A review of AI-based trust management in smart cities

Jessica Ohnesorg[1], *Nazek* Fakhoury[1], *Noura* Eltahawi[1], and *Mouzhi* Ge[1,*]

[1]Deggendorf Institute of Technology, Deggendorf, Germany

Abstract. Given the complexity of trust management in smart cities, this work unfolds the important role of trust management across various domains. Be- yond its traditional roots in human relationships, trust management emerges as a cornerstone in technological, business, and societal contexts. This scoping review first organizes the literature by five most commonly cited indicators, and then derives essential insights from existing literature. This paper not only nav- igates through the challenges presented by technological advancements in trust management, but also offers a comprehensive understanding of the mechanisms and frameworks shaping trust in smart cities.

1 Introduction

In the new era of urbanization, smart cities have emerged as dynamic hubs where cutting-edge technologies are integrated to redefine the urban living [1]. During the digital transformation, these cities evolve into complex ecosystems that demand a sophisticated approach to gov- ernance and technological integration. Within this paradigm, trust management is a critical component with blending artificial intelligence (AI) such as deep learning [2].

Trust management underpins the successful deployment and widespread acceptance of AI-driven solutions within the framework of smart cities. This introduces an exploration of the multifaceted landscape of trust management in these urban environments, where AI serves not only to explore the potential for innovation and efficiency but also creates new challenges. These challenges span the domains of security and privacy, transparency, authen- ticity, communication, and reliability. However, the integration of AI also raises challenging issues related to the security and privacy of citizen data, the transparency of decision-making processes, the authenticity of digital interactions, effective communication strategies, and the reliability of AI-driven systems [3].

This paper therefore navigates through the literature of trust management in smart cities and unravels the implications of AI for trust management. The contribution of this work is framed as a set of research challenges that cover individual preference, transparency, interpre- tation, system reliability and security. This work is to target the question of which research challenges in AI-based trust management should be addressed in smart cities. Beyond the role as a technological enabler, AI becomes a catalyst for reshaping urban interactions and

*e-mail: mouzhi.ge@th-deg.de

governance structures. Furthermore, internet of things simultaneously allow machines, users and sensors to communicate and exchange data with each other over the internet.

Understanding the dynamics of trust in this context is paramount for establishing a resilient and harmonious coexistence between technology, governance, and the urban fabric. By addressing the challenges and potentials of intelligent systems in smart cities, the exploration intends to indicate the future trajectory of urban development. In an era dominated by AI-driven advancements, the trajectory of smart cities depends on the effective manage- ment of trust, which creates a landscape where innovation thrives while citizens' privacy is protected.

The rest of the paper is organized as follows. Section 2 describes the methodology that is used to search and collect paper for our scoping review. Based on the selected papers, section 3 classifies the papers with five criteria. This is then used to derive the insights for trust management in smart cities. From our review results, section 4 proposes a set of research challenges that can be used a research agenda for the research of trust management in smart cities. Finally, section 5 concludes the paper and outlines future research.

2 Methodology

The research methodology adopted in this study involved a comprehensive exploration of research resources, utilizing well-known databases such as Google Scholar, PubMed, and IEEE Xplore. This investigative process commenced with the formulation of carefully selected keywords, encompassing essential themes such as trust management, smart cities, healthcare, and AI. A stringent inclusion criterion was applied, ensuring that identified sources were available in either English or German language and provided open access to full-text articles. To enhance search precision, advanced techniques such as Boolean operations were systematically applied, and a meticulous examination was conducted across titles, abstracts, and complete articles. Importantly, the selection process maintained an inclusive approach regarding publication years, seeking to offer a comprehensive perspective on the subject while maintaining a focused exploration of the most recent investigations.

3 Trust Management in Smart Cities

As trust management has been applied across different domains in smart cites such as interpersonal relationships, business, and technology, it becomes imperative to comprehend and handle trust. This understanding is fundamental for fostering collaboration, establishing partnerships, and ensuring the success of diverse endeavors. In this context, exploring the critical dimensions of trust management becomes critical in smart cities.

After conducting a comprehensive examination of existing research on trust management, this paper uses Table 1 to present an overview considering most relevant previous investigations by highlighting five frequently mentioned aspects in this area. The key aspects such as security and privacy, transparency, authenticity, communication, and reliability were identified. The varied shades within the circles symbolize the degree of content associated with each specific aspect. In this overview, a completely white circle indicates a lack of coverage for the identified aspect, while an entire black-shaded symbol signifies a direct emphasis on the dimension discussed in the correlated research paper. Therefore, this visual representation offers a comprehensive depiction of the diverse content domains encapsulated by the critical aspects.

