Journal of Advanced Robotics, Autonomous Systems and Human-Machine Interaction

PRIME OPEN ACCESS

Volume 2, Issue 1

Research Article

Date of Submission: 12 Dec, 2025
Date of Acceptance: 02 Jan, 2026
Date of Publication: 12 Jan, 2026

Artificial Intelligence and Electoral Cybersecurity: A Technical Case Study of South Africa’s 2024 National and Provincial Elections

John Maphephe[1]* and Surendra Thakur[2]

[1]Independent Electoral Governance & Technology Advisor, South Africa

[2]College of Science, Engineering and Technology, University of South Africa, South Africa

***Corresponding Author:** John Maphephe, Independent Electoral Governance & Technology Advisor, South Africa.

Citation: Maphephe, J., Thakur, S. (2025). Artificial Intelligence and Electoral Cybersecurity: A Technical Case Study of South Africa’s 2024 National and Provincial Elections. *J Adv Robot Auton Syst Hum Mach Interact, 2*(1), 01-07.

Abstract

Artificial Intelligence (AI) is becoming increasingly central to electoral cybersecurity, offering advanced capabilities for real-time anomaly detection, incident response, and disinformation monitoring. This paper presents a technical case study of the Electoral Commission of South Africa’s deployment of AI during the 2024 National and Provincial Elections (NPE). The Commission implemented AI-enhanced firewalls, intrusion prevention systems, endpoint detection and response (EDR), web application firewalls (WAF), and a parallel Security Operations Centre (SOC) to detect, analyze, and mitigate cyber threats. Advanced tools—including User and Entity Behavior Analytics (UEBA), phishing detection, dark-web monitoring, and disinformation tracking—were operationalized at scale using AI and machine learning (ML). Natural language processing (NLP) further augmented threat anticipation and intelligence capabilities. The analysis demonstrates that AI served as a critical force multiplier in safeguarding electoral infrastructure, while highlighting ongoing challenges related to transparency, dependence on external providers, and the explainability of AI-generated outputs. This study contributes to the technical literature on AI-driven cybersecurity by detailing architectural integration, evaluating operational effectiveness, and providing evidence-based recommendations for enhancing resilience in future elections.

Keywords: Artificial Intelligence (AI), Electoral Cybersecurity, Security Operations Center (SOC), User and Entity Behavior Analytics (UEBA), Disinformation Detection, Hybrid Human–AI Security, Election Integrity, South Africa

Highlights

- This paper presents a technical case study of the Electoral Commission of South Africa’s use of AI in the 2024 National and Provincial Elections to enhance cybersecurity and electoral integrity.
- AI-driven tools, including UEBA, EDR, dark-web monitoring, phishing detection, and NLP-based threat intelligence, were deployed at scale to detect anomalies, anticipate threats, and mitigate disinformation.
- The study demonstrates the effectiveness of hybrid human–AI architectures, particularly parallel SOC integration, in strengthening operational resilience.
- Challenges such as algorithmic explainability, dependency on external providers, and ethical considerations are identified, with policy recommendations offered to guide future AI deployment in electoral contexts.

Introduction

Electoral systems worldwide are increasingly exposed to cybersecurity risks due to rapid digitization, the adoption of cloud services, and the growing sophistication of cyber adversaries [1]. Traditional security measures, such as signature-based antivirus solutions and conventional firewalls, are proving insufficient against advanced threats, including novel malware, distributed denial-of-service (DDoS) attacks, and AI-generated disinformation campaigns [2]. In response to these challenges, the Electoral Commission of South Africa implemented AI-enabled cybersecurity measures for the 2024 National and Provincial Elections (NPE). This paper provides a technical case study examining the architectural integration of AI tools within the Commission’s digital ecosystem. It evaluates the effectiveness of these tools in safeguarding electoral infrastructure, explores the interplay between automated systems and human oversight, and situates the findings within the broader scholarly discourse on AI-driven election security. By detailing both operational successes and challenges, this study contributes to the understanding of how advanced AI technologies can enhance resilience, integrity, and trust in contemporary electoral processes.

