Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

AI in Cybersecurity: A Systematic Review and Conceptual Audit Model

Ndaedzo Rananga[1], H.S Venter[2]

[1,2]Faculty of Engineering, Built Environment and Information Technology, Department of Computer Science, University of Pretoria, Pretoria, South Africa

Received:
October 20, 2025
Revised:
March 27, 2026
Accepted:
April 12, 2026
Published:
April 22, 2026

Corresponding Author:
Author Name*:
Ndaedzo Rananga
Email*:
u11329892@tuks.co.za

DOI:
10.63158/journalisi.v8i2.1579

Abstract. Technological advances, particularly in Artificial Intelligence (AI), are accelerating digital transformation while increasing system complexity and exposure to sophisticated cyber threats. These developments challenge traditional cybersecurity audit approaches, which are largely periodic, retrospective, and focused on binary control checks. In response, the adoption of generative AI (GenAI) and predictive AI (PredAI) in cybersecurity auditing is becoming increasingly important. Although AI can improve audit intelligence, scalability, timeliness, and effectiveness, its use also raises concerns about transparency, governance, and auditor independence. This study employed a two-stage methodology. First, a systematic literature review following PRISMA examined studies published between 2021 and 2026, yielding 36 eligible articles. The review found that hybrid AI approaches dominate the literature (58.3%), followed by GenAI (25.0%) and PredAI (16.7%). Despite this growing interest, the literature gives limited attention to risk-based auditing approaches that move beyond binary control confirmation toward context-aware, intelligence-driven cyber risk assessment. Second, using Design Science Research, the study developed the conceptual Anti-Sheriff cybersecurity auditing model. The model shifts auditing from compliance-driven enforcement to intelligence-supported risk governance, enabling continuous auditing, better risk prioritisation, and stronger organisational cyber resilience.

Keywords: Cybersecurity auditing; Artificial intelligence; Generative AI; Risk-based auditing; Continuous auditing

1. INTRODUCTION

The information and communication technology (ICT) ecosystem has become increasingly saturated with interconnected devices and systems [1], [2]. This rapid interconnectivity is largely driven by the proliferation of emerging technologies, including the Internet of Things (IoT), mobile cloud computing, ubiquitous computing, 5G, and artificial intelligence (AI) [3]. As demonstrated by Rananga and Venter [1], a key factor supporting the adoption and appreciation of modern technologies lies in improved access to computational capabilities. Even remote regions in developing economies now have access to such capabilities. Faturoti [4] indicates that changes in how people conduct business and exchange information following the COVID-19 pandemic have accelerated internet adoption across African regions. AI technologies are also being adopted in sensitive sectors such as healthcare, as reported by Bharati et al. [5].

Although small organizations, such as small and medium enterprises (SMEs), continue to encounter several challenges in adopting modern technologies [6], the adoption of technology adavancement within the SME ecosystem remains promising [1]. In essence, no sector is exempt from technological advancements such as AI and machine learning (ML). Digital transformation has become central to the global economy, with an estimated value of approximately $7 trillion of the United States gross domestic product (GDP) expected to rely on or arise from AI adoption, as established by Mo and Ouyang [7].

The advantages associated with modern technological advancement should be recognized; however, digital innovation and operational efficiency are also increasing system complexity and intensifying regulatory scrutiny. The extensive interconnection of digital assets has contributed to the rapid evolution of cybersecurity threats [8]. Cybersecurity threats are becoming more sophisticated, adaptive, and harder to detect. Of concern, these days, cyberthreats may also affect mental health, as cyberbullying has direct psychological consequences [9]. Notably, the application of smart manufacturing and intelligence in autonomous vehicles and health care is experiencing a surge in adoption [3]. Cyberthreats targeting intelligent transportation systems and autonomous vehicles [10] also present a risk to human life. As also strongly emphasized in Calvo et al. [2], cybersecurity threats, such as denial of service, can have a severe impact on business operations, costing organizations millions of dollars due to lost hours from downtime.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Rananga and Venter [11] emphasize that technological adoption should be undertaken with due diligence, ensuring that operational, technological, and managerial dimensions are considered. Traditionally, the effectiveness of information systems (IS) controls has been assessed through a binary audit process. This approach relies heavily on checklists to verify whether controls exist. Changes in business operations and patterns of human interaction increasingly challenge the relevance and effectiveness of traditional IS auditing approaches, which were designed for more static and bounded environments. The relevance of conventional cybersecurity controls in the modern AI era has also been questioned by Irshad et al. [8] and Darwish et al. [12].

Beyond traditional IS auditing challenges, existing specialized cybersecurity audit approaches, including cybersecurity audit and data analytics practices. These areas are also under strain, as they rely on periodic, control-centric, and retrospective assessments that struggle to capture dynamic risk exposure and emerging threat behaviors in real time. The increasing sophistication of cybersecurity threats, which frequently occur in real time, renders overreliance on the existence of cybersecurity controls and historical cyber incidents potentially misleading. In most instances, reactive control assessment can misguide the actual security posture of organizations being audited (referred to herein as auditees). This misalignment raises concerns about the adequacy of current audit practices in providing meaningful assurance within modern ICT environments. Hossain et al. [8] indicate that modern cybersecurity threats, such as zero-day exploits, polymorphic malware, and large-scale distributed denial-of-service attacks, call into question the relevance of traditional network intrusion detection systems and intrusion prevention systems.

Conversely, the influence of GenAI and PredAI within the IS auditing domain is increasingly evident. These approaches present significant opportunities to enhance audit intelligence, scalability, and timeliness, particularly in cybersecurity-focused audits. Their adoption also introduces uncertainty related to governance, accountability, transparency, and auditor independence. If these uncertainties remain unresolved, the use of GenAI and PredAI may be perceived by stakeholders as an extension of surveillance or excessive policing mechanisms (referred to herein as the Sheriff audit approach), potentially leading to resistance and reduced stakeholder participation.

In response to these challenges, this study adopts a twofold approach. First, it evaluates the contributions of GenAI and PredAI to advancing cybersecurity resilience and advocates for more robust, intelligence-driven cybersecurity auditing practices. Second, it proposes a balanced approach to AI integration that prioritizes explainability, accountability, and socially responsible deployment. Morales-Navarro et al, [13]. emphasize the importance of embedding human-centered computing principles within the AI era, reinforcing the governance imperative underpinning this research.

