AI-Enabled Cybersecurity Implementation: A Case Study of Critical Success Factors in a South African State-Owned Entity

Awonke Mamane [0009-0006-1530-0606] and Rennie Naidoo [0000-0001-8392-1136]
Department of Information Systems, University of the Witwatersrand, Johannesburg, South Africa
2138857@students.wits.ac.za, Rennie.Naidoo@wits.ac.za

Abstract. The integration of AI into cybersecurity is essential for addressing complex and evolving threats. However, much of the existing AI implementation research emphasises either technical or social dimensions, neglecting their socio-technical interdependence. This study addresses this gap by identifying the critical success factors (CSFs) for AI-driven cybersecurity implementation through a socio-technical lens. Using an interpretive case study of a South African state-owned entity, the research draws on thematic analysis of in-depth interviews with technical staff and end-users. Findings reveal that successful implementation depends on the interplay between technical and social elements. Key technical CSFs include data quality, scalability, automation, and efficiency, while social CSFs encompass change acceptance, top management support, user awareness, ethical considerations, human oversight, and usability. Crucially, the study confirms that neither technical nor social factors alone are sufficient and that effective implementation depends on their interdependence. By applying a socio-technical perspective, the research offers a more balanced understanding of AI-driven cybersecurity and presents a framework to support practitioners in implementing socially integrated, technically robust solutions. Future research should further examine how human-AI collaboration can be socio-technically integrated to enhance trust, ensure ethical compliance, and improve the operational reliability of AI-enabled cybersecurity systems within organisational settings.

Keywords: artificial intelligence, cybersecurity, socio-technical systems, case study, critical success factors, implementation.

1 Introduction

Cybersecurity is a growing concern for organisations worldwide, with AI increasingly viewed as a critical enabler for enhancing cyber resilience. In South Africa, state-owned entities (SOEs) have experienced a surge in cyberattacks, exposing national data and threatening public trust [1, 2]. According to Microsoft Threat Intelligence [3], the scope, speed, and sophistication of cyberattacks are escalating, coinciding with the rapid adoption of AI technologies. These developments highlight the urgent need for more advanced, adaptive cybersecurity mechanisms. AI offers promising capabilities in this domain, particularly through machine learning, deep learning, and natural language processing, which can support real-time threat detection and automated response [4-6]. A recent CIO survey by Logicalis [7] found that over 85% of global organisations are investing in AI-driven cybersecurity. However, the effective implementation of AI-driven cybersecurity remains a complex undertaking, involving more than just technological deployment. As Taddeo et al. [8] and Pollini et al. [9] argue that such implementations are embedded within socio-technical ecosystems, systems in which human, organisational, and technological components must align. Despite this recognition, much of the IS and cybersecurity implementation literature continues to focus on either technical or social aspects in isolation [10, 11]. This fragmented approach risks overlooking the interdependent nature of the social and technical subsystems. In South Africa, several recent cyber incidents involving SOEs such as Transnet, the Department of Justice, the Companies and Intellectual Property Commission (CIPC), and the Government Employees Pension Fund illustrate how socio-technical weaknesses, including skills shortages, legacy infrastructure, and organisational resistance to change, can compromise AI-driven cybersecurity efforts [1, 12-14]. To address this gap, this study explores the critical success factors (CSFs) for AI-driven cybersecurity implementation using a socio-technical systems (STS) lens.

The research is situated in a South African SOE operating in the higher education sector, responsible for managing sensitive national qualification data. The objective is to identify the social and technical CSFs, as well as their interplay, that contribute to successful AI-driven cybersecurity implementation. Drawing on an interpretive case study, the research involved in-depth interviews with ICT professionals and end-users, with findings analyzed through thematic analysis. This study contributes to both theory and practice. Theoretically, it challenges the dominant tendency in AI-driven cybersecurity research to treat technical and social factors as isolated variables, instead assuming that implementation success is embedded in the mutual shaping of socio-technical systems. By advancing this integrated perspective, the study extends the application of socio-technical thinking to a domain often dominated by technological determinism. Practically, it offers a framework to support public sector institutions in implementing AI-driven cybersecurity solutions that are not only technically robust but also socially responsive.

2 Case Study Context

2.1. AI-Driven Cybersecurity Implementation

The case study is based on a South African SOE (referred to as EduX) operating in the higher education and qualifications verification sector. For confidentiality purposes, the name EduX is used as a pseudonym. The organisation plays a critical national role in managing and safeguarding qualification records for both domestic and international recognition. EduX employs over 150 staff members across various departments, including ICT, Administration, Evaluation Services, and Records Management. The ICT department is responsible for maintaining the institution's digital infrastructure, cybersecurity, and data integrity.

The focal point of the study is the implementation of an AI-driven cybersecurity system designed to enhance the organization's resilience to modern cyber threats. The system was introduced in response to growing national cybersecurity risks, including high-profile attacks on other SOEs such as Transnet, the Companies and Intellectual Property Commission (CIPC), and the Department of Justice in recent years. These incidents highlighted the urgency of adopting more intelligent, adaptive security solutions capable of detecting and responding to sophisticated threat vectors in real time [1, 11, 13]. The organization's pre-existing cybersecurity framework comprised legacy endpoint protection tools, perimeter-based firewalls, and manual security event monitoring, managed by a small internal cybersecurity team.

