JURNAL ILMIAH DINAMIKA SOSIAL

VOLUME 9 NUMBER 1, FEB 2025 | pp.103-118 | E-ISSN : 2581-2424 | P-ISSN : 2597-3657 |
Website: https://journal.undiknas.ac.id/index.php/fisip/index
DOI: https://doi.org/10.38043/jids.v9i1.6278

AI IN CYBER DEFENSE: PRIVACY RISKS, PUBLIC TRUST, AND POLICY CHALLENGES

Abdullah Azizi[1*], Mohammad Qias Mohammadi[2], Abdul Wahid Samadzai[3]

[1*] Computer Science Department, Baghlan University, Baghlan, Afghanistan
[2] Information Systems Department, Badakhshan University, Badakhshan, Afghanistan
[3]Software Engineering Department, Kaul University, Kabul, Afghanistan

E-mail: [1]Abdullah.azizi1368@gmail.com, [2]m.qias@badakhshan.edu.af, [3]samadzai@gmail.com

ABSTRACT

The rapid integration of Artificial Intelligence (AI) into cybersecurity systems, particularly AI-based cyber defense systems, is reshaping the landscape of digital security. This study explores the social impacts of these systems, focusing on privacy, security, and public trust. The purpose of this research is to examine the effects of AI-driven cybersecurity on individuals and society, addressing concerns such as privacy risks, security breaches, and trust in digital platforms. A systematic literature review (SLR) methodology was employed, synthesizing relevant academic studies, conference proceedings, and reports from credible databases, including IEEE Xplore, ACM Digital Library, and ScienceDirect. The results reveal that while AI-based systems improve threat detection and response times, they also raise significant concerns about data privacy, surveillance, and the potential for algorithmic bias. Additionally, the integration of AI in cyber defense has prompted debates on the ethical implications of automated decision-making and the transparency of these systems. In conclusion, while AI offers transformative benefits in cybersecurity, careful attention is required to balance its advantages with ethical and privacy considerations. This study emphasizes the need for ethical frameworks and public awareness to ensure that AI-based systems are deployed in a manner that fosters trust and protects citizens' rights.

Keywords: *AI-Based Systems; Cybersecurity; Privacy; Public Trust; Ethical Implications*

INTRODUCION

Artificial Intelligence (AI) has significantly transformed the cybersecurity landscape, bringing unprecedented capabilities to defend against evolving digital threats. As cyberattacks become increasingly sophisticated, AI-based cyber defense systems have emerged as a critical component in securing industries such as banking, healthcare, and smart technologies (AL-Dosari, Fetais, & Kucukvar, 2024; Nasnodkar, Cinar, & Ness, 2023). These systems leverage machine learning algorithms, data analytics, and automation to detect anomalies, predict potential threats, and respond more rapidly than traditional methods (Capuano, Fenza, Loia, & Stanzione, 2022). However, while AI offers promising opportunities for enhancing cyber

Diserahkan : 2024-10-11,
Direvisi : 2025-01-11,
Diterima : 2025-02-19,

defense, it also introduces new social challenges, including privacy concerns, algorithmic bias, and ethical dilemmas (Kaloudi & Li, 2020; Malatji & Tolah, 2024).

One of the major societal opportunities presented by AI-based cyber defense systems is the ability to protect sensitive personal and organizational data more efficiently. In sectors like banking, AI-driven security protocols have improved threat detection, reducing incidents of financial fraud and data breaches (AL-Dosari et al., 2024). Additionally, AI technologies provide small businesses and vulnerable sectors with affordable and scalable security solutions, helping democratize cybersecurity access (Bécue, Praça, & Gama, 2021). Moreover, AI enhances the resilience of critical infrastructures against cyberattacks, safeguarding public trust and national security (Khan, Arif, & Khan, 2024).

Nevertheless, the deployment of AI in cyber defense raises critical social challenges. Privacy remains a top concern as AI systems often require vast amounts of personal data to function effectively, potentially infringing on individuals' rights (Edu, Such, & Suarez-Tangil, 2020). The risk of biased AI models can also lead to unequal protection or false accusations, exacerbating existing societal inequalities (Gupta, Akiri, Aryal, Parker, & Praharaj, 2023). Furthermore, adversarial attacks that exploit AI vulnerabilities present a growing threat, highlighting the need for continual advancements in AI robustness (Hussain, Neekhara, Jere, Koushanfar, & McAuley, 2021).

Recent studies have addressed the technical capabilities of AI in cybersecurity and discussed ethical challenges separately. However, there remains a notable research gap: limited attention has been paid to evaluating the combined social impact of AI-based cyber defense systems—particularly how these systems simultaneously shape public trust, data privacy, and societal equity in digital spaces. Few works have systematically connected technical, ethical, and social dimensions in a comprehensive way. This study aims to bridge this gap by offering a holistic exploration of the societal consequences of AI-driven cybersecurity, integrating technical advancements with broader social implications.

