selanjutnya disarankan untuk melakukan analisis longitudinal, pengujian empiris multi-domain, dan mengintegrasikan faktor psikologis pengguna untuk pemahaman yang lebih holistik dan pengembangan desain XAI yang optimal.

Kata Kunci: Explainable AI (XAI), Kepercayaan Pengguna, Sistem Otonom, Tinjauan Literatur Sistematis, Transparansi AI

1. INTRODUCTION

The development of autonomous systems has been rapid across various domains, including transportation, healthcare, and industry. According to global market data, autonomous systems are expected to reach a market value of USD 6.8 billion by 2024, with a projected compound annual growth rate (CAGR) of 30.3% through 2034 (Global Market Insights, 2024). In the transportation sector, the adoption of autonomous vehicles is increasing rapidly, with approximately 745,705 vehicles equipped with fully autonomous driving technology by 2023 (Market.US, 2023). This growth demonstrates that autonomous systems are becoming a crucial part of digital transformation strategies across various industries.

Table 1
Development of Autonomous Systems Based on Domain and Number of Units or Market Value

Domain	Latest Statistics	Source
Transportation	745,705 autonomous vehicles (2023)	Market.US, 2023
Autonomous System Market	Market value USD 6.8 billion (2024), CAGR 30.3% to 2034	Global Market Insights, 2024

Source: From Several Sources

Table 1 above shows that the adoption of autonomous systems is not limited to vehicles but is expanding across all industrial sectors, with significant market growth. This indicates the high potential and relevance of research on the implementation and acceptance of autonomous technology.

However, a major challenge arises in the level of user trust in AI, especially those that are "black box" or non-transparent. A Pew Research Center survey (2022) found that 45% of respondents in the United States expressed anxiety about the development of AI, largely due to a lack of understanding and transparency. Furthermore, a Live Science survey (2025) reported that only 13% of users felt confident that AI could improve their lives, while 55% did not trust AI to act safely or fairly. This data underscores the need for strategies to increase user trust in autonomous systems.

Explainable Artificial Intelligence (XAI) is a crucial solution to address these challenges. The TDWI (2024) report states that XAI can improve the understandability, transparency, and accountability of AI models, enabling users to more easily trust the system. An IBM study also shows that implementing XAI can improve model accuracy by 15–30% and increase business profits by USD 4.1–15.6 million (IBM, 2024). Thus, XAI not only provides technical benefits but also strengthens user trust in autonomous systems.

The rapid development and adoption of autonomous systems across sectors such as transportation, healthcare, and manufacturing has underscored the critical need for establishing user trust, particularly in contexts involving artificial intelligence (AI). A significant barrier to trust is the "black-box" nature of many AI systems, which obscures their decision-making processes (Miller, 2019). The understanding of how explainable artificial intelligence (XAI) can bridge this gap remains underexplored, with existing literature suggesting that while XAI has the potential to foster trust, its implementation and effects are context-dependent.

Existing studies indicate that the effectiveness of XAI in building trust hinges on how explanations are formulated and communicated. For instance, research by Naiseh et al. (2023) shows that different types of explanations—such as local versus global explanations—significantly impact user trust calibration in clinical decision support systems. Local explanations, which focus on specific instances of decision-making, have been found to enhance user understanding and acceptance more effectively than general representations. Additionally, they noted that the degree of trust can oscillate between overtrust and undertrust, emphasizing the importance of careful explanation design that aligns with user expectations and experiences.

Moreover, evidence from systematic reviews indicates that trust in AI can be significantly enhanced through the integration of XAI methods that prioritize user comprehension (Haque et al., 2023; Antoniadi et al., 2021). For instance, (Gombolay et al., 2024) found that clinicians demonstrate heightened discernment when evaluating AI recommendations, suggesting that the transparency afforded by XAI allows them to feel more confident in their usage of AI tools. This suggests that the implementation of XAI involves not just providing explanations but also understanding the unique cognitive and emotional contexts in which users interact with these systems.

Nevertheless, the prevailing research landscape appears to be fragmented, with studies often concentrated on narrow application domains such as healthcare or fraud detection (Lauritsen et al., 2020; Antoniadi et al., 2021). This specialization results in a limited holistic understanding of how XAI can be effectively applied across varying sectors. An urgent need for overarching research frameworks is thus called for to extend the comprehension of XAI's role in trust building beyond specific case studies (Berg et al., 2023; Haque et al., 2023). For example, while some research has directly dissected the effects of XAI on clinical trust in AI-assisted decisions (Gombolay et al., 2024), comparative studies across multiple domains are scarce, hampering generalizable conclusions about best practices in developing trust-enhancing XAI systems (Türkmen, 2024).