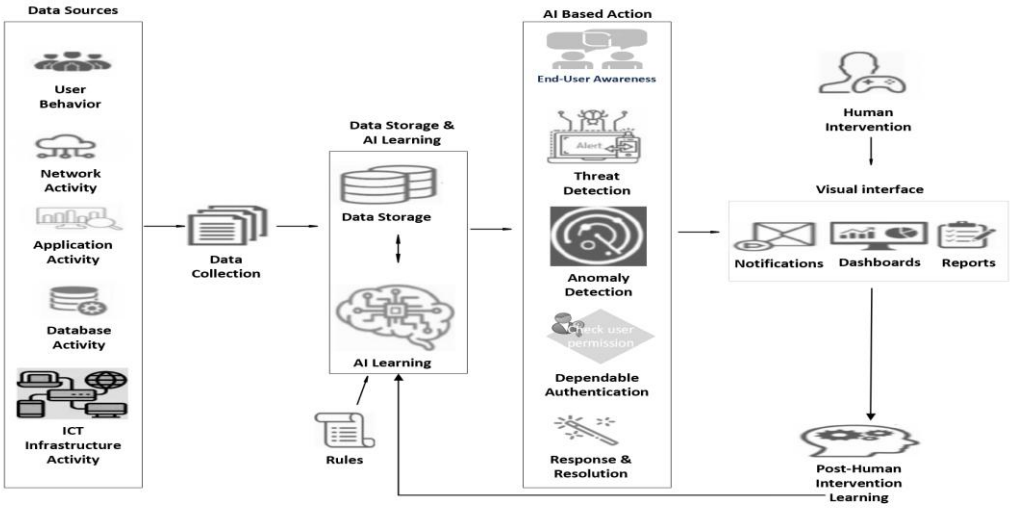

Figure 1: EduX's AI-Driven Cybersecurity architecture

The AI-driven cybersecurity project introduced a set of intelligent tools integrated with machine learning, anomaly detection, and automated response capabilities. The

implementation involved transitioning from reactive, manually intensive security operations to a more proactive, automated monitoring environment. These new tools were procured and then deployed alongside existing infrastructure, configured to integrate threat intelligence feeds, system logs, and user behaviour data. The tools came with pre-configured training models and were further trained through rules and the organisational data they ingested. As shown in Figure 1, the upgraded architecture at EduX includes a cloud-based AI threat detection engine, on-premises monitoring agents, and a central dashboard that security analysts use to triage and respond to incidents. A phased rollout was adopted to allow gradual user adaptation and technical calibration. The EduX case study explores how this AI-driven cybersecurity system was implemented within the unique context of a public sector organisation and identifies the critical social and technical factors that enabled or constrained its success.

2.2. Related Studies on AI-Driven Cybersecurity Implementation

Although AI-driven cybersecurity is a socio-technical phenomenon, not just a technological upgrade, much research treats the technical and social dimensions in isolation. It represents a fundamental transformation in how cyber risks are detected, mitigated, and managed across organisations [8]. As organisations increasingly rely on AI-driven systems, there is a need to assess both the benefits and risks associated with their integration, particularly in socio-technical environments. Several researchers have explored the technical advantages of AI in cybersecurity. For instance, Apruzzese et al. [4] and Sarker [16] highlight the efficiency of AI in enabling real-time threat detection, leveraging machine learning and anomaly detection to improve cyber defense capabilities. Bécue et al. [17] and Naseer [18] argue that AI's ability to adapt to emerging threats gives it a strategic edge over traditional, rules-based systems. Malik et al. [19] further emphasise AI's role in managing vast amounts of security data and reducing response times, noting that its predictive capabilities can help pre-empt cyberattacks before they escalate. While the technical merits are widely acknowledged, researchers are increasingly turning to the social and organisational factors that influence AI implementation outcomes. Malatji and Tolah [20] argue that AI for cybersecurity is a "double-edged sword," capable of improving security while also introducing new risks, particularly related to ethics, trust, and algorithmic bias. This aligns with Pollini et al. [9], who contend that the success of AI solutions depends on how well they are integrated into the organisational and human contexts in which they operate.

From a socio-technical perspective, Sarker et al. [21] argue that an imbalance between technical and social priorities in information systems (IS) implementation often results in suboptimal outcomes. Similar concerns are raised by Merhi [22], who highlights the importance of change acceptance and managerial support in AI implementation. However, scholars and practitioners tend to treat technical and social dimensions separately, without fully examining their interplay[23]. In contrast, Taddeo et al. [8] offer a more holistic view by conceptualising AI and cybersecurity within socio-technical ecosystems, but they do not address CSFs in detail. There is also a growing recognition that organisational readiness, leadership commitment, and user awareness play a pivotal role in determining implementation success [18, 19]. However, gaps remain in

understanding how these social factors interact with technical elements such as data quality, automation, and system scalability. This research addresses that gap by using a socio-technical lens to explore the factors for AI-driven cybersecurity implementation within a public-sector context.

The novelty of the present study lies in its dual focus: it identifies both technical and social success factors, and crucially, it examines the interdependence between them. By capturing stakeholder perspectives through qualitative interviews, the study reveals not only what matters in implementation but also why these factors succeed or fail in combination. This emphasis on organisational implications adds depth to a literature base that has traditionally prioritised technological capabilities in isolation.

3 Methodology

This case study involved qualitative fieldwork at EduX between November and December 2024. Walsham [26], Yin [27] and Runeson & Host [28] contend that case studies are suitable for investigating a complex phenomenon in its natural settings. The research focused on understanding the socio-technical CSFs underpinning the implementation of AI-driven cybersecurity. The fieldwork was structured in three main stages:

Stage 1: Preliminary Engagement and Documentation Review
Relevant internal documents, including policies, system implementation plans, and organisational cybersecurity protocols, were gathered and analysed to understand the context and evolution of AI implementation at EduX. These documents helped identify key stakeholders and shaped the semi-structured interview guide. Consent for this study was granted by Edu X and ethical clearance was granted by the University of Witwatersrand Ethics Committee (CBUSE2287). Written informed consent was also obtained from all individual participants involved in the study.