In other words, despite the growing body of research on AI in cybersecurity, a critical gap remains in cybersecurity auditing and assurance. Existing studies predominantly emphasize AI-driven capabilities such as threat detection, anomaly classification, and predictive intrusion modelling, with limited attention to how these capabilities can be systematically integrated into the cybersecurity audit lifecycle. As a result, current approaches remain fragmented, with binary compliance verification, AI-driven intelligence, and human governance treated as isolated processes rather than components of a unified assurance framework. This fragmentation reinforces a checklist-driven, enforcement-oriented ("Sheriff") audit paradigm that is increasingly inadequate for evaluating dynamic risk exposure in complex digital environments. Consequently, there is a lack of structured, governance-oriented models that integrate AI capabilities with contextual analysis and human judgment to support continuous, intelligence-driven cybersecurity assurance. This study addresses this gap by proposing a conceptual cybersecurity audit model that repositions auditing from static compliance verification toward an integrated, risk-aware, and intelligence-driven assurance approach.

At its core, this paper presents a systematic review of the application of AI within cybersecurity auditing and introduces the conceptual Anti-Sheriff cybersecurity auditing approach. The proposed conceptual model is designed to shift the paradigm from enforcement-driven control expansion toward intelligence-driven augmentation of audit processes. Rather than positioning AI as a policing mechanism, the proposed model integrates GenAI and PredAI across the cybersecurity audit lifecycle to enhance analytical depth, contextual interpretation, and adaptive risk assessment. By integrating cybersecurity auditing with AI-driven intelligence, the approach supports continuous assurance, risk-based prioritization, and strengthened organizational resilience within modern information ecosystems. More importantly, the proposed approach is not an

attempt to replace automated cybersecurity auditing assessment, but rather to enhance automation with a more risk-based approach that also takes into account human in the loop. In summary, the objectives of the present study are articulated as follows:

1) To systematically review the application of GenAI and PredAI in the cybersecurity ecosystem, with a specific focus on cybersecurity auditing.
2) To identify opportunities and limitations in existing AI-driven cybersecurity approaches, while promoting a robust cybersecurity auditing culture.
3) To propose a governance-aligned, anti-sheriff, *artificial intelligence-driven cybersecurity audit model*.

The remainder of this paper is structured as follows: Section 2 presents the background to the study and the theoretical concepts underpinning the research. Section 3 outlines the research methodology adopted in this study. Section 4 presents the results of the systematic review, together with a detailed discussion and interpretation of the findings as the primary contribution. Section 4.6 proposes the Anti-Sheriff cybersecurity auditing model as the secondary contribution of the study. Finally, Section 5 concludes the paper by summarizing the key findings, contributions, limitations, and directions for future research.

2. Background

Authors use terminology differently depending on their area of focus. Terminology in the ICT literature is often subjective and may lack precision, necessitating a clear definition of the key terms used in this study. Hermann and Puntoni [14] indicate that definitions in literature are extensive and should be refined to align with the scope of the study. Terms such as cybersecurity, ICT, and IoT are not defined in detail in this section, as they are widely understood within the ICT community and are not central to the specific focus of this study.

2.1. Cybersecurity auditing

Owing to the evolution of modern technologies, cybersecurity auditing has become an integrated domain within the IS auditing landscape, ensuring that adopted systems and services comply with defined standards and that implemented cybersecurity controls are

adequate and effective [15], [16]. Cybersecurity auditing is a specialized area that has developed from traditional IS auditing. An IS audit assesses the integrity of systems, storage, processing, and the transmission of valuable information [15], ensuring that stored information remains intact and trustworthy. IS auditing may be defined as a formal, independent, and objective examination of an organization's information technology infrastructure to determine whether activities, including procedures and controls involved in collecting, processing, storing, distributing, and using information, comply with established guidelines [17].

The concept of cybersecurity auditing was introduced at a high level in the IS Audit and Control Association manual published in 2015 through the Certified IS Auditor framework, a recognized authority in IS auditing. The concept was more fully integrated into auditing practice in the Certified IS Auditor review manual, 26th edition (2022). While rooted in traditional IS auditing, this demonstrates that cybersecurity auditing has only recently emerged as a distinct audit domain. Cybersecurity auditing is a specialized process that assesses information technology infrastructure against an organization's security policies, controls, governance structures, and compliance with defined internal and international standards.

2. 2. Predictive artificial intelligence

PredAI involves using mathematical modeling, statistical analysis, and ML to identify patterns, anticipate behavior, and forecast future events based on historical data. Božić [14] explains that PredAI is a process-oriented concept that applies advanced mathematical methods to identify patterns, trends, and relationships in structured and unstructured datasets to support informed decision-making and prediction. PredAI presents a promising approach in cybersecurity for detecting advanced threats and supporting behavioral analytics in insider threat detection [18]. It is also applied in financial analysis to predict profitability, liquidity, and operational efficiency [19].

2. 3. Generative artificial intelligence

The concept of GenAI can be traced to earlier developments in AI research, including foundational statistical modeling work associated with scientists such as Gauss in the mid-twentieth century [20]. The practical significance of GenAI increased following the release of GPT-1 in 2018 [20]. Wessel et al. [21] indicate that recent developments in the

AI ecosystem, particularly generative AI, represent a significant shift in digital platforms and in how information exchange is understood. Unlike conventional AI approaches, including PredAI, which rely on pattern recognition and prediction, GenAI can learn from examples and produce novel outputs. As the term suggests, GenAI refers to a class of systems designed to create original content, including text, images, audio, and code [22].

The use of GenAI in cybersecurity is evident in areas such as the detection of complex attacks, including zero-day attacks. Some researchers, including Metta et al. [23], indicate that the term GenAI is sometimes applied loosely and may function as a marketing label rather than a clearly defined technical concept. Having established the distinction between GenAI and PredAI, this study presents a comparative table that contrasts these approaches and clarifies their respective applications within the cybersecurity landscape. As reflected in Table 1, there is a greater appreciation of AI within the cybersecurity landscape. A broader description of Table 1 is presented after the table.