Although the identified aspects may appear self-evident, it is important to delve into a comprehensive description of the conceptualization of each aspect. Security and privacy

emerge prominently as pivotal factors in the establishment of trust, assuming a critical role in protecting individuals' data within intelligent systems. Their significance lies in the prevention of illicit breaches and unauthorized use, thereby reinforcing the foundational elements of trust in the context under consideration.

With regard to transparency, we underscored the importance of fostering clarity and openness within intelligent systems. This encompasses a detailed consideration of various elements, including explaining the processes of data utilization, delineating the objectives of data collection, explicating the operational mechanisms inherent in intelligent systems, and describing their overall productivity. Furthermore, we extended their emphasis to include the establishment of clear policies and delineation of data ownership as integral components of transparency.

Regarding the concept of authenticity, its definition encapsulates the ability to authenticate and authorize actions impacting individuals and their data. Further, authenticity is a multifaceted concept that includes the broader notion of trustworthiness and reliability in various aspects of interactions and transactions. Also, authenticity involves not only validating the legitimacy of an individual's identity but also ensuring the integrity of associated data.

In this context, communication is important in how individuals consume critical information, which plays a pivotal role in the interactions and trust in intelligent systems. It involves effectively conveying relevant details, ensuring individuals to be comfortable and confident in engaging with these systems. Also, reliability pertains to individuals' understanding of the efficiency, precision, and dependability demonstrated by these intelligent systems.

Table 1. Overview of content coverage for selected trust management indicators

	Security and privacy	Transparency	Authenticity	Communication	Reliability
[4]	◓	◔	○	◔	○
[5]	●	○	●	●	◔
[6]	●	○	○	◓	◔
[7]	◐	◔	○	○	◓
[8]	◓	◔	○	○	◓
[9]	○	◓	○	◓	◓
[10]	●	○	●	○	◔
[11]	◓	●	○	●	○
[12]	◔	○	●	○	◓

Based on the literature review, we found that most articles highlighted the research significance of privacy and security. Thus, from Table 1, it can be seen that this topic is one of the most relevant aspects in terms of trust and being trustworthy. Concerning this topic, an investigation in [4] involving 500 companies in the product industry stated that nearly 75 percent of these companies currently lack a well-developed strategy for addressing the cyber risks linked to data-intensive development processes. Substantial investments are needed in both cyber security measures and protecting sensitive data.

In the context of supply chain security and risk management, the term "trustworthiness" is defined by the supplier‘s capability to reliably fulfill the expectations of the potential contract partner in a verifiable manner. Furthermore, security, viewed as an attribute of an organization's trustworthiness, has gained research significance in light of the escalating value of information. The secure sharing of information among participants in the global supply

chain has become a critical factor, especially when stakeholders seek to exploit vulnerabilities within the supply chain, potentially compromising the product or its constituent components.

An additional observation, centered on issues related to security and privacy, indicates that vehicular ad hoc networks (VANETs) are configured to incorporate two forms of communication—specifically, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, where the problem stays in securing the communication channels. Also, privacy in VANET entails that exclusive individuals possess the right to access and control vehicle information, encompassing the genuine identity of the vehicle and its location profile [5].

Moreover, alternative perspectives presented a trust management framework integrating security measures to ensure consumer satisfaction. The discussions within these articles explored how to implement such methods to enhance trust among system users. The encryption techniques implemented in the Trust-FTSR model not only protect the system from malicious attacks but also increase the tolerable fault in case of unpredictable events [10]. Consequently, applying security as an inherent component of the organization's overall trustworthiness is crucial [5, 6]. This finding motivates the essential need for investments in cyber security and the protection of sensitive data [4].

In various contexts, multiple scholars highlight transparency and communication as fundamental pillars for establishing trust, particularly concerning public trust. One specific study introduced a methodology emphasizing a thorough understanding to effectively communicate vital information, thereby enhancing public trust in smart cities. Recognizing the critical role of transparency in conveying essential data, we propose to foster a shared understanding and active community engagement during decision-making processes [9]. Some work such as [11] pointed that having the opportunity to ask question and receive efficient, distinct answers from the system directors is highly valued by the users. Besides, [10] proposed mechanism enforced trust evaluation at the end of any procedure, this leads to trustworthy communication.

Another approach to ensure trust and a facet of an organization's trustworthiness is the aspect of authenticity, which forms the basis for communication with the other entity. For example, an organization's authenticity can be verified through its globally unique ID and the associated digital certificate. [5] describes authentication as a protective mechanism for VANETs in the intelligent transportation systems against malicious entities and is regarded as the primary defense against a multitude of attacks in VANETs such as Sybil Attack, Tunneling Attack, GPS Spoofing or Free-Riding Attack. Moreover, this study highlights the Identity Authentication Certificate (IAC), which is provided by a intimated third party known as Certifying Authorities (CAs). They sign the digital certificates upon validating the authenticity of the certificate's owner, whether it can be an entity, a person, a procedure, or any tangible presence. A further illustration of an identity authentication process is through symmetric and asymmetric password schemes [12].