Stage 2: Data Collection through Semi-Structured Interviews
Seventeen participants were purposively selected based on their roles as technical staff, cybersecurity professionals, and business users within EduX. In line with Myers and Newman [29], semi-structured interviews were conducted to elicit insights into the challenges, enablers, and perceptions surrounding AI-driven cybersecurity implementation. Interviews lasted between 18 and 70 minutes and were audio-recorded and transcribed with Microsoft Teams with participants' consent. All transcripts were manually reviewed against the original recordings and corrected where necessary to ensure accuracy. Pseudonyms were used to ensure confidentiality.

Stage 3: Thematic Analysis and Stakeholder Interpretation
Interview recordings were analysed using ATLAS.ti, a qualitative data analysis software. Thematic analysis followed Braun and Clarke's [30] six-phase approach to identify recurring socio-technical themes and patterns across the dataset. Emergent themes were grouped into three dimensions: social, technical, and socio-technical interplay. A form of stakeholder impact interpretation was applied during analysis, which involved:

1. Identifying primary stakeholder groups (e.g., cybersecurity team, ICT support, executive managers, end-users);

2. Mapping stakeholder concerns to social and technical categories;
3. Exploring how changes (e.g., automation, ethics, usability) affected stakeholders' values, workflows, and decision-making;
4. Contextualising the findings using a socio-technical lens. This allows us to view these integral components of the system that directly affect implementation outcomes;
5. Interpreting how stakeholders viewed AI-based cybersecurity implementation, emphasising core social concerns that must be negotiated alongside technical functionality. Reflexivity was maintained throughout the analysis, with the researcher critically examining their own positionality and potential biases.

The next section presents the findings of the case study, organised according to the identified socio-technical themes and supported by illustrative quotes from participants. To protect confidentiality, all participants are referred to using pseudonyms.

4 Results

This study explored the socio-technical CSFs influencing the implementation of AI-driven Cybersecurity at EduX. Figure 2 shows the three overarching themes that structured our analysis: social factors, technical factors, and their socio-technical interplay.

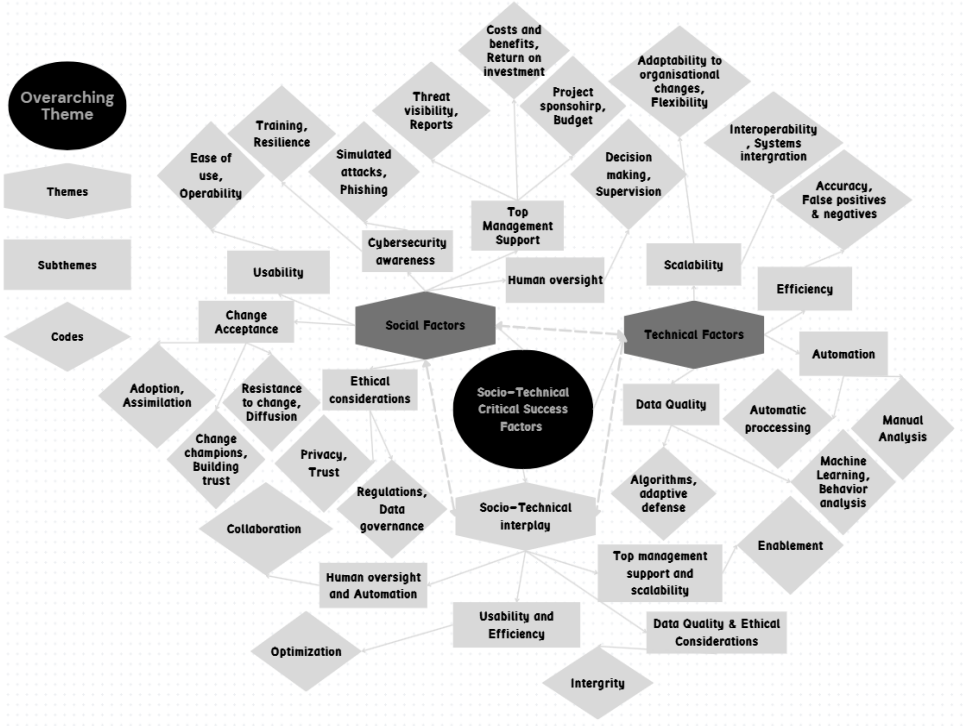

Figure 2: Thematic analysis map

4.1 Social Factors

Six key social factors critical to the successful implementation of AI in cybersecurity were identified: change acceptance, top management support, cybersecurity awareness, ethical considerations, human oversight, and usability.