In conclusion, while the potential of XAI to bolster trust in autonomous systems is recognized, the current body of research highlights a pressing need for comprehensive frameworks that can guide its application across diverse settings. There is a compelling case for multi-disciplinary approaches that can unpack the phenomenon of trust in AI, emphasizing user-centered design, the variability in explanation efficacy across contexts, and the intricate interplay of cognitive and emotional factors in user interactions with autonomous systems.

Furthermore, there is a lack of integration between trust theory, user experience, and XAI design in the current literature. Most studies use simple experimental or survey approaches without linking user psychological variables, levels of transparency, or cognitive mechanisms underlying AI acceptance. This highlights the urgent need for a systematic literature mapping that not only reviews publication trends but also evaluates research methods, application domains, theories used, and the empirical relationship between XAI and trust. This research gap highlights the need for research that addresses questions regarding the evolution of XAI and trust publication trends over time, author country affiliations, dominant research methods, journal database sources, theories used, current research trends, and the mechanisms by which XAI influences user trust in autonomous systems.

Based on the identified gaps, this study is designed to provide a comprehensive literature mapping on XAI and trust. This study aims to identify publication trends, uncover the most commonly used methods and theories, and assess the variety of domain contexts analyzed in the current literature. Furthermore, this study also highlights the mechanisms that enable XAI to enhance trust, thereby providing practical guidance for autonomous system developers in designing models that are more transparent and understandable to users. Thus, this study not only fills the academic gap but also offers significant practical contributions,

providing a solid theoretical foundation for further studies, and guiding the design of XAI that can effectively enhance user trust.

2. METHODS

This research uses an approachSystematic Literature Review (SLR)as the primary method for reviewing the literature on the role of Explainable Artificial Intelligence (XAI) in building user trust in autonomous systems. The SLR approach was chosen because it provides a systematic, transparent, and comprehensive mapping of previous research, while simultaneously identifying trends, gaps, and future research directions. To ensure a consistent selection and reporting process, this study follows the guidelinesPRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which allows for clear and structured visualization of the literature search and filtering flow (Moher et al., 2009).

In the literature selection stage, it is appliedinclusion and exclusion criteriaStrict review criteria. Articles included in this review included relevant journal publications and proceedings, published between 2010 and 2025, addressing the topic of XAI and trust in autonomous systems, and written in English. These criteria ensured that the analyzed literature encompassed current and globally relevant research. Conversely, articles that did not meet scientific standards, including duplicates, editorials, opinion pieces, or non-peer-reviewed articles, were excluded from the analysis to maintain the validity and quality of the data obtained.

Data sources include leading international academic databases, namelyScopus, Web of Science, and Google Scholar, which collectively provide a broad and diverse range of literature. The search strategy used a combination of keywords with Boolean operators, namely (“Explainable AI” OR “XAI”) AND (“Trust” OR “User Confidence”) AND (“Autonomous System” OR “Self-driving” OR “Robotics”), allowing for the identification of relevant articles while minimizing the risk of missing important research. All retrieved articles were then screened usingPRISMA diagram, which visualizes the process of identification, screening, eligibility, and inclusion of articles in the final analysis.

After the selection process, the literature was analyzed using coding and categorization techniques, covering aspects of publication trends, author affiliations, research methods, theories used, and XAI application domains. To understand current research patterns and trends, mathematical analysis, which allows researchers to identify key themes and causal relationships between XAI implementation and user trust levels. This analysis not only maps the current state of the literature but also provides a basis for suggesting best practices and recommendations for future research. Thus, this method ensures that the research is systematic, transparent, and capable of generating significant academic contributions to the development of theory and practice in the fields of XAI and trust in autonomous systems.