Literature Review

Global Trends in Electoral Cybersecurity

Electoral systems are increasingly reliant on digital infrastructure, including electronic voter registration, online result transmission, and digital communication with citizens. While digitization offers efficiency and transparency, it also introduces complex cybersecurity risks. Globally, electoral commissions face threats from malware, ransomware, distributed denial-of-service (DDoS) attacks, and AI-generated disinformation campaigns aimed at influencing voter behavior or undermining public trust [1,2]. Traditional cybersecurity measures, such as signature-based antivirus software and static firewalls, are often inadequate to detect novel, sophisticated attacks, highlighting the need for more advanced, adaptive security solutions.

AI in Cybersecurity Architectures

AI-driven cybersecurity tools increasingly combine anomaly detection, predictive analytics, and automated remediation to address the complexity of modern cyber threats [3]. Traditional signature-based detection systems, while effective against known malware, are increasingly inadequate in environments characterised by zero-day exploits and polymorphic attacks. AI, through supervised and unsupervised learning, enables systems to adapt dynamically to evolving threat landscapes. User and Entity Behaviour Analytics (UEBA) exemplifies this shift. By establishing baselines of "normal" activity, UEBA systems can identify anomalies such as login attempts from unusual geographies, irregular access patterns, or sudden surges of machine-to-machine traffic that may indicate Distributed Denial of Service (DDoS) attacks [4]. These systems not only detect threats but also contextualise behaviours across different layers of the network, thereby reducing false positives and increasing the efficiency of human analysts. Endpoint Detection and Response (EDR) solutions are another critical innovation in AI-based cybersecurity. Unlike traditional antivirus software, which relies heavily on predefined signatures, EDR platforms harness machine learning to analyse behavioural telemetry at scale, detecting novel malware and advanced persistent threats (APT) in near real time [5]. Integration with AI allows these systems to provide predictive insights, anticipating potential compromises before they manifest fully. Beyond anomaly detection and EDR, AI has also been applied in intrusion prevention systems (IPS), web application firewalls (WAF), and network traffic analysis. These systems increasingly employ deep learning algorithms to identify complex attack vectors such as SQL injection, cross-site scripting, and encrypted malware payloads, which may bypass traditional controls [6]. Moreover, AI-driven orchestration platforms support automated remediation, triggering predefined responses—such as isolating infected endpoints or blocking malicious IP ranges—without requiring manual intervention [7]. However, while AI enhances detection and response, it also introduces challenges. Issues of explainability and transparency remain central: many AI-driven systems operate as "black boxes," producing outputs that are difficult for analysts to interpret or audit [8]. This raises concerns about accountability in high-stakes environments such as elections. Furthermore, adversarial AI—where attackers manipulate algorithms through poisoning or evasion techniques—represents an emerging threat that complicates the security benefits of AI adoption [9]. Thus, while AI architectures hold promise for strengthening resilience against sophisticated cyber threats, their deployment must be balanced with considerations of governance, human oversight, and adversarial risk mitigation.

AI and Electoral Integrity

AI plays a dual role in electoral contexts, contributing both to emerging threats and to enhanced defenses. On the one hand, deepfakes and generative AI enable scalable disinformation campaigns, undermining public trust and influencing voter behaviou [10]. On the other hand, AI-driven monitoring tools facilitate the detection of fake accounts, botnets, and coordinated inauthentic behaviour, supporting the identification and mitigation of malicious activity [11]. Electoral commissions worldwide have increasingly integrated AI into Security Operations Centers (SOCs) to automate alert triage, enrich incident response, and improve situational awareness [12]. Recent scholarship highlights AI's dual function in electoral systems: it enhances operational efficiency and resilience while introducing novel risks and ethical challenges. AI can act as a force multiplier, enabling the detection of large-scale threats, real-time anomaly monitoring, and rapid threat prioritization [13,14]. Nonetheless, concerns persist regarding algorithmic explainability, reliance on external providers, potential biases in automated systems, and privacy implications associated with monitoring voter-related data [7]. These dynamics underscore the need for hybrid human–AI security architectures and robust governance frameworks to balance effectiveness with accountability and public trust.