4.1.1 Change Acceptance

Change emerged as both a technical and emotional process, requiring deliberate planning, communication, and sensitivity to resistance. Participants described how AI implementation challenged existing routines and skillsets, requiring cultural shifts and structural adjustments. Alfred Manyika (ICT Senior Manager) described the initial resistance faced and the structured strategy used to overcome it: "*Firstly, it is the people as people aspect. People didn't want to change… there was not enough skill to support the changes… Then the second aspect was the processes… we had to create new policies and procedures… then lastly the technology part."* His emphasis on a "three-pronged" strategy, people, process, and technology, highlighted the socio-technical framing of the initiative. He further explained that an agile, phased rollout helped ease the transition: *"The strategies that we used were sort of like an agile approach… implement them slowly so that the users can get used rather than… a big bang approach."* This phased approach was essential not just to reduce resistance, but to align people with the evolving systems. Joice Mphahlele (ICT Senior Manager) emphasised the importance of advance communication to build readiness: *"Communication was prepared well in advance, to sensitize these users as to what to expect… I think that has provided some level of comfort."* Amelia Smith (Manager, User) echoed this perspective, noting the influence of formal communication from leadership: *"It was an e-mail from the senior manager IT to inform us that these interventions will be enrolled…"* These comments reflect the central role of communication in managing uncertainty and shaping perception. Participant Rebamang Lekoelea (ICT Manager) further underlined the importance of readiness: *"We have to look at the readiness of the organisation… is the organisation ready or maybe is it a big jump?"* Resistance was also tied to established organisational norms. Manyika acknowledged this cultural inertia: *"People are naturally resistant to change… We slowly changed the culture of the organisation."*

To manage this transformation, the organisation implemented a change champion strategy. John Mokwena (ICT Change Champion, User), one such champion, explained: "*It is basically a group of people… to soften the blow… letting them [users] know of any major changes… also giving feedback."* Together, these insights show that change acceptance was not a single intervention but a continuous, socially embedded process involving communication, engagement, and empowerment.

4.1.2 Top Management Support

Support from top management at EduX emerged as both a symbolic and practical enabler. While executives may not always possess deep technical knowledge, their role in sponsoring initiatives and allocating resources was seen as non-negotiable. Lynn Manuel (Chief Executive) articulated this dynamic: *"The executives are very much*

supporting and sponsoring these initiatives whilst not necessarily holding the technical competence… Certainly from the executive's perspective it's a non-negotiable." This symbolic sponsorship was matched by tangible decisions around budget and staffing. Adrian Makwetla (ICT Change Champion, User) stressed the financial realities: *"You must make sure that you got budget… You need money to get enough people with expertise."* The success of AI-driven cybersecurity was directly tied to budgetary allocations, as well as leadership buy-in. However, the analysis also revealed gaps, when executives lacked technical insight, they could underestimate risk or delay decisions. Manyika pointed to this challenge: *"Some of the sponsors don't really know… the impact… if these stakeholders don't fully understand, it becomes difficult to secure a buy-in."* Executives also played a crucial role in advocating for security culture. Mphahlele described a unified strategy: *"Make sure that we have top management buy-in… so that we move on the same page with the organisation."* The executive role extended into accountability and enforcement. Alex Bhingwa (IT & Finance Executive) described how top leadership intervened when awareness was lacking: *"We have even gone to the extent of saying those people that would not have completed their awareness trainings should be noted."* Their engagement also had strategic foresight. Gauta Mothapo (Operations Executive) noted the formation of a task team to explore AI's broader role*: "We put a test team together for them to be able to look at what is the impact of AI within our work environment."* These accounts highlight the critical gatekeeping role played by top management, providing not only funding, but also political will, symbolic endorsement, and strategic direction.

4.1.3 Cybersecurity Awareness

Participants consistently linked successful implementation with organisational awareness of cybersecurity risks. AI tools, while powerful, depend on users' understanding of their role in the security chain. Andiswa Modise (Senior Manager, User) underscored the democratization of cybersecurity knowledge: *"It has really helped a lot… it has given an understanding that you cannot think that cyber threats… are only touching on ICT people."* Smith echoed this need for inclusive awareness: *"Especially those of us who are not in the IT space… make us aware of possible risks."* Beyond training, awareness was personalized through interactive tools like phishing simulations and gamified learning. This translated into observable changes in user behaviour, as noted by Anele Moloto (ICT Specialist):

"We haven't had much of our users that will just click on a link… or provide information to impersonators…" Participants also reflected on how awareness contributed to reducing threat detection noise. Jacob Motaung (ICT Specialist) observed: *"Cyber awareness is important… unaware users can lead to compromises or a large number of alerts."* Pule Molefe (User) shared an instance where he encountered impersonation prevention due to poor awareness: *"I got a notification that it's like a person is impersonating me… So I recalled this is one of the AI security tools that has been implemented."* In sum, the analysis suggests that AI effectiveness is amplified by human awareness. Educated users make fewer errors, raise fewer alerts, and support system efficacy.

4.1.4 Human Oversight

Participants consistently expressed the need for human judgment in tandem with AI. While automation enabled rapid response, it was human oversight that ensured contextual accuracy. Moloto stated the necessity plainly: *"While automation can be useful… it's important to ensure that a human is still involved in the decision-making process."* This need was driven by complexity, explainability and error correction. Lekoelea warned that humans must be skilled to oversee AI effectively: "*The right skill set in terms of managing that and other organisational challenges."* Bhingwa added that technical and business acumen must co-exist: *"People that… have a deep understanding of both the technical and the business side [of] AI projects are needed."* Participants also shared instances where AI blocked legitimate communication, requiring human intervention. Akani Mathebula (User) recalled: *"We could not locate that email… but then after the intervention from IT we were able to locate it."* David Radebe (User) noted the operational risks of delayed intervention: *"It becomes a bit of a challenge… if maybe there was no quick response time… it would have had an impact on my performance."* These accounts point to a human-AI partnership where oversight is not optional but essential, both to uphold system reliability and to mitigate unintended disruptions.