Table 1. Comparative applications of GenAI and PredAI in cybersecurity

Proportion	GenAIb(GenAI)	PredAI (PredAI)	Cybersecurity applications
Primary purpose	The primary objective of GenAI models and capabilities is to generate new content, artifacts, or representations based on learned patterns and produce sound outputs.	These models are predominantly used to predict future events or outcomes based on historical and real-time data.	In a quest to combat modern cyber threats and adaptive cybersecurity attacks, GenAI supports adaptive defense artifacts, while PredAI enables proactive risk anticipation. They both show promise in enhancing cybersecurity controls.
Main function	Content synthesis and creation (e.g., text, code, logs, attack scenarios).	Forecasting, classification, and probability estimation.	Enables both creative threat simulation in real life and forward-looking security analytics, providing a true reflection of threats in near real time.
Learning approach	Typically, unsupervised or self-supervised	Primarily supervised and	Determines suitability for exploratory vs deterministic security use cases. These

Proportion	GenAIb(GenAI)	PredAI (PredAI)	Cybersecurity applications
	learning is used with large-scale datasets.	semi-supervised learning	models are also critical to the cybersecurity strategy, as they can identify high-risk areas that need to be prioritized.
Anticipated Outcome	Non-deterministic and probabilistic; outputs may vary for identical inputs.	Deterministic or statistically bound predictions.	These models are essential to expand the scope, efficiency, and accuracy. They enhance auditability, explainability, and assurance levels.
Augmented and explainability	Due to their unpredictable nature, their explainability capabilities require additional governance and interpretability layers.	Higher explainability due to structured features and labels.	Critical for regulatory compliance and audit traceability. In cybersecurity auditing, the explainability of adopted models is crucial to ensuring stakeholder buy-in for the use of AI in the audit.
Data source	Learn from broad, often unstructured datasets.	Rely on structured, labeled historical data.	Recently, there has been a need for data-driven cybersecurity controls and audits. These models influence data governance and quality requirements.
Compatibility and adaptability	Highly adaptive and capable of novel scenario generation	Limited to patterns present in training data	GenAI enhances resilience testing; PredAI improves operational stability within the cybersecurity controls implementations and enhancements.
Potential associated risks	Susceptible to hallucinations, prompt injections, and misuse	Vulnerable to bias, concept drift, and data poisoning	Requires differentiated AI risk controls and governance. While these models are key for cybersecurity defense mechanisms, the same approach can be used by

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Proportion	GenAIb(GenAI)	PredAI (PredAI)	Cybersecurity applications
			attackers to initiate more complex attacks.
Governance complexity	High governance overhead due to unpredictability and ethical concerns.	Moderate governance complexity with clearer validation paths.	While these approaches can align with AI governance maturity models and frameworks, such as the National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF), there is a need for a narrower set of governance measures within the cybersecurity fraternity.
Role in Cybersecurity Operations	Augments human analysts and simulates adversarial behavior	Automates the detection and prediction of threats	Supports both defensive creativity and operational foresight. In addition, these models are key to enhancing sound cybersecurity audits for operational and compliance purposes.

Table 1 highlights key distinctions and complementary roles between GenAI and PredAI within cybersecurity. GenAI primarily focuses on content generation and simulation using large, often unstructured datasets, enabling adaptive threat modeling and creative defense mechanisms, although it introduces challenges related to explainability, governance complexity, and risks such as hallucinations. While access to the training dataset remains a serious challenge [24], PredAI relies on structured, labeled data to perform forecasting, classification, and risk prediction, offering more deterministic and explainable outputs suitable for operational stability and decision-making. GenAI enhances resilience through scenario generation and adversarial simulation, while PredAI strengthens proactive risk anticipation and detection capabilities. Together, these approaches provide a balanced and integrated cybersecurity strategy, improving auditability, intelligence-driven decision-making, and overall assurance, while also

necessitating robust governance frameworks to manage their distinct risks and ensure regulatory compliance.

The foundational background and key terminologies employed throughout this study have been established. A clear distinction between GenAI and PredAI, together with their respective applications within the cybersecurity ecosystem, has also been defined. The subsequent section presents the research methodology adopted to systematically review the existing literature and, thereafter, to propose a conceptual model aimed at addressing the identified shortcomings. This methodological foundation supports the development and presentation of the proposed Anti-Sheriff cybersecurity auditing framework as the second component of the study.

3. METHODS

As highlighted by Rananga and Venter [24], to ensure that a systematic review is sound, transparent, and reproducible, a clearly defined research methodology should be presented. Clearly defining the research methodology also guards against the temptation of deviating from the main objective of the study [25]. Furthermore, digital academic databases contain a vast amount of information; therefore, it is essential to design a structured strategy that ensures the study objectives are achieved scientifically and rigorously. A detailed description of the methodology adopted in the present study is provided in the following subsections, beginning with the research workflow.

3. 1. Research Design and Workflow

The present study adopts a twofold research methodology that integrates a systematic literature review through PRISMA and proposes a scientific solution through the DSR approach. The objective of PRISMA is to systematically and comprehensively review the literature to identify key shortcomings, while the DSR approach translates these identified gaps into the development of the proposed Anti-Sheriff cybersecurity auditing framework, ensuring that it is grounded in empirical evidence and supported by methodological rigor and reproducibility. The PRISM approach has proved to be a useful tool to provide a checklist and protocol for selecting academic studies from a wide range of available studies [26]. On the other hand, DSR is a rigorous scientific methodology widely adopted for the development and evaluation of innovative artefacts aimed at

addressing complex, real-world problems [27]. More importantly, the authors observe that many systematic review studies focus primarily on identifying gaps in the literature and seldom propose foundational models or approaches to address these gaps. In other words, in the context of increasingly complex cybersecurity threats, it is essential not only to identify limitations in existing approaches but also to simultaneously propose feasible and structured solutions, hence the adoption of two-fold in the present study.

To explicitly articulate the relationship between the PRISMA and DSR methodologies adopted in the present study, a structured research workflow is developed to demonstrate the transition from systematic literature synthesis to framework design as follows:

1) Literature identification through database search.
2) The PRISMA is expanded through screening and eligibility assessment based on predefined criteria.
3) Data extraction of key study attributes
4) Thematic coding and synthesis from selected studies.
5) Gap identification across core dimensions
6) Mapping identified gaps to framework design requirements.
7) Development of the Anti-Sheriff framework using DSR.

The methodological workflow follows a sequential and traceable process, beginning with literature identification through database searches, followed by PRISMA-based screening and eligibility filtering. Relevant data are then extracted and subjected to thematic coding and synthesis to identify key gaps across core dimensions. These gaps are subsequently mapped to design requirements, leading to the development and conceptual evaluation of the Anti-Sheriff framework using the Design Science Research approach. The adoption of a twofold approach ensures that the proposed model is directly derived from systematically identified shortcomings in the literature, thereby enhancing transparency, traceability, and scientific rigor. The subsequent section presents a detailed description of the PRISMA methodology and its application within the present study.