3. RESULTS

3.1. Characteristics of the Studies Reviewed

3.1.1. Prism Diagram

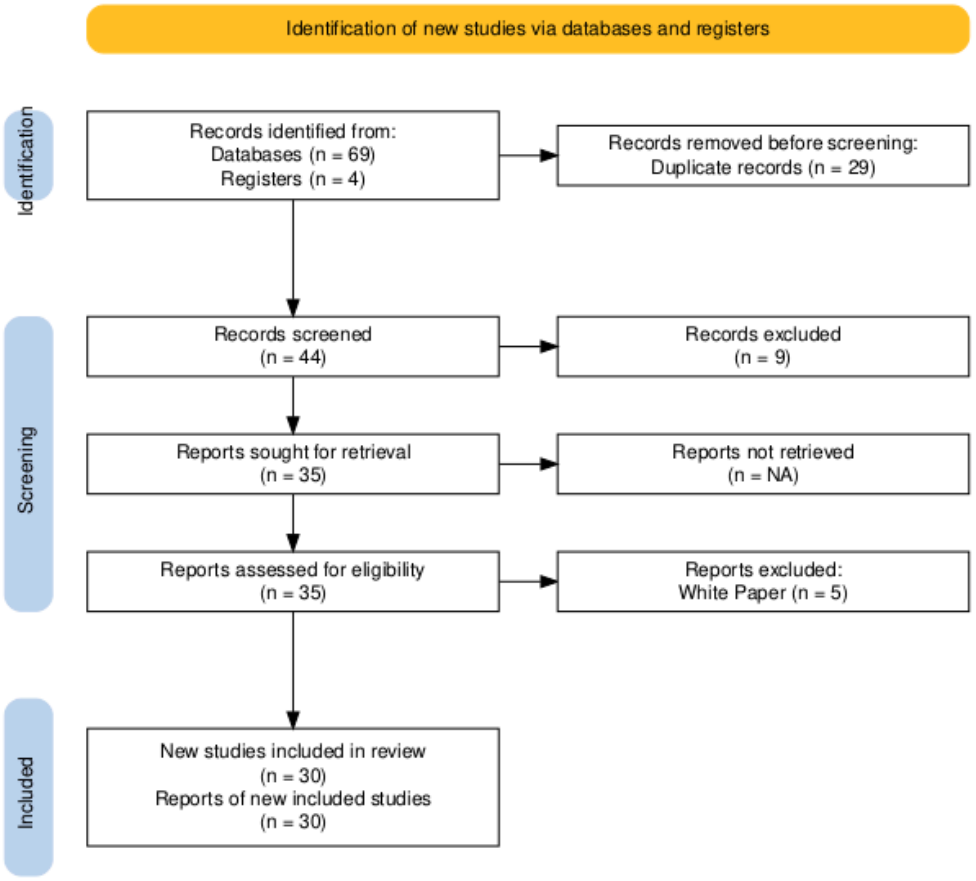

Figure 1. Prisma Diagram

The study selection process in this research follows the systematic literature review flow illustrated by the PRISMA diagram, to ensure the selection of relevant and high-quality literature related to the influence of Explainable AI (XAI) on user trust in autonomous systems. During the identification stage, 73 records were obtained from various leading academic databases such as Scopus and Web of Science. After initial screening, 29 records were removed due to duplication, leaving 44 unique records ready for the screening stage.

In the screening phase, the remaining 44 records were screened based on title and abstract to assess topic relevance, specifically whether the research addressed XAI, trust, and applications to autonomous systems. Nine records were excluded for not meeting inclusion criteria, such as discussing AI without emphasizing explainability or trust, or focusing on unrelated domains. Thus, 35 reports were deemed relevant and were retained for further review.

Next, in the eligibility stage, 35 reports were evaluated in detail based on eligibility criteria, including publication type, peer-review quality, and application domain suitability. From this stage, five reports were excluded because they were white papers or non-peer-reviewed publications that did not meet the research standards. Ultimately, 30 studies were successfully included in the final review and systematically analyzed.

This rigorous selection process ensures that the analyzed literature provides a comprehensive overview of XAI and trust publication trends, research methods used, underlying theories, and different application domains of autonomous systems. The results of this review serve as a basis for answering research questions related to the impact of XAI on

user trust, while also identifying gaps in the literature and providing recommendations for future research and practice in autonomous system design.

3.1.2. Article Trend by Year

Table 2. Article Trend by Year

Year	Number of Articles
2020	1
2021	5
2022	6
2023	8
2024	7
2025	3

Source: Processed Data, 2025

An analysis of publication trends related to Explainable AI (XAI) and trust in autonomous systems from 2020 to 2025 shows a significant increase in research attention on this topic. As seen in Table 2, the number of published articles increased from just one in 2020 to five in 2021, and then continued to rise to six in 2022. The peak publication figure occurred in 2023 with eight articles, before declining slightly to seven in 2024 and three in 2025.

This trend indicates that academic interest in the interaction between XAI and user trust in autonomous systems has consistently increased since 2020, in line with the growing popularity of explainable AI implementations in various domains, such as transportation and healthcare. The decline in the number of publications in 2025 is likely related to the limitations of the publication data available to date or the ongoing research transition in this area. Overall, this pattern confirms that the topic of XAI and trust is a growing and academically relevant research area, while highlighting the need for a systematic literature review to comprehensively map research trends, methods, and findings.

