Organizational and Social Impact of AI-Driven Fake Review Detection in e-Commerce

Naren Chandra
Acton, MA, USA

naren1712@gmail.com

Abstract: The rapid growth of e-commerce and mobile transactions has fueled an increase in fraudulent activities with fake online reviews posing a major threat to businesses and consumers. Such reviews, often generated by seller, third parties, or bots, distort reputations, mislead consumer decisions, and erode trust. To address this, organizations are adopting AI-driven detection systems that offer scalability and efficiency while also raising challenges around bias, transparency, and ethical oversight. Drawing on insights from a doctoral dissertation focused on real-time detection of fake reviews using AI, this article examines how these systems shape platform governance, seller behavior, and consumer perceptions of fairness and trust. It highlights implications for risk management, compliance, and regulation, while also assessing social consequences such as false positives, opacity, and perceived bias. The paper recommends strengthening platform governance, ensuring fair treatment of sellers, and enhancing internal risk management at the organizational level, while at the social level, emphasizing consumer trust-building, fairness and legitimacy, and embedding ethical safeguards to protect privacy, accountability, and literacy. Finally, this paper calls for a responsible, human-centric approach to AI deployment that balances automation needs with ethical oversight, enhances consumer confidence, improves platform integrity, and promotes organizational efficiency.

Keywords: Fake online reviews, AI-Driven fake review detection, e-Commerce platform governance, Algorithmic bias and transparency, Consumer trust in digital marketplaces, Ethical AI deployment

1. Introduction

The rapid expansion of e-commerce has transformed the retail landscape by offering consumers convenient access to a vast array of products and services. At the same time, digital platforms are increasingly turning to artificial intelligence (AI) to manage scale, optimize operations, and improve user trust. The integration of artificial intelligence (AI) into digital platforms has profoundly transformed organizational structures, platform governance, and user interaction. Far beyond automating backend operations, AI now influences how content is moderated, how trust is managed, and how decisions are made across industries (Brendel et al., 2021). Organizations are increasingly adopting frameworks like reciprocal symmetry - a principle that calls for balancing technological advancement with human values to guide ethical AI deployment (Carnero, 2017, Srđević, 2025). This shift has led to cultural and structural changes within companies, requiring ethical leadership, transparency in communication, and alignment of AI initiatives with corporate social values (Carnero, 2017, Aakula et al., 2024).

AI's adoption also necessitates strategic investments in infrastructure, talent, and governance. Barriers such as inadequate systems, lack of skilled professionals, and high implementation costs continue to challenge organizations (Brendel et al., 2021). However, studies show that successful integration can foster employee engagement, enhance decision-making processes, and build resilient digital cultures. This necessitates that organizations develop a value system based on established governance guidelines that prioritize ethical AI practices and have oversight mechanisms that prevent biases from being developed and perpetuated. (Aakula et al., 2024).

However, this growth in e-commerce has also given rise to a pressing challenge: the proliferation of fake online reviews. Fraudulent reviews—whether overly positive, maliciously negative, or autogenerated by bots—distort consumer perceptions, mislead purchasing decisions, and erode trust in digital marketplaces. These deceptive practices can result in significant financial losses for consumers, damage to brand reputation, reduced sales performance, and increased operational costs for businesses (Paul & Nikolaev, 2021).

Fake reviews undermine the credibility and integrity of e-commerce platforms, making the detection and removal of such content a strategic necessity. Traditional rule-based or manual moderation systems often fail to distinguish between genuinely critical reviews and those engineered to manipulate, making the task both complex and resource-intensive. As Paul and Nikolaev (2021) note, conventional methods struggle to differentiate malicious content from legitimate user feedback. In this context, artificial intelligence (AI) emerges as a necessary and powerful tool. AI systems can process large volumes of data, detect patterns, and adaptively flag suspicious content in real time. Cao (2021) highlights AI's role in fraud detection through automation, pattern recognition, and continuous learning to address evolving fraudulent behaviors.