In the research of intelligent transportation systems, VANETs employ a standardized communication system that integrates features across the entire spectrum, from the physi- cal to the application layer. The principal standards governing VANETs include for instance Dedicated Short Range Communications and Wireless Access in Vehicular Environments. Behind those approaches are different channels presenting various characteristics like embodying control or service functions. Moreover, they can additionally include an established architecture and a complementary suite of standardized protocols, services, and interfaces to facilitate V2V and V2I communications [5].

In the process of establishing secure communication between business partners, the evaluation of the trustworthiness of communicating entities can be facilitated through the utilization of Secure Communication Certificates (SCCs), which are usually employed in printed form, featuring security elements like stamps and holograms to guarantee authenticity. How-

ever, these printed certificates pose challenges for electronic utilization, lacking ease and reliability in electronic applications.

In an alternative perspective, we delve into the crucial significance of incorporating highly dependable resources during the development of new intelligent systems. This integration is pivotal for sustaining both the productivity and accuracy of these systems. Furthermore, researchers conduct an extensive inquiry into reliability, especially its connection with privacy. They explore how these systems must ensure the security and privacy of user data, aiming for a thorough comprehension of the diverse aspects of reliability concerning data privacy in intelligent systems [7]. In addition, reliability is also defined and used in previous research as the measure of the stability and validity of the system and proving that the system will provide consistent accurate results. As an example Lyapunov theory (a mathematical theory which serves as a tool to analyze stability of a dynamic system) was used to prove stability and thus reliability of a trust management mechanism [8, 12].

In trust management, there are a variety of key factors, which are security, privacy, transparency, authenticity, communication, and reliability. These dimensions collectively shape the foundation of trust in smart cities. Security and privacy ensure the protection of sensitive information, while transparency enables openness and clarity. Authenticity is essential for verifying actions affecting individuals and their data, and effective communication is important for trust building. Reliability assures consistent and dependable performance. Although each dimension plays a distinctive role, they intersect in the collective contribution to trust management. Thus, one research challenge is to understand how these dimensions interact within specific domains and adapt trust management frameworks to address evolving technological landscapes.

4 Research Challenges

Based on our reviews, we delineate the obstacles faced while examining the proposed trust management indicators. Individual preference refers to the diverse spectrum of preferences among individuals. Transparent and proper communication highlights how transparency and effective communication may contribute the trust in a specific system. Personalized interpretation signifies the varied understandings regarding mechanisms and trust. The system reliability represents the challenge of managing the system's utility. Lastly, security denotes the protection of the integrity of data used in smart cities.

4.1 Capturing individual preference

Establishing a reliable environment is crucial in maintaining trust among individuals within smart cities. However, this is a challenging scenario due to two essential factors; While intelligent systems inspire trust in certain scenarios, the absence of the human factor remains a concern in real-world applications, where individual preferences seem to favour human interaction over smart systems. Another major challenge is in the potential of the non-adoption of the proposed trust mechanism by some users. This challenge highlights that a one-size-fits-all approach to trust mechanisms might not align with the diverse preferences and perspectives of individuals within smart cities.

4.2 Transparent and proper communication

Addressing another concern involves the lack of proper communication and transparency. This challenge revolves around establishing transparent and user-friendly communication

practices, particularly regarding individuals' privacy. This situation leaves individuals uninformed about their data management, keeping them in the dark. However, the challenge extends beyond this aspect. Smart systems are expected to earn trust through clear communication regarding their reliability and the users' comprehensive understanding of such systems.

4.3 Personalized interpretation

Understanding trust in technological systems requires to consider the dynamics of individual preferences and contexts. The existing trust mechanisms often prioritize indirect trust over personalized, direct trust, overlooking the variability in what individuals seem trustworthy. Achieving a comprehensive trust management mechanism that addresses this challenge involves shedding light on the interaction history between end-users. This mechanism should maintain indirect trust between end-users and edge service providers while acknowledging the diverse needs and perceptions of users. The ultimate goal is to create a system that caters to varied user preferences, ensuring trust is cultivated based on personalized interactions and experiences.

4.4 System reliability

One tendency in research is overlooking the implications of mechanisms on energy consumption is noteworthy. This challenge intends to establish a reliable and valid trust management mechanism within a smart city without compromising its limited resources. This balance is crucial for ensuring the effectiveness and sustainability of the trust system. The reliability and validity of the trust management mechanism can ensure seamless interactions between devices, systems, and inhabitants. Thus, the stakeholders in turn have the confidence of using the systems in smart cities.