The objective of this study is to explore the social impact of AI-based cyber defense systems, focusing on both their benefits and challenges. It examines how AI enhances cybersecurity effectiveness while raising ethical and privacy concerns. Additionally, the study highlights emerging opportunities for strengthening societal resilience through advanced AI solutions and proposes recommendations for the responsible and equitable deployment of AI in cybersecurity.

Problem Statement

Although AI-based cyber defense systems offer transformative potential for improving the speed, accuracy, and scalability of cybersecurity operations, their rapid integration into critical sectors has introduced complex social and ethical challenges that are insufficiently addressed in existing research. While AI enhances threat detection and response capabilities, concerns about data privacy violations, algorithmic bias, lack of accountability, and the misuse of autonomous systems are increasingly prominent. Many AI-driven cybersecurity solutions operate as "black boxes," with opaque decision-making processes that undermine transparency, public trust, and regulatory compliance.

Moreover, as adversaries also adopt AI to conduct more sophisticated, adaptive cyberattacks, traditional defense strategies become less effective, necessitating new approaches that account for evolving threats. Despite growing scholarly attention to the technical capabilities of AI in cybersecurity, limited work has systematically evaluated its broader societal impacts, particularly issues related to equity, fairness, and the protection of vulnerable populations.

This growing dependence on AI, without robust ethical and governance frameworks, risks exacerbating social inequalities, marginalizing already at-risk communities, and escalating cybersecurity vulnerabilities. Therefore, there is a critical need to fill this gap by examining the social implications of AI in cyber defense holistically and proposing strategies for its ethical, transparent, and socially responsible deployment.

Research Question

The following research questions guide the exploration of the social impacts, challenges, and opportunities associated with AI-based cyber defense systems. These questions aim to address key issues surrounding privacy, ethics, and citizen engagement in the context of emerging cybersecurity technologies.

RQ1: What are the key social impacts of AI-based cyber defense systems on privacy, security, and public trust?

RQ2: What ethical challenges arise from deploying AI in cybersecurity, and how can they be addressed?

RQ3: How can AI-based cyber defense systems enhance citizen engagement and promote inclusion in digital spaces?

RQ4: What policy recommendations and regulatory frameworks are necessary to ensure the ethical and transparent deployment of AI in cybersecurity?

LITERATURE REVIEW

Artificial Intelligence (AI) has increasingly become integral to modern cyber defense systems, offering novel capabilities for detecting, preventing, and responding to cyber threats. However, its deployment raises diverse technical, ethical, and social considerations, as highlighted in the literature.

Privacy and Data Security

Privacy remains a central concern in AI-based cyber defense. AL-Dosari, Fetais, and Kucukvar (2024) illustrate how AI improves threat detection in the banking sector but simultaneously introduces complex issues around data management and system transparency. Similarly, Edu, Such, and Suarez-Tangil (2020) discuss how smart home assistants, though enhancing convenience, expose users to significant privacy risks. Bécue, Praça, and Gama (2021) further emphasize that while AI fortifies Industry 4.0 security, it also broadens the attack surface for cybercriminals. There is general consensus that AI enhances security, but debate persists around the adequacy of current privacy safeguards against large-scale data collection.

Bias, Explainability, and Trust

The opaque nature of AI decision-making, often referred to as the "black box" problem, undermines user trust. Capuano, Fenza, Loia, and Stanzione (2022) advocate for Explainable AI (XAI) to increase transparency and public confidence in AI decisions. Kaloudi and Li (2020) also highlight the rapid evolution of AI threats, stressing the importance of explainability to adapt defenses. While most studies agree on the necessity of transparency, there is contention regarding how explainability can be operationalized without compromising system performance or security.

Emerging Threats and Adversarial Risks

Adversarial attacks exploiting AI vulnerabilities are a growing threat. Hussain et al. (2021) demonstrate that even AI-based deepfake detectors can be manipulated, revealing the fragility of current defenses. Gupta et al. (2023) warn that generative AI models like ChatGPT can be weaponized, creating new vectors for cyberattacks. This strand of literature agrees that AI's dual-use nature (defensive and offensive) necessitates continual adaptation and innovation in cybersecurity.

Social Engineering and Human-Centric Threats

AI's role in combating social engineering attacks is increasingly recognized. Fakhouri et al. (2024) propose AI-driven detection frameworks, while Khan, Arif, and Khan (2024) highlight AI's transformative potential in enhancing social engineering defenses. However, Dash, Ansari, Sharma, and Ali (2022) caution against over-reliance on AI, noting that human judgment remains essential in complex attack scenarios. Here, a balanced approach combining AI efficiency with human oversight is widely recommended.

System Robustness and Future Directions

Finally, the literature identifies the need for more resilient AI systems. Sarker (2023) emphasizes adversarial learning as a key defense strategy, while Malatji and Tolah (2024) stress the importance of aligning AI innovations with ethical frameworks. Hakimi et al. (2024) also call for future research on systemic robustness and ethical deployment.