Failure to address fake reviews has far-reaching implications. For businesses, this includes loss of consumer trust, reputational harm, reduced marketing effectiveness, regulatory scrutiny, and declining competitive advantage (Mînăstireanu & Meșniță, 2019). For consumers, it leads to misguided purchases, returns, financial loss, and a growing sense of distrust toward online platforms (Jha et al., 2020).

This study seeks to address this problem through two research questions:

RQ1*: How can AI-based methods identify and remove fake online reviews in real time to improve trust, brand reputation, and business outcomes?*

RQ2*: What factors influence organizations to adopt AI for fake review detection, and what challenges and strategic considerations shape their implementation decisions?*

By exploring these questions, the study aims to evaluate the effectiveness and organizational implications of AI in combating fake reviews and sustaining e-commerce integrity.

2. Literature Review

The rise of e-commerce has enhanced the importance of user-generated content, with online reviews serving as a primary source of consumer decision-making (Vidanagama et al., 2019). Yet the proliferation of fake reviews whether overly positive, maliciously negative, or bot-generated poses a significant threat to digital marketplaces. Such deceptive practices distort consumer perceptions, misguide purchasing decisions, and erode trust in platforms (Christopher & Rahulnath, 2016). For organizations, fake reviews can result in reputational damage, reduced sales performance, regulatory scrutiny, and higher operational costs, while for consumers they contribute to financial loss and dissatisfaction (Crawford et al., 2015). Detecting and mitigating fake reviews is therefore not only a technical challenge but also a strategic necessity for sustaining platform integrity and competitiveness.

Another critical strand of literature focuses on trust and reputation systems, especially in e-commerce platforms where user-generated content heavily influences purchase decisions (Pavlou & Dimoka, 2006). Trust in AI systems is critical for acceptance and is multifaceted - spanning accuracy, transparency, ethical handling of data, and user empowerment (Oyekunle et al., 2024). Consumers tend to trust platforms more when AI-generated insights are presented with precision and clarity, enhancing the perceived reliability of information (Kim et al., 2021). Moreover, algorithmic transparency plays a key role in maintaining long-term engagement and minimizing resistance to automated systems (Choung et al., 2023).

In response to concerns about AI opacity and biased decision making, recent advancements in explainable AI (XAI), such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model Agnostic Explanations) aim to make AI decisions more interpretable (Perdigao et al., 2025, Desai et al. 2024). These tools offer visibility into how AI models evaluate data, helping users understand why a piece of content is flagged. This is particularly important for review moderation systems, where false positives can alienate consumers and create reputational risks for platforms. Still, XAI alone may not be sufficient without complementary efforts in digital literacy and consumer education (Swart, 2021).

There is a growing recognition in both academic and industry literature that moderation technologies shape not just platform safety but also user behavior and commercial dynamics. Consumers who perceive AI moderation as arbitrary or opaque may disengage from platforms altogether (Brendel et al., 2021, Marti et al., 2024). Similarly, sellers impacted by inaccurate flagging may lose trust in the system, which can ultimately harm brand equity and platform integrity. These concerns are echoed in the domain of misinformation detection and social media moderation, where scholars stress the need for human-in-the-loop systems and appeal mechanisms to maintain perceived fairness (Perdigao et al., 2025, Srđević, 2025).

Ethical and policy dimensions are central to discussions about AI deployment in digital environments. Literature points to challenges involving data privacy, bias mitigation, and regulatory compliance (Srđević, 2025). Frameworks such as the EU's General Data Protection Regulation (GDPR) and proposed AI Act aim to standardize accountability and promote ethical AI across sectors ((Maddula, 2018). However, these frameworks often lag the pace of AI innovation, making organizational self-regulation and ethical leadership even more important (Srđević, 2025).

This study contributes to this evolving body of knowledge by examining the organizational and social consequences of AI-driven fake review detection in e-commerce. Drawing from a doctoral dissertation that studied how machine learning models could be used to identify fraudulent reviews and understand the factors impacting AI adoption, this research shifts the focus beyond technical accuracy to explore how AI moderation impacts platform governance, seller behavior, user trust, and ethical compliance. It positions fake review detection not just as a fraud prevention mechanism, but as a window into broader questions of responsible AI deployment, algorithmic legitimacy, and stakeholder engagement in digital commerce.