4.1.5 Ethical Considerations

The participants viewed ethical considerations as foundational, not only to regulatory compliance, but to trust in AI systems. Concerns spanned data sovereignty, transparency, and fairness. Mphahlele emphasised data localization: *"The tools that we use… originate from outside the country… where does [the data] reside?"* To mitigate risk, policies were developed internally. Alex Bhingwa explained: *"We are putting up ethical guidelines in the form of policies so that we can regulate this."* Trust was a major concern. Johan De Bruyn (Software Developer) questioned data transparency: *"The provider can be spying on you… they're not transparent on how data is used and that is not ethical."*

Concerns about surveillance were echoed by Banele Mahlangu (User), who raised issues around email interception: *"There is a concern for a privacy… information from supply chain that must not be shared with other units."* These comments highlight the ethical fragility of AI, where decisions made at design and policy levels can undermine trust and inclusivity at the user level.

4.1.6 Usability

Usability concerns revealed a tension between security and user experience. While AI improved protection, participants stressed the importance of seamless integration into daily workflows. Manuel reflected frustration at untimely disruptions: "*When these things happen… it's inconvenient… it delays us… and creates a bit of frustration."* Mahlangu echoed this operational impact: *"We are working on an RFQ and then the email gets blocked… we have to go back to ICT to request access."* From the technical side, however, usability was positively framed. Moloto described her experience: *"I find it easy to view data, incidents and other resources."* This divergence in perception underscores that usability is role-specific. What is user-friendly for ICT staff may feel

obstructive to administrative staff. Still, the analysis shows that without smooth usability, even secure systems can encounter resistance.

4.2 Technical Factors

Four key technical factors critical to the successful implementation of AI in cybersecurity were identified: data quality, scalability, automation, and efficiency.

4.2.1 Data Quality

Participants consistently raised concerns about the integrity and fairness of the data that feeds AI-driven cybersecurity systems. Lynn Manuel asserted that AI systems may carry built-in cultural and geographic biases. She noted, *"People that develop these cutting-edge tools have a certain perspective. That perspective is fed into the AI algorithms, and the African marginalized people are the ones that generally take the brunt of that."* Echoing this concern, Alex Bhingwa expressed hesitation about over-reliance on AI due to the risk of poor data integrity: *"I've always worried about data integrity issues... I mean it has inherent biases, if we over depend on this AI this could give problems."* Participants also acknowledged the influence of compromised or poor-quality training data on AI's decision-making. Adrian Makwetla remarked, *"The information that we are using in the AI technology might be compromised or biased… I have concerns of integrity in terms of what the data produces and how the information was sourced."* Johan De Bruyn introduced a cybersecurity angle, raising concerns about malicious manipulation: *"They can get hacked, and your data can come up in other people's searches."* These concerns point to an overarching unease about the ethical, accurate, and secure training of AI models in cybersecurity.

4.2.2 Scalability

Scalability was viewed as essential to ensure that AI solutions grow alongside organisational needs. Alex Bhingwa highlighted this imperative, stating: "*As the organisation we are growing, we have to ensure that AI is scalable and not static... We need a robust integration strategy and architecture."* Participants stressed that scalability included not only the expansion of user base and processes but also compatibility with new technologies. Jacob Motaung affirmed the importance of interoperability: *"They [AI-Driven cybersecurity systems] have to be interoperable with existing infrastructure."* However, some raised challenges around outdated systems. Mphahlele admitted: "*I think I have picked up that there were issues related to lack of compatibility or integration."* Alfred Manyika added: *"It has been very challenging given that the existing systems [use] legacy technology... So it has led us now to start looking into a holistic road map… to replace them with newer and faster systems."* This reflects the tension between strategic ambition and operational legacy, reinforcing the need for phased infrastructure renewal.

4.2.3 Automation

Automation was widely seen as a positive feature of AI, easing the workload on technical teams. Anele Moloto described the change: *"We once used [a tool that] made it*

difficult to detect insider threats… But thanks to AI, which is able to do this for us… and provide security orchestration and automated response." Jacob Motaung confirmed that automation reduced manual effort: *"AI actually helps a lot because it doesn't require one to always be manually checking... it alleviates all that hard work."* Manyika added a strategic view, highlighting how automation optimized service delivery: *"So once we have managed to automate certain processes, it becomes easier… and make sure that we deliver our services quite fast."* From a resourcing perspective, Bhingwa saw automation as a budgetary lever: *"AI can help us address the cybersecurity skills gap... maybe we need more hands, but we don't have the budget. So, then we can then automate some processes."* Together, these accounts reinforce the dual value of automation, operationally for speed and accuracy, and strategically for resource optimization.

4.2.4 Efficiency

Participants recognized efficiency as a multidimensional outcome, but one heavily dependent on accurate AI performance. Rebamang Lekoelea emphasised that "efficiency is not one-size-fits-all," noting that perceptions of success varied between individuals and roles. Concerns about false positives and false negatives emerged repeatedly. Pule Molefe warned that *"AI can bring about false negatives or false positives. It needs training… the data that was used to train this particular technologies might still need to be improved."* Akani Mathebula linked this to user experience: "*The shortcomings will be sometimes missing the information employees require because of false blacklisting of an e-mail."* John Mokwena drew a line between detection errors and client service: *"It can increase turnaround time for that specific client... but with quick response time from ICT, then it gets solved."* Participants' insights confirm that efficiency is not simply speed but also about precision and minimal disruption. Accuracy in anomaly detection was key to maintaining trust and continuity in business operations.