Overall, while there is strong consensus on AI's potential to transform cyber defense, debates continue regarding data privacy protections, bias mitigation, system transparency, and ethical safeguards. This growing body of research highlights the need for an integrated approach that balances technological innovation with social responsibility.

Table 1. Research Trends in AI-Based Cyber Defense Systems

Theme	Research Focus	Key Findings	Key References
Privacy and Data Security	Impact of AI on privacy in smart systems and banking sectors	AI strengthens security but raises concerns over data transparency and user privacy.	AL-Dosari et al. (2024); Bécue et al. (2021); Edu et al. (2020)
Bias, Explainability, and Trust	Explainable AI (XAI) to enhance transparency and trust	Strong advocacy for XAI to increase interpretability; debates on operationalizing it effectively.	Capuano et al. (2022); Kaloudi & Li (2020)
Emerging Threats and Adversarial Risks	AI vulnerabilities and weaponization of generative AI models	Growing consensus on the dual-use risk of AI; urgent need for adaptive defenses.	Hussain et al. (2021); Gupta et al. (2023)
Social Engineering and Human-Centric Threats	AI applications in detecting and mitigating social engineering attacks	AI-driven tools are promising, but human oversight remains crucial.	Fakhouri et al. (2024); Khan et al. (2024); Dash et al. (2022)
System Robustness and Ethical Deployment	Building resilient and ethically aligned AI-based cyber defense systems	Importance of adversarial learning and ethical frameworks to ensure sustainable cyber resilience.	Sarker (2023); Malatji & Tolah (2024); Hakimi et al. (2024)

METHODS

This study adopts a systematic literature review (SLR) methodology to investigate the social impacts, challenges, and opportunities of AI-based cyber defense systems. The SLR approach was chosen to ensure a structured and comprehensive analysis of the most relevant academic studies, conference proceedings, and high-impact reports on AI and cybersecurity. Following established review protocols, the study involves identifying, selecting, and synthesizing available research to derive critical insights and detect patterns across different sectors and geographical contexts (Sarker, 2023; Capuano et al., 2022).

Research Design and Search Strategy

The research process began with the formulation of specific research objectives and keywords related to AI in cyber defense, societal impacts, ethical considerations, and technological challenges. Databases such as IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, and Taylor & Francis Online were systematically searched using keywords like "AI in cybersecurity," "social impact of AI defense systems," and "ethical challenges in AI security." Clear inclusion and exclusion criteria were developed: only peer-reviewed English-language articles published between 20190 and 2025 were included.

Selection and Evaluation of Sources

The selected papers were then evaluated based on their relevance, methodological rigor, and contribution to the research questions. After filtering duplicates and low-quality studies, a final dataset of 20 highly relevant papers was synthesized. Data extraction focused on major themes such as opportunities created by AI-based defense, social risks, privacy issues, bias in AI systems, and frameworks for ethical AI deployment.

Analysis of Methodical Framework (Figure 1)

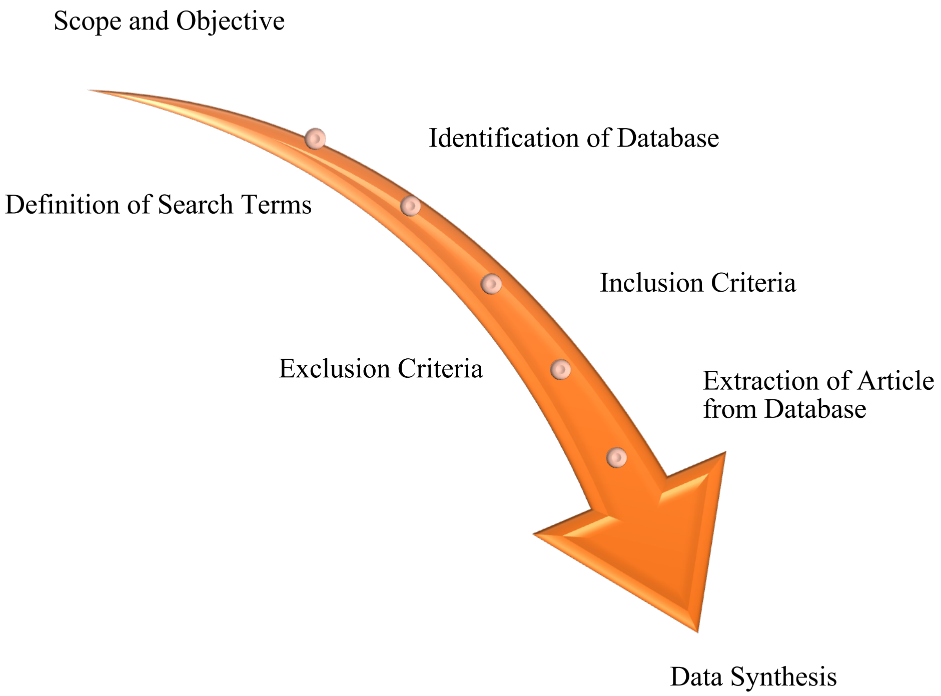
Scope and Objective
Identification of Database
Definition of Search Terms
Inclusion Criteria
Exclusion Criteria
Extraction of Article from Database
Data Synthesis

Figure 1. Methodological Framework for Literature Review

The methodical framework depicted in Figure 1 represents a systematic approach to conducting the literature review for the study of AI-based cyber defense systems. The figure outlines a structured flow, beginning with identification of databases and followed by the definition of search terms. The process emphasizes transparency and thoroughness, starting with clear inclusion and exclusion criteria, which filter out irrelevant or low-quality articles to ensure that only high-impact and peer-reviewed studies are considered.