3. 2. Systematic Literature Review (PRISMA)

The studies were selected based on the PRISMA guidelines. The PRISMA framework provides guidance on effective resource selection for conducting systematic literature

reviews and serves as a vital tool for identifying relevant academic papers across a wide range of available sources [28], [26], [29]. In the present study, the databases were chosen for their popularity, extensive article collections, and ease of access within the cybersecurity landscape. The libraries consulted include IEEE Xplore, ScienceDirect, SpringerLink, Web of Science, and the ACM Digital Library. The included databases do not represent an exhaustive list of all academic databases that could be used to perform a comprehensive systematic review. However, these are among the most renowned and widely used databases in the field. This approach is consistent with prior studies, such as those conducted by Kamruzzaman et al. [30] and Ofusori et al. [31], among many others. This selection was also influenced by the work of Valente et al. [32], who performed a systematic review to identify the most applicable academic databases for computer science research. Their findings [32] indicate that these databases are the most widely used within the computer science community. While there is no doubt that other academic databases that are not included can assist the review process, these databases are generally also adopted for AI systematic review, as was also the case in Rananga et al [29]. Another criterion adopted was the age of the papers at the time of the review. Only publications from these sources published within the last five years at the time of this study were considered for systematic review. It is, however, important to note that studies not meeting the selection criteria were still considered for contextual background and supporting citations, without being included in the final review sample. Table 2 summarizes the eligibility criteria, adapted from Vareta et al. [33].

Table 2. Article selection criteria

No	Criteria
1	Articles written in English and/or have a ready-to-use English version.
2	Articles published from 2021 to 2026.
3	Only peer-reviewed conference and journal articles are considered.
4	Articles with accessible full texts.

After defining the eligibility criteria for the selection of databases and papers for review, the adopted methodology is further expanded through the search strategy to ensure that only papers relevant to the study are considered.

3. 3. Search strategy

The strategy adopted by Muyambo et al. [26] emphasized that keyword search strategies are essential to ensure that the results retrieved from the databases are of high quality and thematically relevant to the study. The keywords and search strings were designed to capture the convergence of GenAI, PredAI, IS auditing, and specialized cybersecurity audits. This focused search strategy was deliberately designed to prioritize conceptual relevance over exhaustive breadth, as broader queries tend to retrieve a high volume of studies that address AI or cybersecurity independently without contributing to cybersecurity auditing. This trade-off aligns with established systematic review practices where domain-specific precision is preferred when the research objective targets a specialized intersection of fields. Furthermore, the inclusion of multiple synonymous AI and auditing terms mitigates the risk of excluding relevant studies while maintaining thematic alignment. This approach assists in effectively identifying relevant papers, without being tempted to include studies that may not directly contribute to the objectives of the study. The primary search string used to identify relevant studies is presented below:

(("Artificial Intelligence" OR "Generative Artificial Intelligence" OR "Predictive Artificial Intelligence" OR "Machine Learning" OR "Deep Learning")

AND

("Cybersecurity" OR "Information Security" OR "Network Security")

AND

("Audit" OR "Cybersecurity Audit" OR "Information Systems Audit" OR "Continuous Auditing" OR "Risk Assessment"))

As aforementioned, the search strategy adopted in this study follows the conventional use of Boolean “AND/OR” operators, commonly applied in database query languages. In a conventional database query setting, Boolean operators are logical connectors used to refine, broaden, or narrow search results [34]. This approach ensures comprehensive coverage of both technical AI-driven methodologies and their application within cybersecurity-focused IS auditing contexts, while minimizing irrelevant or non-domain-specific results. The structured use of Boolean logic also enhances the precision and relevance of retrieved studies across the selected academic databases. While the search strategy was essential for retrieving relevant studies from the selected databases, it is

also important to note that the authors did not indiscriminately include articles for review.

Since different databases use different structures for filtering the search, it is essential to clearly demonstrate how each search was designed for different databases, as shown in Table 3.

Table 3. Search Strategy Across Different Databases Used.

Database name	Search Query	Filter Applied
ScienceDirect	(("Artificial Intelligence") AND ("Cybersecurity" OR "Information Security" OR "Network Security") AND ("Audit" OR "Cybersecurity Audit" OR "Information Systems Audit" OR "Continuous Auditing" OR "Risk Assessment"))	Years: 2021–2026; Article type: Research articles; Publication title: Computer & Security, Future Generation Computer Systems, Journal of Information Security and Application; Access Type: Open access & Open Archive
SpringerLink	("Artificial Intelligence" OR "Generative Artificial Intelligence" OR "Predictive Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND ("Cybersecurity" OR "Information Security" OR "Network Security") AND ("Audit" OR "Cybersecurity Audit" OR "Information Systems Audit" OR "Continuous Auditing" OR "Risk Assessment")	Years: 2021–2026; Content type: Research article; Languages: English; Publishing model: Open access; Discipline: Computer Science
Web of Science	TS=(("Artificial Intelligence" OR "Generative Artificial Intelligence" OR "Predictive Artificial Intelligence" OR	Years: 2021–2026; Indexes: Web of Science Categories; Discipline: Computer Science Information Systems

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Database name	Search Query	Filter Applied
	"Machine Learning" OR "Deep Learning") AND ("Cybersecurity" OR "Information Security" OR "Network Security") AND ("Audit" OR "Cybersecurity Audit" OR "Information Systems Audit" OR "Continuous Auditing" OR "Risk Assessment"))	
ACM Digital Library	Abstract: (("Artificial Intelligence" OR "Generative Artificial Intelligence" OR "Predictive Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND ("Cybersecurity" OR "Information Security" OR "Network Security") AND ("Audit" OR "Cybersecurity Audit" OR "Information Systems Audit" OR "Continuous Auditing" OR "Risk Assessment"))	Years: 2021–2026; Content type: Journals & Conference
IEEE Xplore	(("All Metadata": "Artificial Intelligence" OR "All Metadata": "Generative Artificial Intelligence" OR "All Metadata": "Predictive Artificial Intelligence" OR "All Metadata": "Machine Learning" OR "All Metadata": "Deep Learning") AND ("All Metadata": "Cybersecurity" OR "All Metadata": "Information Security" OR "All Metadata": "Network Security") AND ("All	Years: 2021–2026; Content type: Journals & Conferences

Database name	Search Query	Filter Applied
	Metadata": "Audit" OR "All Metadata": "Cybersecurity Audit" OR "All Metadata": "Information Systems Audit" OR "All Metadata": "Continuous Auditing" OR "All Metadata": "Risk Assessment"))	

Now that the database and eligibility criteria have been defined, the next section provides the screening process. However, it is essential to note that before the formal screening process, a high-level assessment of each paper's structure, including the abstract and conclusion, was conducted to ensure its relevance to the study. In other words, as was the case in Rananga et al [29], authors need to apply a manual logical assessment before the actual screening to avoid a large number of irrelevant studies. A detailed screening process is subsequently discussed next.