3. Methodology

3.1 Qualitative Research

A qualitative research approach was adopted to gain in-depth insights into the factors influencing the detection of fake online reviews using AI-based Fake Review Detection (FDR) systems. Given the limited prior research in this area, grounded theory was chosen to develop conceptual understanding and generate theory relevant to the research questions (Muzari et al., 2022). The study aims to enhance current understanding of the risks fake reviews pose to consumers and organizations, and the role AI can play in mitigating these risks.

Data was collected through one-on-one semi-structured interviews conducted via Zoom, each lasting up to 60 minutes. A total of 12 participants were interviewed, and this sample size was deemed sufficient upon reaching saturation. Interviews explored participants' experiences, motivations, and perspectives on AI and fake review detection.

Participants represented a range of industries—technology (58%), finance (17%), education, consulting, and healthcare (8% each)—and included professionals from large enterprises and startups. Roles spanned management, director, and C-suite levels, with backgrounds in AI, cybersecurity, and IT. The cohort was well-educated: 25% held doctorates, 50% had master's degrees, and 25% held bachelor's degrees.

The age distribution included individuals from 25 to over 65 years old, offering both early-career and seasoned viewpoints: 25–34 (17%), 35–44 (17%), 45–54 (33%), 55–64 (25%), and 65+ (8%). Gender diversity was limited, with 11 men and one woman.

Participants' AI experience ranged from intermediate (58%) to advanced (33%) and expert (8%), with most using AI tools on a weekly or daily basis. This diversity in demographic, professional, and experiential backgrounds provided rich perspectives on the organizational and technological dimensions of AI-driven fake review detection.

Table 1: Demographic Information and Interview times (in minutes)

Participant	Gender	Age	Education	Industry	AI Experience	AI Use	Interview Time
P1	Male	35-44	Masters	Technology	Intermediate	Weekly	59 mins
P2	Male	55-64	Masters	Technology	Intermediate	Monthly	51 mins
P3	Female	55-64	Doctorate	Education	Intermediate	Monthly	49 mins
P4	Male	35-44	Masters	Finance	Intermediate	Weekly	44 mins
P5	Male	45-54	Bachelors	Consulting	Advanced	Weekly	57 mins
P6	Male	55-64	Masters	Technology	Intermediate	Daily	59 mins
P7	Male	45-54	Bachelors	Finance	Intermediate	Monthly	41 mins
P8	Male	25-34	Masters	Technology	Expert	Daily	57 mins
P9	Male	25-34	Bachelors	Technology	Advanced	Daily	68 mins
P10	Male	45-54	Masters	Technology	Advanced	Weekly	51 mins
P11	Male	45-54	Doctorate	Healthcare	Intermediate	Weekly	36 mins
P12	Male	65+	Doctorate	Technology	Expert	Daily	27 mins

3.2 Data Collection and Alignment with Research Objectives

To explore the broader implications of AI-based fake review detection, this study used a semi-structured interview guide grounded in two core research questions. The guide was designed to elicit insights into how AI alters traditional risk management—such as compliance, accountability, and regulatory alignment—as well as the social consequences of AI enforcement, including trust erosion, bias, and ethical concerns.

For **Research Question 1** (How can AI-based fake review detection methods identify, detect, and remove fake reviews in real time?), the interview focused on technical and operational aspects. Questions were aimed at uncovering real-world applications, system effectiveness, and the trade-offs between automation and accuracy.

For **Research Question 2** (What factors influence organizational adoption of AI for fake review detection?), the interview explored strategic, ethical, and implementation challenges. Topics included algorithmic transparency, user privacy, organizational readiness, and stakeholder trust—key factors in evaluating the governance and legitimacy of AI deployment.

This structured yet flexible format allowed participants to share perspectives across industries and roles, ensuring a nuanced understanding of how AI systems are applied, perceived, and governed in e-commerce contexts.

4. Findings

This study's findings are organized into two key dimensions reflecting the broader implications of AI-based fake review detection: **Organizational Impact** and **Social Impact**.