Once the articles are selected, they are extracted from databases such as IEEE Xplore and SpringerLink, based on their relevance to the research questions. The framework then highlights the data synthesis process, where the articles are categorized into key themes such as ethical challenges, privacy concerns, and technological opportunities. This framework ensures that the review is both comprehensive and methodologically rigorous, facilitating the identification of emerging trends and gaps in the existing literature.

Definition of Search Terms

Search terms were defined to capture the broad spectrum of issues related to AI-based cyber defense systems. Keywords such as "AI in cybersecurity," "social impact of AI defense systems," "ethical challenges in AI security," and "AI for cyber threat detection" were used. These terms were refined to ensure that all relevant research articles were retrieved, focusing on both the technological aspects and societal consequences of AI in cyber defense.

Inclusion Criteria and Exclusion Criteria

The following table outlines the Inclusion and Exclusion Criteria applied to select relevant studies for this systematic literature review.

Table 2. Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Time Frame	Articles published between 2019 and 2025	Articles published before 2019
Language	Peer-reviewed studies published in English	Non-English articles
Content Relevance	Studies focusing on AI in cybersecurity and its social impacts	Studies not directly related to AI in cybersecurity
Methodology	Empirical research, case studies, and theoretical discussions	Opinion pieces, non-peer-reviewed sources
Article Type	Peer-reviewed journal articles, conference papers	Non-peer-reviewed materials, blog posts

The Inclusion Criteria ensure that only high-quality, relevant, and recent research is considered for this study. Articles published within the last five years (2019–2025) provide an up-to-date perspective on AI in cyber defense, reflecting current technological advancements and social impacts. Peer-reviewed English-language articles were prioritized for their credibility and academic rigor. The content was selected based on its direct relevance to AI applications in cybersecurity, focusing on both the technological and societal aspects.

Conversely, the Exclusion Criteria filter out studies that do not meet these standards. Articles published before 2019 were excluded to avoid outdated perspectives, and non-peer-reviewed materials were omitted to maintain academic integrity. Studies that lacked a clear focus on AI or cybersecurity were also removed, ensuring that the literature review remains targeted and coherent.

Data Selection

The PRISMA flow diagram in Figure 1 outlines the systematic process used to identify, screen, and select studies for this review on the social impact of AI-based cyber defense systems. Initially, 300 records were retrieved from academic databases. After removing 150 duplicates, 150 records remained for screening. Of these, 100 records were excluded for irrelevance, leaving 50 reports sought for retrieval. However, 10 reports could not be retrieved due to accessibility issues. Out of the 40 reports assessed for eligibility, 19 were excluded based on predefined criteria: 10 were non-English studies, 5 were unrelated to AI-based cybersecurity, and 4 were outdated (published before 2015). Finally, 21 high-quality, relevant studies were included in the final synthesis. This rigorous screening process ensured that only recent, peer-reviewed, and thematically aligned literature contributed to the study, enhancing its reliability, focus, and overall academic integrity.