4.3 Socio-Technical Interplay

Several participants at EduX highlighted tensions and synergies across four key interlinkages: human oversight and automation, top management support and scalability, usability and efficiency, and data quality and ethical considerations.

4.3.1 Human Oversight and Automation

While automation was largely celebrated for improving detection and reducing manual effort, participants stressed that it cannot fully replace human judgment. Instead, the findings reveal a complementary relationship, one where AI and human decision-making must co-evolve. Alex Bhingwa reflected on this dynamic: *"I think one of the things that I'm still grappling with is how do we find an optimal balance in human and AI interaction, striking the perfect balance between automation and human intervention."* His comment encapsulates a central tension in AI-driven cybersecurity, trust in machine efficiency must be balanced with the unpredictability and nuance that only human oversight can resolve. Across roles, participants understood automation as an enabler, but

not a substitute for informed human reasoning, particularly in ambiguous or complex threat environments.

4.3.2 Top Management Support and Scalability

A second key interplay lies between strategic sponsorship from top management and the scalable design of AI infrastructure. Participants explained that scaling AI-driven cybersecurity systems is not merely a technical exercise but a resource-driven one, and executive backing is essential. Alex Bhingwa again articulated this strategic concern: *"I am very positive and convinced that there is a return on investment, and we shall continue to invest in this area to make sure that our systems are really watertight."* This reflects a shared belief that robust, scalable AI systems require not only technical flexibility, but continued organisational commitment. Without executive vision and sustained budget allocation, participants indicated, even well-designed AI systems might fail to evolve with changing operational demands.

4.3.3 Usability and Efficiency

Usability emerged as both a technical and experiential theme. It was not just about functionality but also the user's emotional and cognitive experience. The analysis found that user perceptions of system efficiency were directly tied to how usable and unobtrusive the technology felt. For non-technical users, false positives were a major frustration. Lynn Manuel explained: *"When these things happen, it's late at night or on a weekend and then you have to revert to a human being when the AI is not helping you… it does create delays and a bit of frustration."* Banele Mahlangu shared a similar sentiment: *"There's been hiccups… especially the system that is used to detect any threats on the emails… we have to go back to ICT to request for access… it does delay us."* These reflections demonstrate how perceived inefficiencies, even when rare, can erode user trust. For AI to be effective in practice, it must not only detect anomalies but do so without impeding routine tasks or requiring frequent manual intervention. The success of AI is, therefore, as much about social acceptance and workflow alignment as technical accuracy.

4.3.4 Data Quality and Ethical Considerations

The final interplay explored concerns the relationship between data quality and ethics. Participants drew attention to how poor or biased data not only reduced detection accuracy, but also raised significant ethical concerns around fairness, trust, and transparency. Mphahlele expressed concern about data jurisdiction: *"One would have to look at when they collect this data, where does it reside and all those kinds of things."*

Johan De Bruyn voiced his skepticism more directly: "*I think they're [AI-driven cybersecurity providers] not transparent on how data is used."* Participants were especially sensitive to ethical risks around surveillance, data sovereignty, and implicit bias in training data. For instance, Lynn Manuel warned against algorithmic discrimination: *"Africa[n] marginalized people are the ones that generally take the brunt of that… like AI blocking all African things because Africa is bad."* These examples illustrate how data quality cannot be divorced from ethical reflection. The effectiveness of AI models

is shaped not just by computational performance, but by their cultural awareness, legal compliance, and moral grounding.

5 Discussion

AI is increasingly recognized by scholars as a disruptive capability in the cybersecurity domain, with the potential to enhance threat detection, automate incident response, and augment human expertise [14, 15]. The findings of our EduX case study suggest that, while AI offers significant functional and strategic advantages for cybersecurity, its successful implementation within an SOE is contingent on critical socio-technical factors. As suggested by previous research, these factors encompass both technical enablers, such as data infrastructure and model interpretability, as well as social and organisational dynamics, including trust, skill development, change acceptance, and ethical alignment [12, 13]. Similar to Merhi [22], the case study revealed that AI for cybersecurity implementation was not merely a technological deployment but rather a process of organisational transformation. Although stakeholders acknowledged the performance advantages of AI (e.g., real-time analysis, pattern recognition, anomaly detection), their adoption was moderated by uncertainty regarding accountability, transparency, and role redefinition. These concerns were particularly salient among operational staff who feared deskilling and increased reliance on opaque decision systems. Despite these reservations, participants recognized AI as a valuable augmentation tool, provided its integration is supported by adequate training, ethical safeguards, and human oversight mechanisms. This reinforces the argument that AI, when applied to critical domains such as cybersecurity, must be positioned not as a replacement but as a collaborator within existing human–machine workflows. The socio-technical stakeholder analysis demonstrated a divergence in perceived benefits and risks across organisational levels. Senior leadership and technical managers largely viewed AI as a catalyst for innovation and institutional modernization. In contrast, administrative and compliance officers identified risks such as role redundancy, organisational silos, and unintended bias in algorithmic decision-making. This fragmentation in perception underscores the importance of inclusive, participatory implementation strategies that involve end-users from early stages and embed co-design principles [21].

The EduX study has clear practical implications for organisations pursuing AI-enabled cybersecurity. First, the identified CSFs, especially those relating to human readiness, ethical alignment, and organisational coordination, can serve as a checklist for practitioners seeking to build resilient implementation plans. Second, the risks identified in this study should be documented and incorporated into a project-specific risk register, to ensure appropriate mitigation strategies are developed. Finally, the stakeholder mapping and analysis method presented here can be adopted by other public or private organisations to anticipate barriers and identify enablers specific to their institutional context.