4.1 Organizational Impact

At the organizational level, participants identified three primary areas affected by AI adoption: **platform governance**, **seller impact**, and **internal risk management**.

4.1.1 Platform-Level governance

The integration of AI systems into existing infrastructures presents significant operational challenges. Participants cited high implementation costs, lack of technical expertise, and limited organizational readiness as major barriers. Successful deployment requires long-term vision, financial resources, and a culture supportive of AI adoption.

AI systems must also evolve continuously through adaptive learning to counter increasingly sophisticated fraudulent tactics. Participants emphasized that failure to keep pace with these developments' risks undermining platform integrity, brand reputation, and consumer loyalty.

A recurring theme was the necessity of **human oversight** to complement algorithmic review moderation. Human reviewers provide contextual understanding such as cultural nuances that AI systems often miss. However, excessive human intervention can slow down real-time operations. Striking the right balance between automation and human judgment was seen as crucial to effective governance.

4.1.2 Impact on sellers

Participants noted that trust in AI systems hinges on their measurable performance. There was strong consensus on the need for standardized accuracy metrics such as false positive/negative rates, sensitivity, and specificity to assess system reliability. Poor accuracy not only harms seller reputation but also erodes consumer trust.

Balanced and representative datasets were highlighted as essential for ensuring fair and accurate detection. In practice, however, achieving dataset diversity remains a challenge, particularly across product categories, languages, and review styles.

Concerns about algorithmic fairness were also raised. Participants expressed apprehension about potential biases that could allow AI systems to be gamed by sellers or inadvertently penalize honest businesses—posing risks to both equity and trust.

4.1.3 Internal risk management

Participants agreed that regulatory frameworks have not kept pace with the adoption of AI. There is a clear need for comprehensive, enforceable guidelines to clarify business responsibility, define consumer protections, and support compliance.

The global nature of fake reviews was another challenge - while businesses may remove reviews from their own sites, they cannot control content on external platforms or social media. Several participants suggested that stricter penalties for hosting or spreading fake reviews could accelerate organizational adoption of AI-based solutions.

4.2 Social Impact

AI-driven fake review detection systems have far reaching social implications, particularly concerning consumer trust, perceptions of fairness, and broader ethical concerns surrounding data use and automated decision-making.

4.2.1 Consumer trust and user experience

Participants emphasized that transparency is essential for building consumer trust in AI systems. While participants acknowledged that fake review detection can enhance user experience by improving content quality, they also noted that misclassifying genuine reviews as fraudulent can have the opposite effect, eroding trust not only in the system but in the platform itself. Instances of false flags were seen to contribute to brand alienation and consumer dissatisfaction.

Participants highlighted the need for clear redressal mechanisms to guide users whose reviews are flagged or removed. Providing users with the ability to appeal or request human review was seen as critical for restoring confidence and preserving platform legitimacy. Consumer education and proactive communication about how AI is used and what safeguards exist were also considered vital for fostering trust.

4.2.2 Perceived fairness and algorithmic legitimacy

Perceptions of fairness were strongly linked to system explainability and transparency. Participants stressed that users must be able to understand why a review was flagged, and how to respond. A lack of visibility into algorithmic processes undermines confidence and fuels skepticism.

Several participants raised concerns about bias in review moderation, particularly in relation to cultural nuances, linguistic variation, and region-specific expressions. These subtleties are often missed by AI models, leading to misinterpretations that challenge the legitimacy of moderation outcomes. Additionally, the potential for algorithmic manipulation by sellers—such as gaming the system to suppress legitimate negative reviews—was seen as a threat to both fairness and platform integrity.

4.2.3 Ethical considerations

Ethical concerns emerged as a central theme in the study. Participants pointed to algorithmic fairness, data privacy, and bias mitigation as key areas requiring attention. A lack of transparency in how user data is collected and used, coupled with the absence of user consent, raised concerns about perceived censorship and consumer rights violations.