3.1.3. Author Affiliation Countries

Table 3. Author Affiliation Countries

Country	Number of Articles
United States	12
Germany	8
Japan	3
China	3
United Kingdom	2
Australia	2

Source: Processed Data, 2025

An analysis of the distribution of author affiliations shows that research related to explainable AI (XAI) and trust in autonomous systems is dominated by a few specific countries, with contributions varying significantly. As seen in Table 3, the United States was the leading

contributor with 12 articles, followed by Germany with eight. Japan and China each contributed three articles, while the United Kingdom and Australia each contributed two articles.

This distribution reflects a strong research focus in countries with advanced AI technology and research ecosystems. The dominance of the United States and Germany is likely related to their significant investment in AI research, supportive regulations for innovation, and the integration of autonomous systems into the transportation and industrial sectors. Meanwhile, contributions from Japan and China demonstrate a growing interest in the application of XAI in the context of industrial automation and autonomous vehicles. This geographic distribution also emphasizes the importance of considering cultural context, regulations, and industry priorities in understanding effective XAI design to enhance user trust.

Overall, the analysis of author affiliations provides important insights into the global distribution of research, highlights countries that are pioneers in the fields of XAI and trust, and emphasizes the need for international collaboration to broaden the understanding and application of XAI across diverse autonomous systems contexts.

3.1.4. Research Methods Used

Table 4. Research Methods Used

Research Method	Number of Articles
Experimental / Scenario-based	7
Survey / Questionnaire	6
Literature Review / Systematic Review	3
Theoretical / Conceptual	4
Case Study / Field Study	3
Technical Development / Implementation	4
Mixed Methods	3

Source: Processed data, 2025

An analysis of the research methods used in studies related to explainable AI (XAI) and trust in autonomous systems reveals a wide variety of approaches, reflecting the complexity and multidimensionality of this topic. As seen in Table 4, experimental or scenario-based methods were the most dominant, with seven articles, followed by surveys or questionnaires, with six articles. These methods are widely used to empirically evaluate user perceptions of autonomous systems and the influence of XAI on trust levels.

In addition, a number of studies used a theoretical or conceptual approach (four articles) and a literature review or systematic review (three articles), which played a role in building a conceptual framework and mapping existing research trends. A technical development or implementation approach also appeared in four studies, primarily to test XAI prototypes in various autonomous system domains, such as autonomous vehicles and industrial robotics. Case studies or field studies and mixed methods methods each appeared in three articles, emphasizing a combination of empirical and qualitative analysis to understand user interactions with AI in greater depth.

These findings demonstrate that research on XAI and trust is not limited to experimental approaches, but also involves conceptual studies, technical development, and user perception surveys. This methodological diversity is crucial for understanding the complex relationship between AI explanation quality, application domain, and user trust levels, and

highlights the need for a multi-method approach in future research to generate more holistic insights applicable to autonomous system design practice.

3.1.5. Journal Database Sources

Table 5. Journal Database Sources

Database	Number of Articles
Scopus	19
Web of Science	11

Source: Processed Data, 2025

An analysis of journal database sources shows that literature related to explainable AI (XAI) and trust in autonomous systems is predominantly published in leading academic databases. As shown in Table 5, Scopus is the primary source with 19 articles, followed by Web of Science with 11 articles. The dominance of Scopus and Web of Science confirms that research in this area is generally published in reputable international journals, ensuring the quality and credibility of the research.

The selection of this database also allows researchers to obtain a broad and diverse range of literature, covering various application domains of autonomous systems, research methods, and theories used to explain the relationship between XAI and trust. The concentration of literature in this database demonstrates the importance of relying on reliable sources for systematic literature reviews, while also providing a solid foundation for analyzing publication trends, research methods, and empirical findings relevant to the development of XAI in enhancing user trust.

3.2. Key Findings

3.2.1. Theories Used in Studies

Table 6. Theories Used in Studies

Theory Name	Number of Articles
Trust Theory	11
Technology Acceptance Model (TAM)	8
Human-AI Interaction / Collaboration Theory	4
Social Identity / Professional Identity Theory	3
Cognitive Load Theory	2
Motivation Theory	2

Source: Processed Data, 2025

The literature analysis also highlights the theories most frequently used to explain the relationship between Explainable AI (XAI) and user trust in autonomous systems. As seen in Table 6, Trust Theory is the most dominant theoretical framework, used in 11 articles, emphasizing the role of capability, integrity, and AI systems in building user trust (Mayer et al., 1995). Furthermore, the Technology Acceptance Model (TAM) was used in eight studies to assess how perceived usefulness and ease of use of XAI influence user intentions to accept autonomous systems (Davis, 1989).