This study also contributes to the growing body of socio-technical systems theory by extending its application to the domain of AI-enabled cybersecurity within public

sector institutions [6, 8, 9, 23]. While prior research has focused extensively on technical architectures or threat modelling [4, 18], this study foregrounds the socio-organisational dimensions that shape the success or failure of AI deployments in complex institutional environments. First, the research advances our understanding of socio-technical alignment by identifying the interplay between human factors (trust, resistance, capability-building) and system-level enablers (data quality, algorithm transparency, oversight mechanisms) in a high-stakes domain. The findings support and extend existing frameworks that argue for the co-evolution of social and technical subsystems, particularly in environments marked by institutional inertia, regulatory burden, and hierarchical governance structures [8, 9]. Second, the study introduces a context-sensitive operationalization of CSFs specific to AI-driven cybersecurity. These CSFs move beyond generic IT implementation checklists and include AI-specific elements such as model interpretability, ethical risk calibration, and the procedural embedding of human-in-the-loop controls. In doing so, the study addresses a conceptual gap between AI systems research and applied organisational change theory. Third, by applying and refining stakeholder impact analysis within a socio-technical context, the study demonstrates how perceived risks and benefits of AI adoption vary across organisational strata. This empirical insight contributes to theoretical debates on technology acceptance and organisational sensemaking by showing how perceived value is socially constructed, role-contingent, and shaped by institutional power dynamics. Lastly, the study contributes to the broader theorization of AI not as a purely technological artefact, but as an agent of organisational transformation [6, 17]. It reinforces the proposition that AI systems must be understood as embedded, contested, and negotiated within social contexts, thus supporting a critical, interpretivist view of technology implementation in the public sector.

6 Conclusion and Future Work

AI is transforming the cybersecurity landscape with its potential to detect, deter, and respond proactively to evolving threats. Yet, the socio-technical nature of AI deployment in public sector environments, particularly within state-owned entities, introduces complex challenges that extend beyond technical feasibility. This case study highlights that successful AI-driven cybersecurity implementation demands a balanced consideration of both the social (human, organizational, ethical) and technical (infrastructure, system design, data integration) subsystems. The findings of this research suggest that socio-technical alignment, through inclusive governance, stakeholder engagement, contextual responsiveness, and capacity building, is central to the effective implementation of AI-enabled cybersecurity in resource-constrained public institutions. Despite the theoretical promise of AI, several CSFs emerged as pivotal: the presence of clear policies and leadership support, alignment with organizational priorities, ethical data practices, and the cultivation of trust between human actors and AI systems. These findings reinforce that while AI technologies offer clear operational advantages, their deployment in public sector cybersecurity systems is not purely a technical undertaking, it is a socio-technical transformation process. The limitations of this study include

its single-case design and its focus on perceptions and practices within one South African state-owned entity. Broader generalizability is therefore constrained, although the qualitative depth provides rich insights into emerging challenges and enablers. Additionally, the rapid pace of AI innovation means that any implementation strategy must remain adaptable to technological and regulatory changes.

However, the study is subject to several limitations. It employed a single-case study design focused on one SOE, which constrains the generalizability of the findings to other national or institutional contexts. The sample was relatively small and purposive, concentrating on internal ICT professionals and end-users.

Looking ahead, future theory-building efforts should aim to generalise these insights across domains and geographies, and further explore the dialectical tensions between efficiency gains promised by AI and the ethical, organisational, and relational complexities it introduces. Future research should also investigate longitudinal implementations of AI in cybersecurity across multiple public sector institutions to explore the evolution of trust, governance models, and interdependencies over time. Comparative studies between state-owned and private entities could also illuminate differences in socio-technical coordination. Furthermore, we propose implementation frameworks that incorporate socio-technical dimensions, enabling more grounded, ethically-informed, and context-sensitive AI deployment.

Disclosure of Interests. The authors declare that they have no competing interests.

References

[1] H. Pieterse, "The Cyber Threat Landscape in South Africa: A 10-Year Review," *Afr. J. Inf. Commun.*, vol. 28, 2021, doi: 10.23962/10539/32213.

[2] S. Timcke, M. Gaffley, and A. Rens, "The centrality of cybersecurity to socioeconomic development policy: A case study of cyber-vulnerability at South Africa's Transnet," *Afr. J. Inf. Commun. AJIC*, no. 32, pp. 1–28, Dec. 2023, doi: 10.23962/ajic.i32.16949.

[3] Microsoft Threat Intelligence, "Staying ahead of threat actors in the age of AI," Microsoft Security Blog. Accessed: Feb. 24, 2025. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/

[4] G. Apruzzese *et al.*, "The Role of Machine Learning in Cybersecurity," *Digit. Threats Res. Pract.*, vol. 4, no. 1, pp. 1–38, Mar. 2023, doi: 10.1145/3545574.

[5] I. Lee, "Cybersecurity: Risk management framework and investment cost analysis," *Bus. Horiz.*, vol. 64, no. 5, pp. 659–671, Sept. 2021, doi: 10.1016/j.bushor.2021.02.022.

[6] R. Gafni and Y. Levy, "The role of artificial intelligence (AI) in improving technical and managerial cybersecurity tasks' efficiency," *Inf. Amp Comput. Secur.*, vol. 32, no. 5, pp. 711–728, July 2024, doi: 10.1108/ICS-04-2024-0102.