3. 4. Screening process

After presenting the selection criteria and the search strategy adopted in the present study, the PRISMA flow diagram (Figure 1) shows the full flow and the logic of the selected papers. A literature review is a foundational step toward addressing modern challenges systematically [32]. As depicted in Figure 1, cybersecurity has attracted interest across sectors, including fintech and physical infrastructure. The rise in the adoption of modern technologies, such as IoT and mobile cloud computing, is evident in the literature [35]. Researchers explore opportunities presented by modern technologies to improve business operations and end-user experience.

In this study, only studies aligned with the defined research objectives were considered. As depicted in Figure 1, the initial search across five major academic databases yielded the following results: IEEE Xplore (n = 80), ScienceDirect (n = 113), SpringerLink (n = 84), Web of Science (n = 85), and the ACM Digital Library (n = 88). After removing duplicate records (n = 11), a total of 439 studies were subjected to the screening process. During screening, studies were excluded based on predefined eligibility criteria, including: (i) absence of a direct focus on Generative AI (GenAI) or Predictive AI (PredAI) (n = 113); (ii)

lack of application within the cybersecurity domain (n = 110); (iii) no clear relevance to the scope of cybersecurity auditing (n = 155); and (iv) studies adopting a literature review methodology (n = 55). Studies employing a literature review approach were excluded from the systematic synthesis because this study focuses on primary research contributions that present original findings, models, or empirical results. However, these review-based studies were retained for background and contextual framing to ensure comprehensive coverage of the research domain.

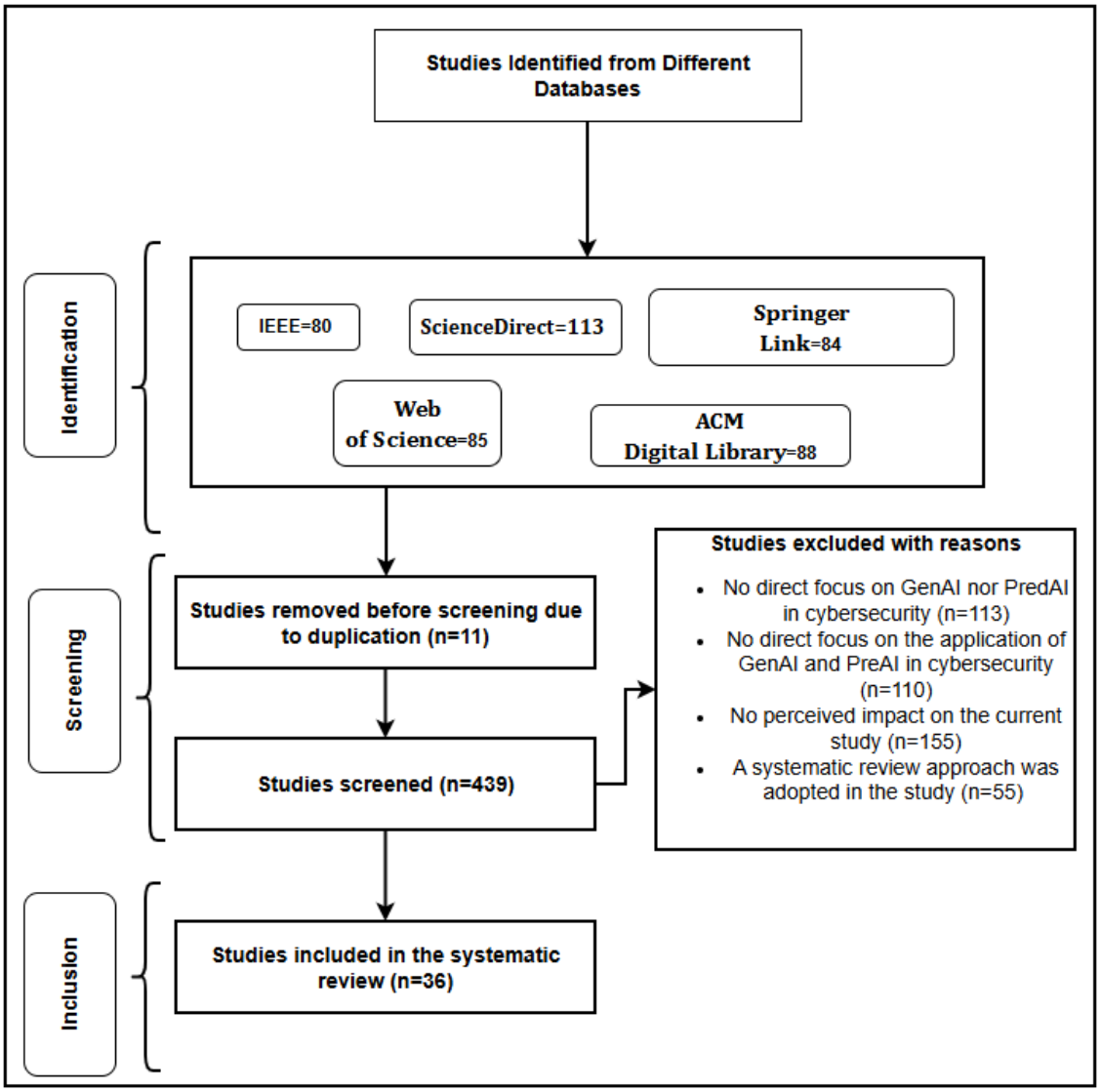

Figure 1. A PRISMA Flow diagram was adopted to select the studies reviewed

The relatively high exclusion rate during screening highlights the fragmented nature of existing research, where AI and cybersecurity are often examined in isolation, with limited

integration within cybersecurity auditing contexts. The importance of filtering studies from a large database was also indicated by Bolbot et al. [36] and Pirbhulal et al. [36]. After applying the filtering criteria, a final set of n = 36 studies was included in the systematic review. This refined corpus represents the most relevant body of literature addressing the intersection of AI capabilities and cybersecurity auditing limitations. These studies form the analytical foundation for the subsequent gap analysis and the development of the proposed Anti-Sheriff cybersecurity auditing model through the DSR approach.