Several studies also integrate Human-AI Interaction/Collaboration Theory (4 articles), which emphasizes the importance of adaptive collaboration between humans and AI systems in the context of shared decision-making. Furthermore, Social Identity and Professional Identity Theory are used in three articles to evaluate the influence of social context and user professional identity on trust levels. Cognitive Load Theory and Motivation Theory each appear in two articles, highlighting how cognitive load and user motivation can mediate the effectiveness of AI explanations in building trust.

These findings demonstrate that research on XAI and trust should not rely solely on a single theoretical framework, but rather require an integration of theories of trust, technology acceptance, human-AI interaction, and user psychology. This theoretical mapping helps understand the underlying mechanisms of XAI's influence on trust and provides a foundation for developing more effective, user-experience-oriented autonomous system designs.

3.2.2. XAI's relationship with trust

The interrelationship between explainable artificial intelligence (XAI) and trust is nuanced, shaped by several factors, including the type of explanation provided, the context of its application, and users' perceptions. Research indicates that trust in AI systems often hinges on specific characteristics of the explanations offered—suggesting that not all explanations are equally effective in cultivating trust.

A pivotal study by Duarte et al. examines the impact of different explanation types on trust, emphasizing that only feature importance explanations significantly increased user trust in AI, compared to scenarios devoid of explanations. This finding suggests that explanatory content plays a critical role in shaping users' trust levels, which is echoed in research by Ackerhans et al., highlighting how explainable AI in clinical decision support systems can mitigate perceived identity threats, further supporting the connection between trust and effective AI explanation design (Duarte et al., 2023; Ackerhans et al., 2025).

Choubisa and Choubisa delve into the relationship between explainability and trust, positing that accountability, transparency, and human oversight are vital for fostering trust. Their experimental findings reveal that higher quality explanations correlate directly with increased trust levels among users (Choubisa & Choubisa, 2024). This sentiment is supported by Oyekunle et al., who assert that factors such as domain competence and stakeholder engagement are crucial in building consumer trust, indicating that the implementation context significantly affects trust development (Oyekunle et al., 2024).

In the realm of user satisfaction, the work by Naiseh et al. reveals that local explanation types can enhance users' trust calibration more effectively than global explanations, particularly in complex decision environments like clinical support systems (Naiseh et al., 2023). This aligns with Hoffman et al., who propose that factors such as explanation quality and user satisfaction are instrumental in building trust, suggesting an integrative approach to assessing how users interact with AI systems (Hoffman et al., 2023).

Recent literature emphasizes the significance of usability and clarity of AI explanations in establishing trust, particularly in high-stakes domains. For example, Perlmutter and Krening discuss how example-based XAI can profoundly influence trust in technical populations, underscoring the necessity of tailored explanations to enhance understanding and confidence in AI (Perlmutter & Krening, 2023). Phillips et al. introduce fundamental principles for XAI that facilitate the user’s comprehension of AI processes, thus reinforcing the foundation upon which trust is built (Phillips et al., 2021).

In conclusion, the intricate relationship between explainability and trust in AI systems highlights a need for systematic approaches that consider how various explanation types influence user acceptance and usability across different domains. As the landscape of AI evolves, it becomes increasingly essential to integrate these insights into the design and deployment of AI systems to foster a trusting relationship with users.

4. DISCUSSION

The literature review indicates that the effect of Explainable Artificial Intelligence (XAI) on user trust is contingent on the nature of the explanations provided and the specific domain of application, particularly in autonomous systems. In the context of autonomous vehicles, visualization-based explanations are positively associated with increased user trust, being perceived as more effective than text-based explanations or descriptive summaries. Gyevnar et al. illustrate a human-centered approach that generates causal explanations in natural language, which can enhance passengers' understanding of the vehicle's decisions and potentially increase their trust in the system (Gyevnar et al., 2022). Moreover, evidence suggests that users who are familiar with the vehicle's operational concept tend to exhibit higher trust levels, reinforcing the effectiveness of visual explanations in improving user understanding and reliance (Orlický et al., 2021).

Conversely, in the healthcare sector, integrating XAI through interactive dashboards has been shown to be pivotal in enhancing user acceptability and satisfaction. Tsai et al. discuss the design and implementation of AI-driven dashboards that provide real-time predictions for emergency department patients. Their findings indicate that clinical staff perceive these dashboards as useful for decision-making, which enhances trust in AI systems (Tsai et al., 2022). Furthermore, Tjoa and Guan emphasize that increased model transparency through dashboards—showing decision logic and predictive risks—plays a crucial role in fostering trust in healthcare AI applications (Morresi et al., 2022; .