Hybrid Human–Machine Security Architectures

Hybrid cybersecurity architectures, combining AI automation with human oversight, are increasingly recommended in critical infrastructure and electoral systems. AI handles large-scale data analysis and real-time anomaly detection, while human analysts provide contextual interpretation, validate alerts, and make operational decisions [13,15]. This approach addresses limitations of fully automated systems, reduces false positives, and enhances decision-making while maintaining ethical oversight.

South Africa in Global Context

South Africa's adoption of AI in NPE 2024 reflects broader global concerns about electoral integrity in the age of AI, where both established and emerging democracies face common challenges in securing digital electoral infrastructure. While countries such as Estonia and the United States have long integrated advanced AI-enabled systems into their election management processes, South Africa's case demonstrates how a developing democracy can strategically leverage AI for resilience without the same scale of resources. Early analysis suggests limited impact of AI-generated disinformation during these elections, partly due to institutional preparedness, proactive monitoring of the digital environment, and coordinated responses between the Electoral Commission and external cybersecurity providers [16]. The South African experience further illustrates how AI adoption in electoral contexts is shaped by regional realities, including high levels of mobile connectivity, growing social media penetration, and vulnerabilities linked to uneven digital literacy. In this regard, South Africa provides an instructive technical example of AI implementation that balances defensive cybersecurity mechanisms with capacity-building imperatives. This case also contributes to the global discourse by showing how democratic institutions in the Global South can both innovate and adapt global best practices in election technology, while simultaneously highlighting issues of dependency on external vendors, data sovereignty, and the importance of contextualised governance frameworks.

Research Gap

While there is extensive literature on AI in general cybersecurity applications, studies focusing specifically on electoral systems remain limited. Few works provide in-depth technical analysis of AI architectures, operational integration, and human–machine interaction within Election Management Bodies (EMBs). This paper addresses this gap by providing a detailed case study of the Electoral Commission of South Africa's AI deployment during the 2024 National and Provincial Elections, contributing both empirical insights and policy-relevant recommendations.

Conceptual Framework

The study is guided by a hybrid intelligence framework, which situates AI as a complementary tool to human expertise in critical cybersecurity operations. This framework emphasizes proactive threat anticipation, continuous learning, and the integration of ethical governance alongside technological deployment, providing a lens for analyzing both operational effectiveness and systemic challenges in electoral AI systems [17].

Methodology

This study applies a qualitative case study research design, which is particularly suited to exploring complex sociotechnical systems in context [18]. The case study method enables in-depth analysis of how Artificial Intelligence (AI) tools were deployed by the Electoral Commission of South Africa during the 2024 National and Provincial Elections (NPE), and how these tools intersect with broader questions of electoral integrity, resilience, and governance. This methodology enables the investigation of multiple dimensions of AI implementation, including operational functionality, cybersecurity measures, stakeholder interactions, and the broader socio-political implications of automated systems in electoral processes. Through in-depth exploration, the study captures not only the technical aspects of AI deployment but also its effects on institutional resilience, public trust, and regulatory compliance [19]. The approach provides a nuanced understanding of how AI technologies shape, and are shaped by, the organizational, political, and social context of South Africa's electoral environment.

Case Selection

The Electoral Commission of South Africa was selected due to the unprecedented scale of AI integration during the 2024 NPE. Its adoption of multiple AI-driven cybersecurity measures—including AI-enhanced firewalls, intrusion prevention systems, EDR, UEBA, phishing detection, dark-web monitoring, and disinformation tracking—provides a comprehensive context for examining operational effectiveness, human–machine collaboration, and ethical considerations.