Participants also cautioned that AI systems may inadvertently favor larger companies with greater resources to invest in advanced tools, putting smaller businesses at a disadvantage and distorting market fairness. The inclusion of human oversight and the involvement of independent third-party audits were seen as effective strategies to ensure ethical accountability and enhance public confidence.

Rather than deleting reviews outright, some participants advocated for a risk-tiered classification system, allowing potentially fake reviews to be flagged for further validation rather than immediately removed—an approach viewed as more balanced and fairer to all stakeholders.

5. Discussion

This study set out to explore the broader organizational and social implications of using AI-based systems to detect and manage fake reviews in e-commerce. The findings reveal that AI is more than just a fraud detection tool; it is a transformative force shaping platform governance, user behavior, organizational strategy, and digital trust ecosystems.

At the organizational level, the adoption of AI alters traditional structures of risk management and accountability. Participants emphasized the importance of aligning AI systems with internal compliance protocols, technological capabilities, and strategic priorities. High implementation costs, data challenges, and organizational inertia emerged as common barriers. However, businesses that adapt effectively by investing in robust infrastructure, training, and oversight are better positioned to reap the benefits of AI moderation. These findings mirror similar transformations seen in social media platforms, where algorithmic content moderation has become integral to managing scale and user safety, but also raises concerns about bias, censorship, and transparency.

The study also highlights how AI detection tools shape user and seller behavior. From a consumer perspective, the presence of automated moderation may deter fraudulent activity, but it can also create anxiety when errors occur such as the flagging of genuine reviews. These errors may cause disengagement or mistrust unless users are given clear feedback and recourse mechanisms. Censorship concerns also need to be addressed to build trust. For sellers, algorithmic moderation introduces reputational risk and increases the importance of maintaining ethical marketing practices. As with misinformation detection on social platforms, perceived fairness and legitimacy are key to ensuring continued participation and compliance from all actors.

AI moderation thus enhances organizational adaptation. Companies must evolve not only technologically, but also culturally and operationally. The integration of human-in-the-loop systems, standardized accuracy metrics, and transparent communication strategies emerged as critical enablers of effective deployment. Organizations that treat AI as a participatory system—engaging stakeholders and embedding ethical safeguards—are more likely to succeed in building trust and competitive advantage.

At a societal level, the use of AI in fake review detection introduces tensions between efficiency and legitimacy. Participants voiced concerns over data privacy, algorithmic bias, and the lack of redressal options, especially for smaller businesses and linguistic user groups. These concerns echo broader debates in the field of AI governance, where the push for automation often clashes with the need for fairness, explainability, and accountability.

In sum, this study reaffirms that fake review detection is not solely a technical challenge, it is a socio-organizational one. AI has the power to enhance platform integrity and consumer trust, but only when deployed with transparency, inclusivity, and ethical oversight. The lessons drawn here resonate far beyond e-commerce, offering insights relevant to other domains of content moderation, including social media, digital advertising, and misinformation control.

6. Recommendations

Based on the study's findings and analysis, the following recommendations are proposed to guide the responsible and effective implementation of AI-driven fake review detection systems in e-commerce:

Organizational Impact

- **Strengthen platform governance:** Invest in adaptive AI systems complemented by human oversight to balance speed with contextual accuracy. Adopt standardized performance metrics using benchmarks such as false positives/negative rates and specificity to evaluate AI effectiveness and build stakeholder confidence.
- **Support sellers fairly**: Invest in diverse training data that reflect the diversity of languages, cultures, and review contexts to minimize bias and improve fairness. Prioritize transparency and explainability by clearly communicating how AI moderates reviews and why content is flagged or removed so sellers understand moderation outcomes and build trust.
- **Enhance internal risk management**: Develop clear AI governance frameworks, establish regulatory standards for AI transparency, data privacy, and algorithmic accountability in consumer-facing platforms. Collaborate across platforms and social media providers to address fake reviews across digital ecosystems.