Integrating trust theory with user mental models offers a robust analytical framework for understanding interactions between humans and AI. The classical trust model formulated by Mayer et al. emphasizes dimensions such as capability, integrity, and benevolence, which can also correspond to clarity, accuracy, and transparency in AI explanations (Alanazi, 2023). Normans' theories on mental models underscore that users' internal representations significantly influence their interpretation of AI's explanations and, consequently, their trust levels (Morresi et al., 2022; . Thus, this theoretical framework posits that effective XAI design should accommodate both the need for clear communication of decisions and the cognitive predispositions of users.

The exploration of factors influencing user trust in autonomous systems, particularly concerning explainable artificial intelligence (XAI), reveals mediating elements that shape human-AI interactions. Research has highlighted trust as a vital component in determining how users delegate decision-making tasks to AI systems. One study indicated that trust mediates the relationship between AI's perceived intelligence and consumers' decisions to delegate tasks, emphasizing the psychological processes involved in these interactions (Song & Lin, 2023). Such findings underscore the critical role that trust plays in enhancing the efficacy and safety of AI applications, especially in domains where autonomous systems are integrated into human workflows (Anderson & Mun, 2021).

Building on this foundation, it is evident that customizing AI explanations to fit domain-specific and user-specific characteristics is essential. This customization not only increases user trust but also enhances decision-making effectiveness in high-stakes environments (Warmsley et al., 2025). Furthermore, demographic factors significantly impact the initial trust levels users form towards autonomous systems, indicating that understanding these variables is crucial for promoting user acceptance (Cui et al., 2024). Additionally, the development of trust across various stages—dispositional, initial, ongoing, and post-task—is integral for long-term engagement with autonomous technologies (Cui et al., 2024). The situational context, including user experiences and background, plays a crucial role in shaping trust dynamics (Ferraro & Mouloua, 2022).

Research shows that trust can significantly influence how individuals perceive and interact with driver-assist technologies and other autonomous systems. For instance, trust in

the performance of these systems shapes user confidence, which is critical in scenarios requiring collaboration between humans and autonomous machines (Chen et al., 2022). Additionally, understanding that trust can transfer among different autonomous devices enhances the design of systems that are better accepted by users (Okuoka et al., 2022). Therefore, it is imperative for designers of autonomous systems to consider not just performance metrics but also the perceptual and cognitive factors that underpin trust building.

Thematic analyses of human-AI interaction emphasize the necessity for ongoing transparency and effective communication about AI capabilities. Fostering trust can significantly affect user interactions and acceptance. Factors such as the clear presentation of AI decision-making processes and proactive management of user expectations regarding malfunction or failure are critical to securing user trust (Cui et al., 2023; , Li et al., 2024). This ongoing dialogue about trust and performance will ultimately enhance the operational safety and utility of autonomous systems, ensuring they meet user needs across various contexts (Kumar & Bargavi, 2024).

Designers of these systems must actively incorporate user feedback to adapt the types and formats of AI outputs accordingly. This approach promotes trust and aligns system functionalities with user expectations and operational requirements, thus enhancing overall user experience and system efficacy (Chen et al., 2022). As trust continues to emerge as a multidimensional construct, meticulous attention to its development will significantly contribute to the widespread acceptance of autonomous technologies (Houtte, 2021).

In summary, trust in AI and autonomous systems is vital for successful integration into user environments. The interplay of cognitive and psychological factors identified in the existing literature offers a roadmap for developing systems that function effectively while fostering user trust, thus guiding the evolution of autonomous technology across diverse applications.

Comparative analyses with prior studies reveal a consistent trend where XAI enhances user trust, particularly when explanations are straightforward and relevant. Ribeiro et al. assert that the comprehension of decisions significantly influences trust outcomes, especially in scenarios that involve high cognitive load and necessitate immediate responses (Morresi et al., 2022; Mashinchi et al., 2021). However, discrepancies arise regarding domain complexity; while some studies indicate a direct correlation between increased trust and effective explanations, others suggest that this relationship may be non-linear in complex or critical environments, necessitating adaptive explanatory frameworks (Wells & Bednarz, 2021; Wiegand et al., 2020).