4.5 Data protection and privacy

Developing a secure and trustworthy environment for sharing sensitive information poses significant challenges. Enhancing privacy protection, particularly regarding system identity and location, becomes pivotal as increased privacy fosters higher trust levels. Addressing issues like data breaches and unauthorized data usage promptly is imperative. However, the central challenge is to maintain the protection of such data. Creating a secure and trustworthy environment for sharing sensitive information is complex in smart cities. Privacy preserving method not only emerges as a central research topic for sensitive data but also plays a crucial role in building and maintaining trust among stakeholders.

4.6 Research synergy

After presenting five fields of possible challenges, which are considered to be critical obstacles in maintaining trust in smart cities. Meanwhile, those challenges also specify the need for personalized trust mechanisms, transparent communication, and efficient resource usage in smart cities' trust systems. One further derived result can be formulated as "how to leverage the human factor with techniques and machines in smart cities?", answering such a question may reveal the importance of individual preferences in gaining and maintaining trust. However, a more suitable way to tackle the mentioned challenges is to create a better communication systems for individuals.

5 Conclusion

In this paper, we have explored the trust management in the context of smart cities and analyzed a complex landscape for trust management. We have also proposed critical dimensions for trust management, including security and privacy, transparency, authenticity, communication, and reliability. Afterwards we have identified the research challenges in individual preference, transparency, communication, interpretation, system reliability, and data protection in the context of trust dynamics. Addressing these challenges is critical for the effective deployment of intelligent systems in smart cities.

Since trust management is important in the successful integration of AI applications within urban ecosystems, where AI introduces both opportunities and challenges, it is important to propose personalized and context-aware trust management mechanisms in the future, which is to cater the human interaction in certain scenarios, transparent communication, and the impact of varying interpretations.

Given the evolving landscape of smart cities and the integration of intelligent systems in further research is essential to address the challenges associated with trust management. It is also important to further explore novel strategies and solutions to ensure the trustworthy deployment of AI technologies in urban environments. Therefore, those aspects open new directions for future investigation, exploring the feasibility of devising a trust management mechanism in smart cities that ensures reliability and validity with limited resources.

References

1. M. Ge, B. Buhnova, DISDA: Digital Service Design Architecture for Smart City Ecosystems, in *Proceedings of the 12th International Conference on Cloud Computing and Services Science,* CLOSER 2022, April 27-29, 2022, pp. 207–214 (2022)
2. H. Bangui, B. Buhnova, D. Kusniráková, D. Halasz, Trust management in social Internet of Things across domains, Internet Things, **23**, p. 100833 (2023)
3. H. Bangui, M. Ge, B. Buhnova, Deep-Learning based Reputation Model for Indirect Trust Management, in *The 14th International Conference on Ambient Systems, Networks and Technologies, Leuven, Belgium*, **220**, pp. 405–412 (2023)
4. R. Geissbauer, J. Wunderlin, S. Schrauf, J.H. Krause, J.T. Morr, A. Odenkirchen, in *Digital Product Development 2025*, pp. 1–45 (2019)
5. Z. Lu, G. Qu, Z. Liu, A survey on recent advances in vehicular network security, trust, and privacy, IEEE Transactions on Intelligent Transportation Systems, **20**, pp. 760–776 (2018)
6. M. Alazab, G. Manogaran, C.E. Montenegro-Marin, Trust management for internet of things using cloud computing and security in smart cities, Cluster Computing, pp. 1–13
7. T.W. Um, J. Kim, S. Lim, G.M. Lee, Trust management for artificial intelligence: A standardization perspective, Applied Sciences, **12**, p. 6022 (2022)
8. S.A. Alowais, S.S. Alghamdi, N. Alsuhebany, T. Alqahtani, A.I. Alshaya, S.N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H.A. Badreldin et al., Revolutionizing healthcare: the role of artificial intelligence in clinical practice, BMC Medical Education, **23**, p. 689 (2023)
9. Kerasidou, C. Kerasidou, Data-driven research and healthcare: public trust, data governance and the NHS, BMC medical ethics, **24**, p. 51 (2023)
10. K. Haseeb, T. Saba, A. Rehman, Z. Ahmed, H.H. Song, H.H. Wang, Trustmanagement with fault-tolerant supervised routing for smart cities using internet of things, IEEE Internet of Things Journal, **9**, pp. 22608–22617 (2022)

11. N. Jacobs, P. Edwards, M. Markovic, C.D. Cottrill, K. Salt, Who trusts in the smart city? Transparency, governance, and the internet of things, Data & Policy, **2**, p. e11 (2020)
12. B. Wang, M. Li, X. Jin, C. Guo, A reliable IoT edge computing trust management mechanism for smart cities, IEEE Access, **8**, pp. 46373–46399 (2020)