Social Impact

- **Build consumer trust:** Educate users on AI moderation by explaining how reviews are evaluated and what steps to take if their feedback is incorrectly flagged. Implement appeal and redressal mechanisms that allow consumers and sellers to contest decisions and request human review. A hybrid model with human-in-the-loop that balances automation with human oversight can address contextual nuance and enhance the system's adaptability over time and improve both accuracy and legitimacy.
- **Ensure fairness and legitimacy:** Develop explainable AI capabilities so users and sellers understand decisions. Require platforms to notify users when content moderation is AI-driven and what recourse is available. Guard against cultural or linguistic bias through continuous dataset refinement and third-party audits to validate fairness, accuracy, and compliance with ethical standards.
- **Embed ethical safeguards:** Protect data privacy and obtain explicit user consent while upholding fairness and transparency in AI-driven moderation. Commit to transparent marketing strategies and avoid exploiting algorithmic loopholes. Hold organizations accountable through independent oversight and third-party audits. Promote digital literacy initiatives to empower consumers to

recognize fake reviews and understand the role of AI in content moderation and shaping their online experiences.

These recommendations aim to foster a responsible AI ecosystem that balances innovation with ethical integrity -enhancing consumer trust, protecting platform reputation, and supporting sustainable digital commerce.

7. Conclusion

The proliferation of fake online reviews presents a serious threat to the credibility and integrity of e-commerce platforms. These deceptive practices distort consumer perceptions, mislead decision-making, and erode trust in digital marketplaces. This study set out to explore the organizational and social implications of using AI-based systems to detect and manage fake online reviews in e-commerce.

Through qualitative analysis this study has shown that artificial intelligence offers powerful, scalable tools to detect and mitigate such fraudulent content. However, its deployment is not without complexity. AI systems influence not only the technical landscape of fraud prevention but also fundamentally alters traditional risk management practices, compels new forms of platform governance, and significantly influences seller behavior and user experience. While AI increases operational efficiency and enables real-time content moderation, it also introduces risks around fairness, transparency, explainability, and trust.

These findings underscore the double-edged nature of AI in e-commerce. On one hand, AI provides scalable, efficient tools for moderating content and protecting consumers. On the other, it can alienate users, unintentionally marginalize smaller businesses, and create new forms of bias and opacity if not carefully governed. These tensions mirror similar challenges in adjacent domains such as misinformation or fake news detection and social media moderation, where the speed of automation often clashes with the demand for procedural legitimacy and ethical oversight. The tension between efficiency and legitimacy is especially acute in environments where trust is both a consumer expectation and a competitive advantage.

Considering these dynamics, the study calls for a responsible, human-centered approach to AI deployment. At the organizational level, platforms must embed ethical safeguards, integrate automation with contextual human judgment, involve diverse stakeholders, and maintain transparency in how moderation systems are designed, implemented, and evaluated. At the social level, consumer trust must be built by providing clear explanations, offering appeal and redress mechanisms, and hybrid oversight, while ensuring fairness through explainable AI, third-party audits and continuous bias mitigation. Embedding ethical safeguards – protecting privacy, securing consent, and promoting digital literacy will be essential to align automation with fairness and accountability.

As fake reviews continue to evolve in sophistication, so too must the tools used to detect them. Yet technical advancement alone is insufficient. The future of AI in e-commerce will depend on how well it is integrated into the social, organizational, and ethical frameworks that sustain trust in digital ecosystems. By addressing both technological capability and social responsibility, AI can fulfill its promise—not just as a tool for detection, but as a foundation for more trustworthy and equitable online commerce.

Ethics Declaration: Ethical clearance was not required for this study.

AI Declaration: During the preparation of this article, the author used OpenAI to improve the paper's readability. After using this tool, the author reviewed and edited the content as needed.

References

Aakula, A., Saini, V., & Ahmad, T. (2024). The impact of AI on organizational change in digital transformation. *Internet of Things and Edge Computing Journal, 4*(1).