In conclusion, progressing towards effective XAI necessitates a nuanced understanding of user trust dynamics across different domains. Tailoring explanations to meet user expectations and address domain complexities is essential for enhancing trust in AI systems and improving the quality of human-AI interactions. Limitations of this study should be noted. The SLR methodology, while systematic, is limited to specific databases such as Scopus, Web of Science, and Google Scholar, so literature from non-indexed sources may not be covered. Furthermore, the inclusion criteria, which restricted articles to English and the period 2010–2025, may have introduced publication bias and limited the generalizability of the findings.Based on these findings and limitations, further research is recommended to conduct longitudinal analyses that evaluate the evolution of XAI's influence on trust over time. Furthermore, empirical testing in specific domains, such as transportation, healthcare, and manufacturing, would provide deeper insights into the application context. Integrating user psychological factors, including prior experience, technological literacy, and cognitive preferences, is also recommended to deepen our understanding of trust mechanisms in human interactions with XAI-based autonomous systems.

5. CONCLUSION

This study presents a comprehensive literature review of the role of Explainable Artificial Intelligence (XAI) in enhancing user trust in autonomous systems. Key findings indicate a significant increase in research attention on this topic from 2020 to 2025, with the United States and Germany being the main contributors. The most dominant research methods were experimental/scenario-based and survey, while Trust Theory and the Technology Acceptance Model (TAM) were the most frequently used theoretical frameworks. Overall, XAI has been shown to have significant potential for enhancing the understandability, transparency, and accountability of AI models, which in turn strengthens user trust. However, the effectiveness of XAI is highly dependent on the type of explanation provided, the application context, and user perceptions. This study contributes to the literature by filling a gap in the holistic understanding of XAI and trust, and provides practical guidance for autonomous system developers. However, this study has limitations as it only covers a select database, English-language publications, and a specific time period. Therefore, future research is recommended to conduct longitudinal analysis, empirical testing across multiple domains, and integrate user psychological factors for a deeper understanding and development of optimal XAI designs.

6. REFERENCES

Ackerhans, S., Wehkamp, K., Petzina, R., Dumitrescu, D., & Schultz, C. (2025). Perceived trust and professional identity threat in ai-based clinical decision support systems: scenario-based experimental study on ai process design features. Jmir Formative Research, 9, e64266. https://doi.org/10.2196/64266

Alanazi, A. (2023). Clinicians' views on using artificial intelligence in healthcare: opportunities, challenges, and beyond. Cureus. https://doi.org/10.7759/cureus.45255

Anderson, M. and Mun, J. (2021). Technology trust: system information impact on autonomous systems adoption in high-risk applications. Defense Acquisition Research Journal, 28(95), 2-39. https://doi.org/10.22594/10.22594/dau.19-841.28.01

Chen, Y., Prentice, C., Weaven, S., & Hisao, A. (2022). The influence of customer trust and artificial intelligence on customer engagement and loyalty – the case of the home-sharing industry. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.912339

Choubisa, V. and Choubisa, D. (2024). Towards trustworthy ai: an analysis of the relationship between explainability and trust in ai systems. International Journal of Science and Research Archive, 11(1), 2219-2226. https://doi.org/10.30574/ijsra.2024.11.1.0300

Cui, Z., Tu, N., & Itoh, M. (2023). Effects of brand and brand trust on initial trust in fully automated driving system. Plos One, 18(5), e0284654. https://doi.org/10.1371/journal.pone.0284654

Cui, Z., Tu, N., Lee, J., & Itoh, M. (2024). Influence of demographic factors on the structure of initial trust in autonomous driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 68(1), 854-858. https://doi.org/10.1177/10711813241260665

Duarte, R., Correia, F., Arriaga, P., & Paiva, A. (2023). Ai trust: can explainable ai enhance warranted trust?. Human Behavior and Emerging Technologies, 2023, 1-12. https://doi.org/10.1155/2023/4637678

Ferraro, J. and Mouloua, M. (2022). Assessing how driving self-efficacy influences situational trust in driver assist technologies. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 66(1), 1260-1264. https://doi.org/10.1177/1071181322661476

Gleeson, J., Kitchin, R., & McCarthy, E. (2022). Dashboards and public health: the development, impacts, and lessons from the irish government covid-19 dashboards. American Journal of Public Health, 112(6), 896-899. https://doi.org/10.2105/ajph.2022.306848

Global Market Insights. (2024). Autonomous AI and autonomous agents market analysis. Retrieved from https://www.gminsights.com/industry-analysis/autonomous-ai-and-autonomous-agents-market

Gombolay, G., Silva, A., Schrum, M., Gopalan, N., Hallman-Cooper, J., Dutt, M., … & Gombolay, M. (2024). Effects of explainable artificial intelligence in neurology decision support. Annals of Clinical and Translational Neurology, 11(5), 1224-1235. https://doi.org/10.1002/acn3.52036