The outcomes of the reviewed studies are subsequently analyzed through a methodological gap analysis, as this approach is discussed in the following section.

3. 5. Gap Analysis and Synthesis

Following the screening process, the selected studies were systematically analyzed to identify key themes, opportunities, and limitations in the application of AI within cybersecurity auditing. The analysis focused on critical dimensions, including predictive capabilities, governance challenges, explainability, data quality, and audit readiness. A thematic synthesis approach was adopted to extract recurring patterns across the reviewed studies, as such approaches are effective in identifying common shortcomings within complex research domains [24]. Through this process, the identified gaps were categorized into key structural limitations, particularly the fragmentation between (i) compliance verification, (ii) intelligence-driven risk assessment, and (iii) human governance within existing cybersecurity auditing practices. To ensure methodological traceability and alignment with the study objectives, each identified gap was explicitly mapped to a corresponding component of the proposed Anti-Sheriff framework. Specifically, gaps related to fragmented compliance verification informed the binary (compliance) layer, limitations in predictive and intelligence integration motivated the AI-enhanced risk layer, and governance and explainability deficiencies justified the inclusion of the human judgment layer.

This structured mapping establishes a direct linkage between the findings of the systematic review and the conceptual design of the proposed framework, thereby ensuring that the model is empirically grounded rather than purely conceptual. The

thematic synthesis was conducted through an iterative coding process. Initially, open coding was applied to extract key concepts from each study, focusing on AI capabilities, auditing functions, and identifying limitations. These codes were then grouped into higher-level themes through axial coding, enabling the identification of recurring patterns across studies. Finally, selective coding was used to align the themes with core dimensions of cybersecurity auditing, resulting in the categorization of gaps into compliance, intelligence, and governance layers. This structured coding approach enhances the transparency and reproducibility of the synthesis process.

3. 6. Design Science Research (DSR) for Conceptual Model Development

As indicated by Al-dhaqm et al. [37], a specific method should be adopted to develop a sound, systematic scientific solution while maintaining the scope of the study. Muyambo et al. [25] also contend that defining and adhering to a rigorous scientific method is essential to manage the scope of the study and its length requirements. As already alluded to, the present study adopted a two-fold approach; the second component of the present study employs the DSR. The DSR methodology represents a widely used scientific approach that supports the development and evaluation of innovative artifacts, including models, methods, and frameworks [38]. The DSR methodology is adopted in this study to systematically translate empirically identified gaps into a structured cybersecurity auditing framework. As depicted in Figure 2, the DSR methodology responds to the defined problem through technological enhancement, process improvement, and advances in scientific knowledge, achieved through the creation of innovative ideas that can be communicated to the broader academic community [39].

Figure 2 presents the DSR methodology that comprises six core stages: problem identification, solution objective definition, design and development, demonstration, evaluation, and communication. These sequential and iterative steps provide a structured approach for developing rigorous and practically relevant artifacts. These steps are essential, particularly in cybersecurity, owing to the complexity of threats associated with advancements in modern technologies such as AI. By progressing systematically through these stages, the problem identified earlier can be addressed through the design of a theoretically grounded and empirically validated solution. The DSR methodology ensures that the proposed solution is not only conceptually sound but also reproducible,

transparent, and capable of refinement through iterative evaluation. The steps of the DSR incorporated in the present study are elaborated next.

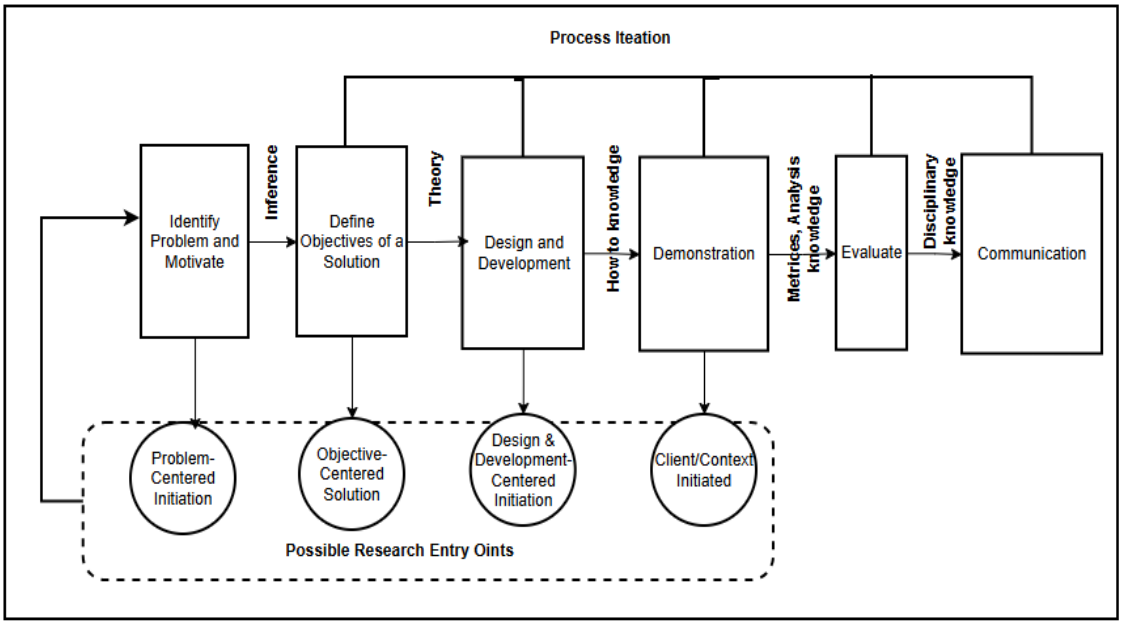

Figure 2. The Design Science Research methodology process model was applied in the present study(Adapted from Vom Brocke et al [39])

1) Problem Identification and Motivation

The systematic review revealed a persistent overreliance on binary compliance-based auditing, with limited integration of intelligence-driven risk assessment and governance interpretation, resulting in a false sense of assurance despite evolving threat landscapes. The authors have realized that if this gap is not addressed as a matter of urgency, the incorporation of AI within cybersecurity will continue to be regarded as a policing approach referred to herein as the "sheriff approach".