DataSources

Three principal data sources underpin this study

- **Primary data:** Official feedback from the Electoral Commission detailing the technical integration of AI-enabled cybersecurity tools, including firewalls, intrusion prevention systems, endpoint detection and response (EDR), web application firewalls (WAF), and chatbot functionality. This data provides insight into the Commission's rationale, operational procedures, and outcomes of AI use [20].
- **Secondary literature:** Peer-reviewed academic studies, policy reports, and technical documents from think tanks, international organisations, and industry specialists were analysed to situate the South African case within broader scholarly and policy debates on AI and election security. These sources enrich the empirical evidence with theoretical and comparative perspectives.
- **Comparative benchmarks:** Practices from other democracies—including Estonia, the United States, and member states of the European Union—were examined as comparative points of reference. This benchmarking allowed identification of global best practices and contextual adaptations relevant to South Africa [21].

AI Tool sand Architectural Overview

Key deployed systems included:

- **Parallel SOC:** Managed by an external provider, augmented with AI for log correlation, anomaly detection, and threat prioritization.
- **UEBA:** Monitored user and network behaviour for deviations indicative of threats.

- **EDR:** Provided granular endpoint telemetry and detected previously unknown malware.
- **Phishing and Email Security:** AI gateways detected linguistic anomalies and social engineering patterns.
- **Attack Surface & Disinformation Monitoring:** AI crawlers scanned open and dark web sources for threats and misinformation.
- **ML/NLP Threat Intelligence:** Aggregated and enriched threat feeds for proactive risk anticipation

Analytical Approach

The study employed a thematic analysis framework (Braun & Clarke, 2006) to categorise and interpret the data. Codes and themes were developed iteratively, with particular attention to the functional domains of AI deployment. Data was mapped against three technical domains:

- **Anomaly Detection** – including User and Entity Behaviour Analytics (UEBA) and EDR tools that identify deviations from normal network or endpoint behaviour.
- **Threat Anticipation and Intelligence** – encompassing AI applications in predictive analytics, machine learning (ML), and natural language processing (NLP) for proactive cybersecurity defence.
- **Human–Machine SOC Integration** – focusing on the hybrid model of combining AI-enabled monitoring with human analyst oversight in the Security Operations Centre (SOC).

This three-domain framework ensured analytical coherence while allowing cross-comparison with global practices. Triangulation of primary, secondary, and comparative data enhanced the study's validity and reliability.

Ethical Considerations

Given the sensitive nature of electoral data and cybersecurity operations, ethical considerations were central to the design and conduct of this study. The research maintained strict confidentiality and anonymized all sensitive information, ensuring compliance with data protection regulations and safeguarding both institutional and individual privacy. The deployment of AI in electoral cybersecurity raises specific ethical concerns, including transparency, explainability, and accountability. AI systems, particularly those integrated into Security Operations Centers (SOCs) and threat intelligence platforms, can produce outputs that are difficult for human operators to interpret. To address this, the study evaluated how AI-supported decisions were validated by human analysts, ensuring that automated outputs were contextualized and subject to oversight. This hybrid human–AI governance model aligns with recommended best practices for ethical AI deployment in critical infrastructure (IEC, 2018). Additionally, the study considered potential biases in AI-driven monitoring, including differential treatment of data inputs, false positives, and algorithmic assumptions that could inadvertently affect operational decisions. Ethical evaluation also included attention to transparency and documentation of AI workflows, enabling reproducibility, auditability, and accountability in cybersecurity operations. Finally, the research adhered to broader ethical principles of integrity and responsibility, ensuring that findings were reported accurately, that risks to participants or institutions were minimized, and that recommendations for AI deployment were grounded in both technical evidence and ethical governance frameworks. By embedding these ethical safeguards, the study not only evaluated technical effectiveness but also promoted responsible and accountable AI adoption in electoral cybersecurity contexts.

Limitations

While the case study approach provides rich contextual detail, it also has inherent limitations [19]. Access to certain operational details of the Electoral Commission's systems was restricted for security reasons, limiting the granularity of technical verification. Furthermore, reliance on official feedback risks potential bias, which was mitigated through triangulation with independent secondary literature and comparative international examples. Despite these constraints, the methodology provides a rigorous and transparent account of AI's role in electoral cybersecurity during South Africa's NPE 2024.