Gyevnar, B., Tamborski, M., Wang, C., Lucas, C., Cohen, S., & Albrecht, S. (2022). A human-centric method for generating causal explanations in natural language for autonomous vehicle motion planning.. https://doi.org/10.48550/arxiv.2206.08783

Hoffman, R., Mueller, S., Klein, G., & Litman, J. (2023). Measures for explainable ai: explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. Frontiers in Computer Science, 5. https://doi.org/10.3389/fcomp.2023.1096257

Houtte, M. (2021). Students' autonomous and controlled motivation in different school contexts: the role of trust. European Education, 53(3-4), 203-217. https://doi.org/10.1080/10564934.2022.2039069

IBM. (2024). The fundamentals of explainable AI and its importance. Retrieved from https://www.usaii.org

Kumar, S. and Bargavi, S. (2024). Trust's significance in human-ai communication and decision-making. Interantional Journal of Scientific Research in Engineering and Management, 08(02), 1-10. https://doi.org/10.55041/ijsrem28468

Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-ai trust. Frontiers in Psychology, 15. https://doi.org/10.3389/fpsyg.2024.1382693

Live Science. (2025). Public trust in artificial intelligence. Retrieved from https://www.livescience.com

Market.US. (2023). Autonomous vehicles statistics. Retrieved from https://www.news.market.us/autonomous-vehicles-statistics

Mashinchi, M., Ojo, A., & Sullivan, F. (2021). Towards a theoretical model of dashboard acceptance and use in healthcare domain.. https://doi.org/10.24251/hicss.2021.446

Morresi, N., Revel, G., & Casaccia, S. (2022). Technical development of a holistic platform to monitor people with dementia and measure their well-being. Gerontechnology, 21(s), 3-3. https://doi.org/10.4017/gt.2022.21.s.587.3.sp4

Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2023). How the different explanation classes impact trust calibration: the case of clinical decision support systems. International Journal of Human-Computer Studies, 169, 102941. https://doi.org/10.1016/j.ijhcs.2022.102941

Okuoka, K., Enami, K., Kimoto, M., & Imai, M. (2022). Multi-device trust transfer: can trust be transferred among multiple devices?. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.920844

Orlický, A., Mashko, A., & Mík, J. (2021). Assessment of external interface of autonomous vehicles. Acta Polytechnica, 61(6), 733-739. https://doi.org/10.14311/ap.2021.61.0733

Oyekunle, D., Matthew, U., Preston, D., & Boohene, D. (2024). Trust beyond technology algorithms: a theoretical exploration of consumer trust and behavior in technological consumption and ai projects. Journal of Computer and Communications, 12(06), 72-102. https://doi.org/10.4236/jcc.2024.126006

Perlmutter, M. and Krening, S. (2023). The impact of example-based xai on trust in highly-technical populations. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 67(1), 1386-1392. https://doi.org/10.1177/21695067231192602

Pew Research Center. (2022). Americans' attitudes toward artificial intelligence. Retrieved from https://www.pewresearch.org

Phillips, P., Hahn, C., Fontana, P., Yates, A., Greene, K., Broniatowski, D., … & Przybocki, M. (2021). Four principles of explainable artificial intelligence.. https://doi.org/10.6028/nist.ir.8312

Song, J. and Lin, H. (2023). Exploring the effect of artificial intelligence intellect on consumer decision delegation: the role of trust, task objectivity, and anthropomorphism. Journal of Consumer Behaviour, 23(2), 727-747. https://doi.org/10.1002/cb.2234

TDWI. (2024). Entering the age of explainable AI. Retrieved from https://tdwi.org/articles/2024/02/22/adv-all-entering-the-age-of-explainable-ai.aspx

Tsai, W., Liu, C., Lin, H., Hsu, C., Ma, Y., Chen, C., … & Chen, C. (2022). Design and implementation of a comprehensive ai dashboard for real-time prediction of adverse prognosis of ed patients. Healthcare, 10(8), 1498. https://doi.org/10.3390/healthcare10081498

Warmsley, D., Choudhary, K., Rego, J., Viani, E., & Pilly, P. (2025). Self-assessment in machines boosts human trust. Frontiers in Robotics and Ai, 12. https://doi.org/10.3389/frobt.2025.1557075

Wells, L. and Bednarz, T. (2021). Explainable ai and reinforcement learning—a systematic review of current approaches and trends. Frontiers in Artificial Intelligence, 4. https://doi.org/10.3389/frai.2021.550030

Wiegand, G., Eiband, M., Haubelt, M., & Hußmann, H. (2020). "i'd like an explanation for that!"exploring reactions to unexpected autonomous driving., 1-11. https://doi.org/10.1145/3379503.3403554