Brendel, A. B., Mirbabaie, M., Lembcke, T.-B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability*, *13*(4), 1974

Cao, L. (2021). Artificial intelligence in retail: applications and value creation logics. *International Journal of Retail & Distribution Management*, *49*(7), 958–976. https://doi.org/10.1108/ijrdm-09-2020-0350

Carnero, M. C. (2017). Asymmetries in the Maintenance Performance of Spanish Industries Before and After the Recession. *Symmetry*, *9*(8), 166. https://doi.org/10.3390/sym9080166

Chandra, N. (2025). *Using the Power of Artificial Intelligence (AI) for Fraud Detection and Prevention in E-Commerce/Online Retail* (Order No. 32043842). Available from Proquest Dissertations & Theses Global. (3222709425). https://www.proquest.com/dissertations-theses/using-power-artificial-intelligence-ai-fraud/docview/3222709425/se-2

Christopher, S. L., & Rahulnath, H. A. (2016). Review authenticity verification using supervised learning and reviewer personality traits. *2016 International Conference on Emerging Technological Trends [ICEET], 9*, 1-7. https://doi.org/10.1109/icett.2016.7873647

Choung, H., David, P., & Ross, A. (2023). Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction*, *39*(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Najada, H. A. (2015). Survey of review spam detection using machine learning techniques. *Journal of Big Data, 2*(1). https://doi.org/10.1186/s40537-015-0029-9

B. Desai, K. Patil, I. Mehta and A. Patil, "Explainable AI in Cybersecurity: A Comprehensive Framework for enhancing transparency, trust, and Human-AI Collaboration," 2024 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, pp. 135-150. 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10762690

Jha, B. K., Sivasankari, G. G., & Venugopal, K. R. (2020). Fraud detection and prevention by using big data analytics. *Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020)*. https://doi.org/10.1109/iccmc48092.2020.iccmc-00050

Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. *Psychology and Marketing, 38*(7), 1140-1155. https://doi.org/10.1002/mar.21498

Maddula, S. S. (2018). The impact of AI and reciprocal symmetry on organizational culture and leadership in the digital economy. *Engineering International, 6*(2), 201-210. https://doi.org/10.18034/ei.v6i2.703

Marti, C. L., Liu, H., Kour, G., Bilgihan, A., & Xu, Y. (2024). Leveraging artificial intelligence in firm-generated online customer communities: a framework and future research agenda. *Journal of Service Management, 35*(3), 438-458. https://doi.org/10.1108/josm-10-2023-0443

Mînăstireanu, E., & Meșniță, G. (2019). An analysis of the most used machine learning algorithms for online fraud detection. *Informatică Economică*, *23*(1/2019), 5–16. https://doi.org/10.12948/issn14531305/23.1.2019.01

Muzari, T., Shava, G. N., & Shonhiwa, S. (2022). Qualitative research paradigm, A key research design for educational researchers, processes and procedures: A theoretical overview. *Indiana Journal of Humanities and Social Sciences*, *3*(1). https://indianapublications.com/Journals/IJHSS

Oyekunle, D., Matthew, U. O., 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

Paul, H. L., & Nikolaev, A. G. (2021). Fake review detection on online E-commerce platforms: a systematic literature review. *Data Mining and Knowledge Discovery*, *35*(5), 1830–1881. https://doi.org/10.1007/s10618-021-00772-6

Pavlou, P. A., & Dimoka, A. (2006). Institutional Feedback Technologies in Online Marketplaces: An investigation of feedback text comments, trust, and price premiums. *SSRN Electronic Journal*. https://doi.org/10.21.39/ssrn.2380738

Perdigão, P. A., Coelho, N. M., & Brás, J. C. (2025). AI-Driven Threats in Social Learning Environments - A multivocal literature review. *ARIS2 – Advanced Research on Information Systems Security, 5*(1), 4-37. https://doi.org/10.56394/aris2.v5i1.60

Srđević, B. (2025). Evaluating the Societal impact of AI: A comparative analysis of human and AI platforms using the Analytic Hierarchy Process. *AI, 6*(4), 86. https://doi.org/10.3390/ai6040086

Swart, J. (2021). Experiencing Algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society, 7*(2). https://doi.org/10.1177/20563051211008828

Vidanagama, D.U., Silva, T.P. & Karunananda, A.S. Deceptive consumer review detection: a survey. *Artif Intell Rev* 53, 1323–1352 (2020). https://doi.org/10.1007/s10462-019-09697-5