Technical Architecture of AI Deployment

Threat Detection and Anomaly Analysis (UEBA & EDR)

The Commission implemented User and Entity Behavior Analytics (UEBA) algorithms to continuously monitor network and user activity. Unusual deviations—such as logins from unexpected geographic locations or machine-to-machine traffic indicative of potential DDoS attacks—automatically triggered alerts for further investigation. Complementing this, AI-powered Endpoint Detection and Response (EDR) agents were deployed across endpoints, providing detailed telemetry and deep system visibility. These agents leveraged advanced code analysis to identify previously unknown malware, moving beyond reliance on traditional signature-based detection and enhancing the overall cybersecurity posture of the electoral infrastructure [15].

Phishing and Email Security

The Commission employed an AI-enabled email gateway to filter inbound messages and proactively detect phishing attempts. Leveraging natural language processing and machine learning, the system identified subtle linguistic anomalies, suspicious social engineering cues, and irregularities in sender metadata. This AI-driven approach enhanced the organization's ability to recognize and block targeted spear-phishing campaigns, reducing the risk of credential compromise and strengthening overall email security. By continuously learning from emerging threats, the system also adapts over time, improving resilience against increasingly sophisticated cyberattacks [22].

Attack Surface and Disinformation Monitoring

To safeguard the electoral process, the Commission collaborated with external cybersecurity service providers to implement AI-driven monitoring across its digital ecosystem [23]. Advanced web crawlers mapped the Commission's attack surface, encompassing both the open internet and the dark web, to identify potential vulnerabilities and threats. These systems continuously scanned for malicious websites impersonating official electoral platforms, leaked or compromised credentials, and emerging disinformation narratives circulating on social media [23]. By correlating data from multiple sources, AI enabled real-time threat detection and prioritized high-risk incidents for rapid human-led investigation and remediation [7]. This proactive approach not only strengthened the resilience of the Commission's online infrastructure but also served a critical role in countering misinformation, protecting voter trust, and maintaining the integrity of the electoral process. The integration of automated monitoring with expert human oversight exemplifies a hybrid cybersecurity strategy, combining the speed and scale of AI with contextual judgment and analytical expertise.

Threat Anticipation Through Machine Learning and Natural Language Processing

The Commission leveraged threat intelligence platforms incorporating machine learning (ML) and natural language processing (NLP) to aggregate and analyze diverse cybersecurity data feeds [15,4]. These platforms processed structured and unstructured information, including logs, vulnerability reports, social media activity, and dark web intelligence, applying AI algorithms to identify patterns, trends, and emerging threats [24]. By enriching raw data with contextual interpretation, AI enabled analysts to anticipate potential risks proactively, prioritize high-risk incidents, and reduce false positives [23]. This approach enhanced situational awareness and improved the decision-making capacity of cybersecurity teams, allowing for more efficient allocation of resources and pre-emptive responses to threats that could impact electoral operations. The integration of ML/NLP-driven analysis with human oversight exemplifies a hybrid intelligence model, in which automated insights are validated and contextualized by experienced analysts to optimize threat anticipation and response.

Parallel Security Operations Center (SOC) Integration

To enhance the cybersecurity resilience of the electoral infrastructure, the Commission deployed a parallel Security Operations Center (SOC) operated by an external provider, complementing its internal cybersecurity team [18]. Advanced AI tools augmented the SOC's capabilities by analyzing and correlating millions of system events in real time, detecting anomalous patterns, and recommending appropriate remediation strategies [14,13]. AI-generated insights were continuously reviewed and validated by human analysts, exemplifying a hybrid human–machine security architecture that combines the computational speed and scale of automation with the contextual judgment of cybersecurity experts [15,7]. This integrated approach improved threat detection, reduced response times, and enhanced the overall resilience of the electoral system against cyberattacks while aligning with international cybersecurity standards.