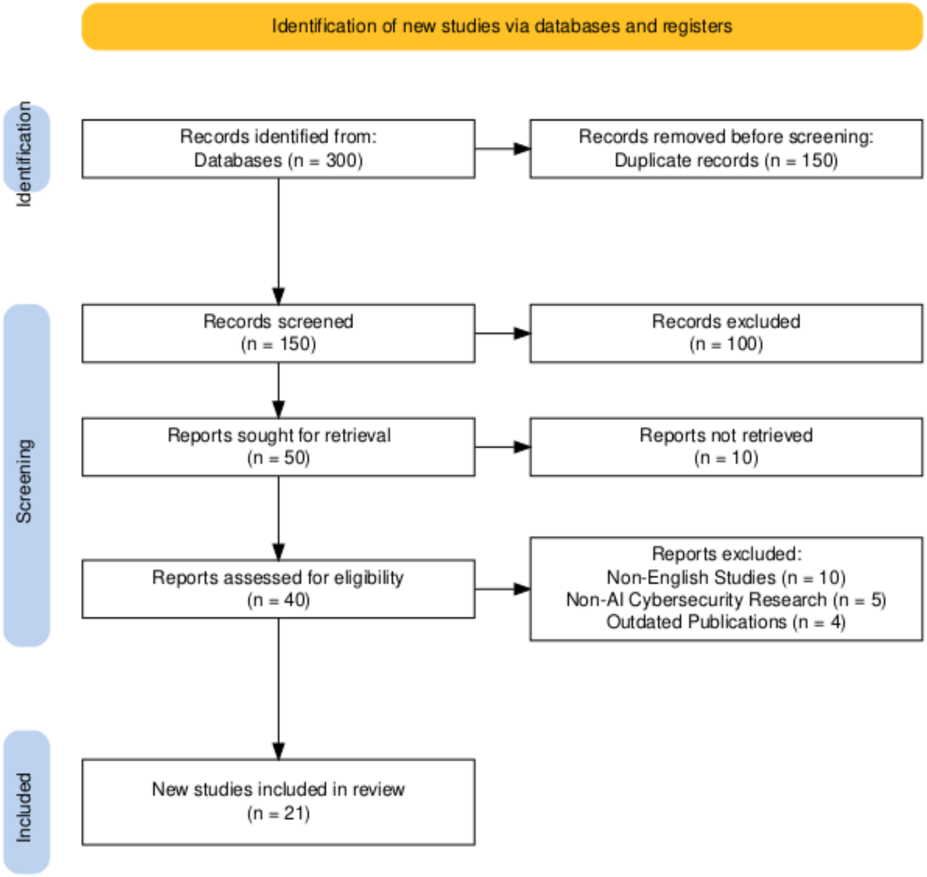
Identification of new studies via databases and registers
Identification
Records identified from:
Databases (n = 300)
Records removed before screening:
Duplicate records (n = 150)
Screening
Records screened
(n = 150)
Records excluded
(n = 100)
Reports sought for retrieval
(n = 50)
Reports not retrieved
(n = 10)
Reports assessed for eligibility
(n = 40)
Reports excluded:
Non-English Studies (n = 10)
Non-AI Cybersecurity Research (n = 5)
Outdated Publications (n = 4)
Included
New studies included in review
(n = 21)

Figure 2. PRISMA Flow Diagram Illustrating the Selection Process for AI-Based Cyber Defense Studies

Data Synthesis

The data synthesis process involved a thorough analysis and categorization of the 30 selected studies. The main themes extracted from the literature focused on the opportunities, challenges, ethical implications, and social impacts of AI-based cyber defense systems. A critical synthesis of the literature revealed several key areas of interest, including the effectiveness of AI in enhancing cyber defense capabilities, the risks of privacy violations, and the ethical concerns surrounding AI's decision-making processes in cybersecurity (Capuano et al., 2022; Gupta et al., 2023; Sarker, 2023).

Studies indicated that AI can significantly improve the accuracy and speed of threat detection and response in cybersecurity, allowing for real-time analysis and mitigating potential risks before they escalate (Gupta et al., 2023). However, ethical issues such as algorithmic bias, transparency, and accountability in AI decision-making processes emerged as major concerns. The literature highlighted the need for robust frameworks that ensure the fair and ethical deployment of AI in cybersecurity, emphasizing that these systems must be designed to prioritize privacy and human oversight (Capuano et al., 2022; Sarker, 2023).

The synthesis also identified a lack of consensus on the regulatory and governance frameworks for AI in cybersecurity, with some studies suggesting that the implementation of AI in defense systems requires rigorous oversight to avoid the risks of misuse or unintended consequences (Sarker, 2023). Finally, opportunities for increasing public trust and citizen engagement through transparent AI systems were discussed, with a focus on developing inclusive, accessible AI-driven cybersecurity systems that promote fairness and equity in digital spaces.

RESULTS AND DISCUSSION

The results section of this study addresses the key social impacts, ethical challenges, and opportunities associated with AI-based cyber defense systems. Through the analysis of various research questions (RQs), the findings highlight the complex interplay between AI's capabilities and its implications for privacy, security, public trust, and inclusion in digital spaces. The study examines how AI can enhance cybersecurity while also raising concerns regarding privacy violations, algorithmic bias, and the lack of transparency. Additionally, the results reveal how AI-driven cybersecurity solutions can promote citizen engagement, inclusivity, and equitable access to digital platforms. The following sections present a detailed analysis of these findings.

RQ1: What are the key social impacts of AI-based cyber defense systems on privacy, security, and public trust?

The integration of AI in cyber defense systems has a significant impact on privacy, security, and public trust. AI enhances cybersecurity by enabling faster threat detection and response, improving overall security. However, concerns about data privacy arise as AI systems may require access to sensitive information to function effectively. Studies have shown that AI systems can sometimes violate privacy rights if not properly governed (Gupta et al., 2023). Furthermore, the transparency of AI decision-making remains an ongoing concern, as the "black-box" nature of some AI algorithms can undermine public trust (Capuano et al., 2022).