2) Define Objectives of the Solution:

Based on the identified gaps, the study defines the objective of developing a unified cybersecurity auditing framework that integrates binary control verification, AI-driven intelligence, and human governance. In essence, the objective of the second fold of the study is to methodologically propose a feasible approach in a quest to move from a cybersecurity sheriff approach to a more risk-based approach.

3) Design and Development:

The Anti-Sheriff framework is designed as a three-layer model consisting of (i) cybersecurity binary compliance verification, (ii) AI-enhanced risk assessment, and (iii) human judgment and governance interpretation. While existing approaches address isolated aspects of cybersecurity assurance, they lack an integrated structure that combines compliance verification, intelligence-driven risk assessment, and governance interpretation within a unified auditing model.

4) Demonstration:

The framework is conceptually demonstrated through its integration within the cybersecurity audit lifecycle, illustrating its ability to differentiate between compliant-but-high-risk and compliant-and-low-risk scenarios. Although empirical validation is reserved for future work, the conceptual evaluation provides analytical evidence of improved risk visibility compared to traditional binary approaches.

5) Evaluation:

The framework is evaluated through a structured conceptual analysis based on its ability to address identified gaps, improve audit interpretability, and enhance risk differentiation compared to traditional compliance-driven auditing models.

6) Communication:

The developed framework is presented as a contribution to the cybersecurity auditing body of knowledge. Following the structured methodology, the results derived from the systematic review are presented and critically discussed in the next section, with a focus on identifying patterns, limitations, and their implications for the proposed Anti-Sheriff framework.

4. RESULTS AND DISCUSSION

The contribution from the extensive body of available literature is duly acknowledged. A structured process for identifying relevant review papers has been clearly presented, leading to the systematic selection of applicable studies. Furthermore, the methodological approach adopted to develop the conceptual model is explicitly outlined.

The study is further strengthened through the incorporation and presentation of the PRISMA results.

4.1. PRISMA Results

To maintain analytical focus and methodological clarity, the detailed list of reviewed studies is provided in Annexure A, while the main body emphasizes the synthesis and interpretation of findings. This approach ensures that the integrity of the two-fold methodology (PRISMA-driven review and DSR-based development) is preserved, without disrupting the narrative flow through extensive listings of individual studies. The use of annexures to manage detailed study-level data has also been supported in prior research [29]. For transparency and ease of reference, the results of the reviewed studies are systematically presented in tabular format in Annexure A (**Error! Reference source not found.**). The table includes the reference for each study, along with a structured classification of whether the study explicitly addresses Generative AI (GenAI), Predictive AI (PredAI), or a hybrid approach. In addition, key opportunities in cybersecurity and shortcomings in cybersecurity auditing identified from each study are incorporated to support the subsequent thematic analysis.

To enhance interpretability, a binary marking scheme is adopted: a check mark (✓) indicates that a study explicitly addresses a given AI category, while a blank cell denotes its absence. For brevity and readability, the categories are abbreviated as G (GenAI), P (PredAI), and H (Hybrid) throughout the table. For example, the study by Mohammad et al. [19] is classified under PredAI (P), as indicated by a check mark, while the GenAI (G) and Hybrid (H) categories remain unmarked. A comprehensive representation of all included studies is provided in Annexure A.

Having established the corpus of included studies, the analysis is further extended through a quantitative assessment of AI adoption patterns within the cybersecurity domain. As illustrated in Table 3, the distribution of GenAI, PredAI, and hybrid approaches is analysed to identify dominant trends and to examine the underlying factors contributing to their prevalence. A detailed interpretation of these findings is provided immediately following the figure.

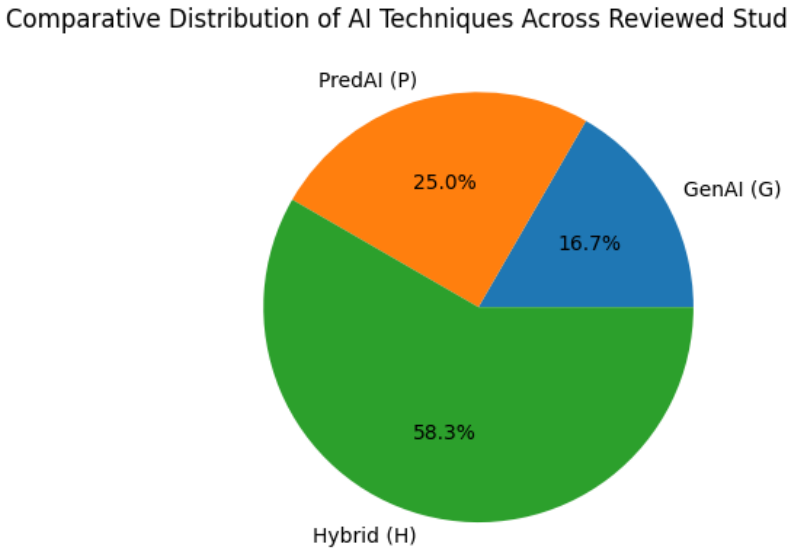

Figure 3. Comparative distribution of artificial intelligence techniques

The distribution illustrated in Figure 3 indicates a clear dominance of hybrid approaches (58.3%), followed by Generative AI (GenAI) (25%) and Predictive AI (PredAI) (16.7%). This distribution reflects a strong research preference for integrated, human-in-the-loop AI models, while also highlighting the comparatively limited emphasis on standalone predictive or generative techniques within cybersecurity auditing research. The prevalence of hybrid approaches suggests that researchers increasingly recognize the limitations of isolated AI capabilities and favour integrated models that combine predictive analytics with generative and contextual reasoning. A key contributing factor to this trend is the inherent complexity of the cybersecurity ecosystem, characterized by evolving and sophisticated threat landscapes. Prior studies, including Guntuka et al. [10], Rananga et al. [1], and Mutalib et al. [40], emphasize that modern cyber threats require adaptive and multi-dimensional analytical approaches. However, despite the dominance of hybrid models, a fundamental limitation persists existing approaches largely remain rooted in compliance-driven, enforcement-oriented paradigms. This reinforces a "policing" or "sheriff-style" approach, where the presence of controls is prioritized over the contextual interpretation of risk.