Voter-Facing AI Chatbot

The Electoral Commission's website chatbot represents an early yet significant step toward integrating AI into public electoral engagement. Although its current functionality is limited, the chatbot serves as a digital interface for voters, providing timely information on voter registration, polling locations, election dates, and procedural guidelines [25]. By automating responses to frequently asked questions, the system reduces the informational burden on human staff while increasing accessibility for citizens seeking electoral guidance. Beyond basic informational functions, the chatbot has the potential to collect data on voter queries, which can inform improvements in public outreach and highlight areas of common confusion or misinformation. Its deployment reflects a broader strategic intent to explore AI-mediated communication as a tool for enhancing transparency, responsiveness, and citizen trust in the electoral process [26]. However, there are notable limitations and challenges. The chatbot's current capabilities are restricted by its pre-programmed responses and inability to handle complex or ambiguous queries, which may lead to incomplete or inaccurate guidance. Ethical considerations, such as data privacy, bias in automated responses, and digital accessibility for users with limited literacy or connectivity, remain critical concerns. Despite these challenges, lessons learned from this pilot initiative can inform the development of more sophisticated AI systems capable of multilingual support, adaptive learning, and real-time issue resolution, ultimately strengthening the democratic process [21].

Discussion and Policy Recommendations

Strengths of AI Deployment

The deployment of AI within the Commission's electoral cybersecurity framework demonstrated several significant strengths. Scalability was a key advantage: AI algorithms processed and analyzed data volumes far beyond the capacity of human analysts, including system logs, threat intelligence feeds, and social media activity. This enabled timely insights and real-time detection of potential threats, such as anomalies in network traffic or emerging disinformation campaigns, which would be impractical with manual monitoring alone [14]. Proactive defense was another notable benefit. The integration of User and Entity Behavior Analytics (UEBA) and machine learning (ML)-based threat intelligence allowed for early identification of suspicious user behavior, endpoint anomalies, and emerging phishing campaigns. By anticipating threats before they escalated, the Commission enhanced the resilience of its electoral infrastructure and strengthened its capacity to maintain election integrity [15,13]. Hybrid human–machine architecture further reinforced the effectiveness of AI deployment. The parallel Security Operations Center (SOC), augmented with AI tools, enabled rapid detection, correlation, and prioritization of alerts, while human analysts provided contextual interpretation and validation. This combination minimized false positives, improved decision-making, and ensured that AI outputs adhered to oFperational and ethical standards [7]. Beyond technical performance, AI deployment improved situational awareness, accelerated response times, and allowed cybersecurity personnel to focus on complex investigations and strategic planning, enhancing operational efficiency and the robustness of electoral processes.

Limitations and Challenges of AI Deployment

Despite these strengths, AI deployment also presented notable limitations and challenges. Technical limitations included dependency on the quality and completeness of input data. Incomplete logs, gaps in threat intelligence feeds, or inconsistent social media data could reduce detection accuracy, while novel attack vectors or zero-day malware required ongoing model updates and expert intervention [14]. Human oversight and expertise remained critical. Effective interpretation of AI-generated alerts required cybersecurity knowledge and contextual understanding of the electoral environment. Limited analyst availability or training gaps could reduce the effectiveness of AI-human collaboration and delay threat response [13]. Ethical and privacy considerations were also significant. The collection and analysis of voter-related data, including online behavior and social media activity, necessitated compliance with data protection regulations and ethical guidelines. Mismanagement in this area could undermine public trust or provoke legal challenges. Resource and infrastructure constraints posed further challenges. Advanced AI systems, such as ML-driven threat intelligence, UEBA, and AI-enhanced SOCs, required substantial computational resources, secure infrastructure, and continuous maintenance. Less resourced electoral bodies might face difficulties replicating this model without external support [4]. Finally, adversarial threats continued to evolve. Sophisticated actors could deploy generative AI, swarm-based attacks, or disinformation campaigns designed to bypass detection, creating an ongoing arms race between AI defenses and emerging threats (Shu, Sliva, Wang, Tang, & Liu, 2017). Operational dependence on external providers introduced additional risks, including potential misalignment in priorities or delayed communication, underscoring the need for robust governance and integration protocols.