[7] Logicalis, "CIO Report." Accessed: Feb. 24, 2025. [Online]. Available: https://www.za.logicalis.com/cio-report

[8] M. Taddeo, P. Jones, R. Abbas, K. Vogel, and K. Michael, "Socio-Technical Ecosystem Considerations: An Emergent Research Agenda for AI in Cybersecurity," *IEEE Trans. Technol. Soc.*, vol. 4, no. 2, pp. 112–118, June 2023, doi: 10.1109/TTS.2023.3278908.

[9] A. Pollini *et al.*, "Leveraging human factors in cybersecurity: an integrated methodological approach," *Cogn. Technol. Work*, vol. 24, no. 2, pp. 371–390, May 2022, doi: 10.1007/s10111-021-00683-y.

[10] B. Guembe, A. Azeta, S. Misra, V. C. Osamor, L. Fernandez-Sanz, and V. Pospelova, "The Emerging Threat of Ai-driven Cyber Attacks: A Review," *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2037254, Dec. 2022, doi: 10.1080/08839514.2022.2037254.

[11] D. Araya, "Cybersecurity and AI," Centre for International Governance Innovation, 2022. Accessed: Feb. 24, 2025. [Online]. Available: https://www.jstor.org/stable/resrep42557.11

[12] Brett Van Niekerk, "An Analysis of Cyber-Incidents in South Africa," *Afr. J. Inf. Commun. AJIC*, no. 20, Dec. 2017, doi: 10.23962/10539/23573.

[13] R. Dagada, "The Advancement of 4IR Technologies and Increasing Cyberattacks in South Africa," *South. Afr. J. Secur.*, vol. 2, p. 27 pages-27 pages, Mar. 2024, doi: 10.25159/3005-4222/15157.

[14] J. Mtsweni and M. Thaba, "Building an Integrated Cyber Defence Capability for African Missions," *J. Inf. Warf.*, vol. 21, no. 1, pp. 17–34, 2022.

[15] N. Siphambili, O. Mahlasela, E. Baloyi, and E. Mukondeleli, "A Review of the South African Public Sector's Capability in Combating Ransomware," in *2024 4th International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, Nov. 2024, pp. 493–499. doi: 10.1109/IMITEC60221.2024.10850969.

[16] I. H. Sarker, "Detecting Anomalies and Multi-attacks Through Cyber Learning: An Experimental Analysis," in *AI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability*, I. H. Sarker, Ed., Cham: Springer Nature Switzerland, 2024, pp. 61–77. doi: 10.1007/978-3-031-54497-2_4.

[17] A. Bécue, I. Praça, and J. Gama, "Artificial intelligence, cyber-threats and Industry 4.0: challenges and opportunities," *Artif Intell Rev*, vol. 54, no. 5, pp. 3849–3886, June 2021, doi: 10.1007/s10462-020-09942-2.

[18] H. Naseer, K. Desouza, S. B. Maynard, and A. Ahmad, "Enabling cybersecurity incident response agility through dynamic capabilities: the role of real-time analytics," *Eur. J. Inf. Syst.*, vol. 33, no. 2, pp. 200–220, Mar. 2024, doi: 10.1080/0960085X.2023.2257168.

[19] S. Malik, P. Malik, and A. Naim, "Opportunities and Challenges in New Generation Cyber Security Applications Using Artificial Intelligence, Machine Learning and Block Chain," 2024, pp. 23–37. doi: 10.1007/978-981-97-1249-6_2.

[20] M. Malatji and A. Tolah, "Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI," *AI Ethics*, Feb. 2024, doi: 10.1007/s43681-024-00427-4.

[21] S. Sarker, S. Chatterjee, X. Xiao, and A. Elbanna, "The sociotechnical axis of cohesion for the IS discipline: its historical legacy and its continued relevance," *MIS Q*, vol. 43, no. 3, pp. 695–720, Sept. 2019, doi: 10.25300/MISQ/2019/13747.

[22] M. I. Merhi, "An evaluation of the critical success factors impacting artificial intelligence implementation," *Int. J. Inf. Manag.*, vol. 69, p. 102545, Apr. 2023, doi: 10.1016/j.ijinfomgt.2022.102545.

[23] M. Malatji, A. Marnewick, and S. von Solms, “Validation of a socio-technical management process for optimising cybersecurity practices,” *Comput. Secur.*, vol. 95, p. 101846, Aug. 2020, doi: 10.1016/j.cose.2020.101846.

[24] N. Mohamed, “Current trends in AI and ML for cybersecurity: A state-of-the-art survey,” *Cogent Eng.*, vol. 10, no. 2, p. 2272358, Dec. 2023, doi: 10.1080/23311916.2023.2272358.

[25] W. Yeoh, S. Wang, A. Popovič, and N. H. Chowdhury, “A systematic synthesis of critical success factors for cybersecurity,” *Comput. Secur.*, vol. 118, p. 102724, July 2022, doi: 10.1016/j.cose.2022.102724.

[26] G. Walsham, “Interpretive case studies in IS research: nature and method,” *Eur. J. Inf. Syst.*, 1995.

[27] R. K. Yin, *Case Study Research: Design and Methods*. SAGE, 2009.

[28] P. Runeson and M. Höst, “Guidelines for conducting and reporting case study research in software engineering,” *Empir. Softw. Eng.*, vol. 14, no. 2, pp. 131–164, Apr. 2009, doi: 10.1007/s10664-008-9102-8.

[29] M. D. Myers and M. Newman, “The qualitative interview in IS research: Examining the craft,” *Inf. Organ.*, vol. 17, no. 1, pp. 2–26, Jan. 2007, doi: 10.1016/j.infoandorg.2006.11.001.

[30] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.