Additionally, while AI improves security capabilities, it also introduces new risks, including algorithmic bias and misuse of AI technology for malicious purposes, which can erode trust in digital systems. Public perception of AI in cybersecurity is shaped by the understanding of these benefits and risks. Building trust through transparency, accountability, and privacy protection is crucial for the successful deployment of AI in cyber defense (Sarker, 2023).

Table 3. Social Impacts of AI-Based Cyber Defense Systems

Social Impact	Description	References
Privacy Concerns	AI systems require access to sensitive data, risking privacy violations if not properly governed.	Gupta et al., 2023
Security Enhancement	AI enables faster threat detection, improving security defenses and response times.	Capuano et al., 2022
Public Trust	AI's decision-making transparency and accountability are critical for maintaining public trust.	Sarker, 2023
Algorithmic Bias	Bias in AI algorithms can negatively impact the effectiveness of cybersecurity systems, leading to mistrust.	Gupta et al., 2023

RQ2: What ethical challenges arise from the deployment of AI in cybersecurity, and how can they be addressed?

The deployment of artificial intelligence (AI) in cybersecurity presents various ethical challenges that need to be carefully considered to ensure fair and responsible technology use. A significant concern is algorithmic bias, where AI systems, depending on the data they are trained on, may reflect biases that lead to discriminatory outcomes in areas like threat detection and identity verification (Capuano et al., 2022). This bias may unfairly target certain groups, jeopardizing equity and fairness in the cybersecurity landscape.

Another ethical challenge is the issue of accountability. With AI systems being highly autonomous, it becomes challenging to attribute responsibility when AI-based cybersecurity tools make incorrect decisions, leading to breaches or security lapses. This raises questions of liability and legal responsibility, especially when harm results from faulty AI interventions (Basit et al., 2021). Clear guidelines on human oversight and responsibility in AI decision-making are crucial in addressing this issue.

Lastly, transparency is a key concern. AI's "black box" nature often makes it difficult to understand how decisions are made, which can result in a lack of trust from users and stakeholders. This opacity, particularly in the deployment of AI systems for detecting and mitigating cyber threats, undermines public confidence (Gupta et al., 2023). To mitigate this, the development of explainable AI (XAI) is essential. XAI models, which allow users to understand and interpret how AI systems make decisions, are increasingly seen as crucial for ensuring ethical compliance and public trust in AI-driven cybersecurity solutions (Sarker, 2023).

Table 4. Ethical Challenges of AI in Cybersecurity

Ethical Challenge	Description	References
Algorithmic Bias	AI systems may make biased decisions based on skewed or incomplete data, leading to discrimination.	Capuano et al., 2022
Accountability	Determining responsibility when AI systems make incorrect decisions, especially in the event of breaches.	Basit et al., 2021
Transparency	Lack of clarity in how AI systems make decisions can undermine trust and ethical compliance.	Gupta et al., 2023
Privacy Violations	AI-driven cybersecurity systems may infringe on personal privacy by surveilling users without consent.	Fakhouri et al., 2024

RQ3: How can AI-based cyber defense systems enhance citizen engagement and promote inclusion in digital spaces?

AI-based cyber defense systems have the potential to foster greater citizen engagement and inclusivity by offering advanced solutions that protect privacy, ensure security, and create equitable digital environments. By proactively addressing cyber threats such as data breaches, identity theft, and phishing attacks, these systems contribute to a safer online experience for citizens. AI-powered systems can also improve the accessibility of digital services, ensuring that underrepresented groups have secure access to online platforms. For example, AI technologies can be used to enhance the security of e-governance systems, making them more transparent and trustworthy, which encourages broader participation in digital democracy (Gupta et al., 2023; Fakhouri et al., 2024).

Furthermore, AI tools can facilitate the identification of barriers faced by marginalized communities in the digital space. By utilizing AI-driven analysis, organizations can tailor security measures that accommodate various socio-economic conditions and enhance inclusivity (Hakimi et al., 2024). However, challenges such as algorithmic bias and the risk of discrimination must be carefully managed to ensure fairness in the implementation of AI technologies (Sarker, 2023).

Table 5. AI-based Cyber Defense Systems and Citizen Engagement

Study	Key Findings
Gupta et al. (2023)	AI systems in cybersecurity can enhance public trust by ensuring data privacy and protection from threats.
Fakhouri et al. (2024)	AI-driven cybersecurity solutions promote secure and equitable access to e-government services.
Hakimi et al. (2024)	AI tools help identify barriers for marginalized communities, enhancing their digital inclusion.

RQ4: What policy recommendations and regulatory frameworks are necessary to ensure the ethical and transparent deployment of AI in cybersecurity?