Broader Implications

This case study highlights AI's dual role in electoral cybersecurity: it enhances resilience and operational efficiency while introducing dependency, ethical considerations, and governance challenges. AI-enabled monitoring, threat anticipation, and hybrid human–machine architectures demonstrate the potential of advanced technologies to detect and mitigate cyber threats, disinformation, and operational anomalies. However, reliance on AI also requires careful oversight to prevent algorithmic bias, overdependence on automation, and privacy violations [17]. As adversaries increasingly leverage generative AI and swarm-based attacks electoral commissions must expand their AI capabilities while embedding robust oversight, human-in-the-loop processes, and ethical governance frameworks. This necessitates investment in AI literacy, algorithmic transparency, and cross-institutional collaboration, ensuring that AI supports democratic processes rather than introducing new risks. The South African case demonstrates that successful AI deployment depends on balancing technological innovation with human judgment, ethical oversight, and adaptive governance—a lesson broadly applicable to electoral bodies worldwide.

Policy Recommendations

Building on the discussion of strengths, limitations, and broader implications, several policy recommendations are proposed to guide the strategic and responsible deployment of AI in electoral systems.

- **Develop In-House AI Capacity:** Electoral commissions should invest in building internal AI expertise to reduce reliance on external SOCs and third-party providers. Local expertise enables agile threat response, contextual understanding, and long-term institutional resilience [13].
- **Mandate Explainability and Transparency:** AI systems must adhere to frameworks ensuring interpretability and transparency of algorithmic decisions. Explainable AI fosters accountability, builds public trust, and supports ethical governance by allowing stakeholders to understand how critical security or engagement decisions are made [15].
- **Expand AI for Voter Engagement:** Deploy multilingual, NLP-driven chatbots and virtual assistants to enhance citizen communication, improve access to electoral information, and counter misinformation. These tools should handle complex queries, maintain privacy standards, and adapt dynamically to evolving voter concerns [25,26].
- **Implement AI Regulation in Electoral Processes:** Adopt measures such as digital watermarking and mandatory disclosure for AI-generated political content. Such regulation prevents the spread of synthetic media, maintains transparency, and reinforces public confidence in electoral processes.
- **Promote Cross-Sector Collaboration:** Partnerships with academia, the private sector, and civil society support knowledge sharing, threat intelligence exchange, and co-development of AI tools. Collaborative approaches ensure technological, ethical, and societal considerations are integrated into AI deployment strategies [17].
- **Conduct Continuous Stress-Testing and Scenario Planning:** Regularly simulate AI-driven cyberattack scenarios, including malware, disinformation campaigns, and coordinated phishing attacks. Stress-testing allows identification of vulnerabilities, refinement of response protocols, and future-proofing of AI-driven security infrastructure [14]. These recommendations collectively emphasize the need for a balanced approach that leverages AI's benefits while addressing technical, ethical, and governance challenges. By combining technological innovation, human oversight, and robust regulatory frameworks, electoral authorities can enhance both cybersecurity resilience and public trust in democratic processes.

Conclusion

The 2024 National and Provincial Elections in South Africa illustrate the critical role of Artificial Intelligence (AI) in safeguarding electoral integrity. By operationalizing anomaly detection, dark-web monitoring, and parallel Security Operations Center (SOC) integration, the Electoral Commission significantly enhanced its resilience against cyber threats, disinformation campaigns, and operational disruptions. AI functioned as a force multiplier, enabling rapid threat detection, proactive response, and the efficient prioritization of security incidents. However, the study highlights that technical deployment alone is insufficient. Challenges remain regarding transparency, explainability, capacity-building, and governance frameworks, all of which are essential to ensuring that AI complements rather than replaces human judgment. As electoral systems worldwide continue to digitize, AI-driven cybersecurity architectures will become indispensable—but only if embedded within accountable, human-centric institutions that prioritize ethical oversight and public trust. Overall, while AI deployment substantially strengthens cybersecurity and supports electoral integrity, its sustained effectiveness depends on careful integration of technical capabilities, human expertise, and ethical governance. Addressing these limitations is critical for maintaining the robustness and reliability of AI-enhanced electoral systems in future election cycles [27].

References

1. Nye Jr, J. S. (2016). Deterrence and dissuasion in cyberspace. International security, 41(3), 44-71.
2. Shires, J., & Smeets, M. (2021). Cyber operations during elections: Threats and responses. Journal of Strategic Studies, 44(6), 789–812.
3. Kumar, R., Singh, P., & Kaur, J. (2022). Artificial intelligence for intrusion detection systems: A survey. Journal of Cybersecurity, 8(2), 145–163.
4. Sommer, R., & Paxson, V. (2010, May). Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE symposium on security and privacy (pp. 305-316). IEEE.
5. Wang, M., Zheng, K., Yang, Y., & Wang, X. (2020). An explainable machine learning framework for intrusion detection systems. IEEE Access, 8, 73127-73141.
6. Shaukat, K., & Ribeiro, P. (2020). Cyber threat detection using supervised machine learning classifiers and ensemble models. Computers & Security, 92, 101715.
7. Conti, G., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. Future Generation Computer Systems, 78, 544–546.
8. Castelvecchi, D. (2016). Can we open the black box of AI?. Nature News, 538(7623), 20.
9. Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I., & Tygar, J. D. (2011, October). Adversarial machine learning. In Proceedings of the 4th ACM workshop on Security and artificial intelligence (pp. 43-58).
10. Woolley, S., & Howard, P. (2019). Computational propaganda: Political parties, politicians, and political manipulation on social media. Oxford University Press.
11. Bradshaw, S., Bailey, H., & Howard, P. N. (2021). Industrialized disinformation: 2020 global inventory of organized social media manipulation. Computational Propaganda Project at the Oxford Internet Institute.
12. IDEA. (2024). Artificial Intelligence for Electoral Management. International IDEA Policy Paper.
13. Basu, S., & Sengupta, I. (2021). Human-in-the-loop AI for cybersecurity: Enhancing detection and response capabilities. Journal of Cybersecurity, 7(1), taab002.
14. Sommer, P., & Brown, I. (2011). Reducing systemic cybersecurity risk. Computers & Security, 30(1), 1–12.
15. Mavroeidis, V., & Bromander, S. (2017). Cyber threat intelligence model: An evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. Journal of Information Security and Applications, 36, 1–14.
16. Gwagwa, A. (2024). Generative AI and electoral resilience in South Africa. DGAP Policy Brief, 12(3), 1–12.
17. Helbing, D. (2025). AI swarms and future cyber threats to democracy. arXiv preprint arXiv:2506.06299.
18. Yin, R. K. (2018). Case study research and applications (Vol. 6). Thousand Oaks, CA: Sage.
19. Bryman, A. (2016). Social research methods. Oxford university press.
20. Wimmer, R. D., & Dominick, J. R. (2013). Mass media research. Wadsworth Publishing Company.
21. Kavanagh, J., & Rich, M. D. (2018). Digital Politics in Western Democracies: AI, Algorithms, and the Future of Elections. Routledge.
22. Al-Sarem, M., & Abbasi, A. (2020). Phishing detection using machine learning techniques: A review. IEEE Access, 8, 149832–149852.
23. Thomas, K., Beresford, A. R., & Rice, A. (2017). Security analysis of emerging cyber-threat intelligence sources. Journal of Information Security and Applications, 34, 1–10.
24. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.
25. Mou, W., & Lin, Y. (2020). Chatbots for public services: Opportunities, challenges, and ethical considerations. Government Information Quarterly, 37(4), 101495.
26. Margetts, H., & Dorobantu, C. (2019). Rethinking public engagement: AI and the future of democratic communication. Policy & Internet, 11(3), 325–345.
27. International Electrotechnical Commission. (2018). IEC 62443-1-1: Industrial communication networks—Network and system security—Part 1-1: Terminology, concepts, and models. IEC.