The rapid adoption of AI in cybersecurity necessitates the development of robust policy recommendations and regulatory frameworks to address ethical, privacy, and transparency concerns. Effective regulation must emphasize transparency by requiring AI developers to implement explainable AI (XAI) systems, enabling stakeholders to understand and trust automated decision-making processes (Capuano, Fenza, Loia, & Stanzione, 2022). Privacy protection should be a cornerstone, mandating strict data governance policies that limit unnecessary data collection and ensure compliance with global standards like GDPR. Additionally, frameworks should enforce regular audits and bias testing of AI models to prevent discrimination and ensure fairness (Gupta, Akiri, Aryal, Parker, & Praharaj, 2023). Governments must also establish guidelines for the responsible use of AI, particularly restricting autonomous systems that could impact individual rights without human oversight. Cross-sector collaboration between policymakers, cybersecurity experts, and ethicists is essential to create adaptive regulations that evolve with technological advancements (Kaloudi & Li, 2020). In developing nations like Afghanistan, capacity building and international support are critical to ensuring AI-driven cybersecurity measures do not widen the digital divide. Overall, proactive policy-making and standardized ethical frameworks are crucial to harness AI's benefits in cybersecurity while safeguarding public trust and democratic values.

Discussion

AI-based cyber defense systems have proven to be transformative tools in enhancing citizen engagement and fostering inclusivity within digital spaces. These systems, leveraging advanced machine learning algorithms and real-time data analysis, can significantly enhance the security and privacy of online interactions, thereby increasing public trust in digital platforms (Gupta et al., 2023). By ensuring that personal data and interactions are protected, AI-driven defense systems provide citizens with a sense of security, encouraging them to engage more actively in online spaces, such as e-governance systems, digital democracy platforms, and social media (Capuano et al., 2022).

Furthermore, AI-based defense systems have the potential to address significant accessibility challenges faced by marginalized communities in the digital sphere. By using AI algorithms to detect and mitigate cyber threats that disproportionately affect vulnerable populations, these systems can level the playing field and ensure secure access to digital services. AI systems can also adapt to different languages, cultures, and socio-economic conditions, ensuring that marginalized groups are not excluded from critical digital services (Hakimi et al., 2024). For instance, AI-driven security systems in e-government services can help citizens from diverse backgrounds securely access public services, making these systems more inclusive and equitable.

The role of AI in enhancing digital inclusion extends beyond security. AI technologies can support the development of accessible user interfaces and adaptive systems that cater to the needs of individuals with disabilities or those in underserved regions. Such inclusivity can help bridge the digital divide by enabling all citizens to engage with online services without fear of discrimination or exclusion due to lack of technical literacy or accessibility barriers (Dash et al., 2022). Moreover, AI can be instrumental in identifying patterns of exclusion and offering personalized recommendations to make digital platforms more user-friendly for diverse populations.

Despite the potential benefits, the implementation of AI-based cyber defense systems must be carefully managed to mitigate the risk of reinforcing existing inequalities. One critical challenge is algorithmic bias, where AI systems may unintentionally perpetuate societal inequalities due to biased data sets (Sarker, 2023). For example, facial recognition technologies used in cybersecurity systems have been criticized for exhibiting racial and gender biases, leading to unequal access for marginalized communities. Therefore, it is essential to incorporate fairness, transparency, and accountability into the design and deployment of AI systems (Bécue et al., 2021).

These findings have critical implications for policymakers, who must establish clear regulatory frameworks to guide ethical AI deployment in cybersecurity. AI developers are encouraged to prioritize transparency, fairness, and inclusivity in system design. For the public, increasing digital literacy and awareness around AI-driven security measures is essential. Together, these efforts can foster a safer and more equitable digital environment.

CONCLUSION

AI-based cyber defense systems have the potential to revolutionize how digital spaces are secured and managed, particularly by enhancing citizen engagement and promoting inclusivity. By utilizing advanced algorithms to detect and prevent cyber threats, these systems play a crucial role in fostering trust and confidence among users. As digital platforms continue to expand, ensuring the security and privacy of users becomes paramount. AI technologies are essential for mitigating growing cybersecurity risks and protecting sensitive data, all while enhancing the overall user experience.

Moreover, AI systems can help break down barriers to accessibility, ensuring that marginalized and underserved communities are not excluded from critical digital services. Adaptive, user-friendly platforms and personalized security measures enable equitable digital participation for individuals with diverse needs and capabilities, helping to bridge the digital divide.

However, the implementation of AI in cyber defense must be approached cautiously. Risks such as algorithmic bias, unintended exclusions, and ethical challenges must be proactively addressed. Fairness, transparency, and accountability must remain central principles in AI system design and deployment.

Collaboration between governments, the private sector, and civil society is urgently needed to develop robust AI ethics standards and regulatory frameworks. By balancing technological innovation with strong ethical governance, AI-based cyber defense can contribute to creating secure, inclusive, and participatory digital environments for all.

Recommendations

To maximize the potential of AI-based cyber defense systems, it is recommended that developers prioritize transparency and fairness in algorithm design to prevent biases. Additionally, continuous monitoring and refinement of AI models should be conducted to ensure they remain effective in addressing emerging threats. Collaboration between stakeholders, including governments, businesses, and cybersecurity experts, is essential to create inclusive policies that ensure equitable access to digital security for all citizens.

Future Research

Future research should focus on developing AI models that are more adaptable to evolving cyber threats while ensuring inclusivity. Investigating ethical frameworks and the social impact of AI in cybersecurity will also be critical to address potential biases and promote fairness.

REFERENCES

AL-Dosari, K., Fetais, N., & Kucukvar, M. (2024). Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges. *Cybernetics and systems*, *55*(2), 302-330. https://doi.org/10.1080/01969722.2022.2112539

Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2021). A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommunication Systems, 76*, 139–154. https://doi.org/10.1007/s11235-020-00733-2

Bécue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. *Artificial Intelligence Review*, *54*(5), 3849-3886. https://doi.org/10.1080/01969722.2022.2112539

Budzinski, O., Noskova, V., & Zhang, X. (2019). The brave new world of digital personal assistants: Benefits and challenges from an economic perspective. *NETNOMICS: Economic Research and Electronic Networking, 20*, 177–194. https://doi.org/10.1007/s11066-019-09133-4

Capuano, N., Fenza, G., Loia, V., & Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access, 10*, 93575–93600. https://doi.org/10.1109/ACCESS.2022.3204171

Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). Threats and opportunities with AI-based cyber security intrusion detection: a review. *International Journal of Software Engineering & Applications (IJSEA)*, *13*(5). https://ssrn.com/abstract=4323258

Edu, J. S., Such, J. M., & Suarez-Tangil, G. (2020). Smart home personal assistants: A security and privacy review. *ACM Computing Surveys, 53*(6), Article 116. https://doi.org/10.1145/3412383

Fakhouri, H. N., Alhadidi, B., Omar, K., Makhadmeh, S. N., Hamad, F., & Halalsheh, N. Z. (2024, February). Ai-driven solutions for social engineering attacks: Detection, prevention, and response. In *2024 2nd International Conference on Cyber Resilience (ICCR)* (pp. 1-8). IEEE. https://doi.org/10.1109/ICCR61006.2024.10533010

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access, 11*, 80218–80245. https://doi.org/10.1109/ACCESS.2023.3300381

Hakimi, M., Ezam, Z., Totakhail, A., & Ghafory, H. (2024). Transformative Impact of Artificial Intelligence on IoT Applications: A Systematic Reviewof Advancements, Challenges, and Future Trends. *International Journal of Academic and Practical Research*, *3*(1), 1-1. https://www.ejournals.ph/article.php?id=24184

Hakimi, M., Suranata, I. W. A., Ezam, Z., Samadzai, A. W., Enayat, W., Quraishi, T., & Fazil, A. W. (2025). Generative AI in Enhancing Hydroponic Nutrient Solution Monitoring. *Jurnal Ilmiah Telsinas Elektro, Sipil dan Teknik Informasi*, *8*(1), 94-103. https://doi.org/10.38043/telsinas.v8i1.6242

Hussain, S., Neekhara, P., Jere, M., Koushanfar, F., & McAuley, J. (2021). Adversarial deepfake: Evaluating vulnerability of deepfake detectors to adversarial examples. In *Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 3347–3356). IEEE. https://doi.org/10.1109/WACV48630.2021.00339

Kaloudi, N., & Li, J. (2020). The ai-based cyber threat landscape: A survey. *ACM Computing Surveys (CSUR)*, *53*(1), 1-34. https://doi.org/10.1145/3372823

Khan, M. I., Arif, A., & Khan, A. R. A. (2024). AI's Revolutionary Role in Cyber Defense and Social Engineering. *International Journal of Multidisciplinary Sciences and Arts*, *3*(4), 57-66. https://www.neliti.com/publications/591913/ais-revolutionary-role-in-cyber-defense-and-social-engineering

Khinvasara, T., Ness, S., & Tzenios, N. (2023). Risk management in medical device industry. *Journal of Engineering Research and Reports, 25*(8), 130–140. https://doi.org/10.9734/jerr/2023/v25i896547

Kreinbrink, J. L. (2019). *Analysis of artificial intelligence (AI) enhanced technologies in support of cyber defense: Advantages, challenges, and considerations for future deployment* (Master's thesis, Utica College). https://www.proquest.com/openview/2ca10115b5be484fc619b2534e01ace0/1

Malatji, M., & Tolah, A. (2024). Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI. *AI and Ethics*, 1-28. https://doi.org/10.1007/s10462-020-09942-2

Nasnodkar, S., Cinar, B., & Ness, S. (2023). Artificial intelligence in toxicology and pharmacology. *Journal of Engineering Research and Reports, 25*(7), 192–206. https://doi.org/10.9734/jerr/2023/v25i795249

Sarker, I. H. (2023). Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview. *Security and Privacy*, *6*(5), e295. https://doi.org/10.1007/s43681-024-00427-4

Xuan, T., & Ness, S. (2023). Integration of blockchain and AI: Exploring application in the digital business. *Journal of Engineering Research and Reports, 25*(8), 20–39.
https://doi.org/10.9734/jerr/2023/v25i895548