This observation highlights a critical research gap regarding the need to transition from rigid, enforcement-centric cybersecurity auditing toward more flexible, intelligence-driven, and risk-based assessment models. While hybrid AI approaches enhance analytical capability, they do not inherently resolve the governance and interpretability challenges

required for effective cybersecurity assurance. This gap provides the foundation for the proposed Anti-Sheriff framework, which seeks to integrate compliance verification, intelligence-driven risk assessment, and human governance into a unified auditing paradigm.

While no single AI capability can address all cybersecurity challenges associated with technological advancement, hybrid approaches are the most promising. Talukder et al. [40] and Suryotrisongko et al. [41] also contend that adopting a hybrid approach is the most feasible option. In addition to the emphasis on the adoption of flexible and hybrid approaches, several findings can be deduced from the reviewed literature. In the next subsection, the study is further expanded through a systematic extraction and analysis of themes related to the opportunities and shortcomings identified in the reviewed literature.

4. 2. Thematic Analysis

From a cybersecurity perspective, AI is predominantly applied in detective and preventative controls. Typical use cases include AI-enhanced anomaly detection, malware classification, fraud detection, and vulnerability exploitation forecasting. Incorporating AI into cybersecurity capabilities enhances early warning by identifying patterns that indicate a higher likelihood of cyber incidents. The integration of AI is largely influenced by complex, sophisticated cyber threats that can bypass traditional detection and prevention controls. A typical application of AI in security information and event management is security orchestration, automation, and response [38]. To systematically present the findings from the reviewed studies, a thematic distribution of key insights is adopted, focusing on both opportunities and shortcomings within AI-driven cybersecurity auditing. This thematic structuring enables a more coherent interpretation of how AI contributes to, and simultaneously challenges, existing auditing practices.

4.2. 1. Opportunity themes derived from reviewed studies

Advancements in modern technologies are associated with a wide range of opportunities, including automation and efficiency. Modern technologies, such as AI, have a positive effect on day-to-day human activities [1], [42]. Recently, researchers have explored

multiple ways to enhance cybersecurity using AI [43], rather than relying on conventional, rigid techniques. The adoption of AI in cybersecurity reflects the growing complexity of the ecosystem and the recognition that no single approach can address the diverse challenges faced by the cybersecurity community. As demonstrated in Table 4, different authors approach AI-driven opportunities for addressing cybersecurity challenges from varying perspectives; however, there is no fundamental disagreement regarding the overarching objective and opportunities. As emphasized by Ndaedzo and Venter [11], the adoption of modern technologies should occur alongside operational, organizational, and managerial considerations at a governance level. It is evident from the reviewed studies that aspects such as explainability cannot be separated from the adoption of AI within the cybersecurity ecosystem.

Table 4 demonstrates that AI enables enhanced predictive and contextual insights, which are particularly valuable in cybersecurity auditing. These capabilities support more targeted audit scoping by prioritising high-risk areas, thereby improving audit effectiveness while optimising resource utilisation. The detailed opportunity themes identified from the reviewed studies are discussed following the table.

Table 4. Opportunity themes derived from reviewed studies

Thematic grouping of opportunities	Description	Representative references
Predictive and contextual insight	Use of AI models to anticipate threats (for example, zero-day attacks, APTs), correlate contextual data, and provide forward-looking audit insights rather than retrospective findings.	[44], [45], [46], [47], [48], [49], [50], [51], [52], [53]
Automation and continuous monitoring (CCA/CCM)	Enables continuous controls auditing and monitoring through real-time data ingestion, adaptive detection, and reduced reliance on periodic manual audits.	[10], [8], [54], [55], [50], [2], [56], [57], [58]
Explainability and governance (XAI)	Focuses on auditability of AI models, transparency of decisions, bias	[59], [19], [60], [47], [61], [62], [63], [64]

Thematic grouping of opportunities	Description	Representative references
	detection, model accountability, and support for regulatory and stakeholder assurance.	
Security and privacy (Blockchain / DIDS)	Ensures integrity, immutability, and admissibility of audit evidence using blockchain, distributed learning, and privacy-preserving AI approaches.	[59], [8], [55], [19], [61], [65], [63]
Participatory and collaborative auditing	Emphasizes human-in-the-loop auditing, stakeholder engagement, and inclusion of non-experts in audit and oversight processes.	[13], [66], [10], [62], [65], [61]

As depicted in Table 4, the thematic grouping demonstrates that AI is transforming the broader cybersecurity landscape through predictive and contextual intelligence. The literature emphasizes the ability of AI models to anticipate emerging threats, such as zero-day vulnerabilities and Advanced Persistent Threats (APTs). AI as a feasible approach to combat zero days cybersecurity has been stressed by different authors, such as Rizvi et al [17] and Kim et al [67]. Another prominent theme in the literature is the growing adoption of continuous monitoring approaches, such as Continuous Control Auditing (CCA) and Continuous Control Monitoring (CCM), indicating a transition from periodic assessments to real-time assurance. The nature of modern cybersecurity threats is such that real-time monitoring should be adopted [53]. While a gap in AI governance remains evident [68], the literature increasingly highlights the importance of explainability, reflecting the need for transparency, accountability, and auditability of AI-driven decisions. The modern digitalization requires that the concept of confidentiality, integrity, and availability (CIA) be prioritized through sound security implementation. Furthermore, the security and privacy mechanisms, including blockchain and privacy-preserving AI, should be incorporated to enhance the integrity and reliability of audit evidence. Human in the loop is one of the influential factors regarding the adoption of AI within organizations [62]. Participatory and collaborative approaches further reinforce the continued importance of human judgment and stakeholder involvement. Collectively,

these themes illustrate a multi-dimensional evolution of the cybersecurity landscape. However, a critical observation emerges while the literature strongly emphasizes AI capabilities and associated opportunities, it does not proportionally address how these capabilities can be systematically integrated into existing cybersecurity auditing methodologies. This disconnect creates a gap between AI potential and audit readiness, thereby justifying the need for structured, integrative frameworks such as the proposed Anti-Sheriff cybersecurity auditing model. To further deepen the analysis, the relative significance of the identified opportunity themes is quantified through a thematic weighting approach. This enables the identification of dominant themes and provides additional insight into the distribution of research emphasis across the reviewed studies. The theme weights are calculated as mean percentages, as defined in Equation 1.

$$\text{Theme Weight } (\%) = \frac{\text{Number of references supporting the theme}}{\text{Total theme referenves}} \times 100 \quad (1)$$

Despite challenges, there is notable adoption of AI within the cybersecurity landscape. Figure 3 demonstrates that cybersecurity auditing can be significantly enhanced through the effective use of AI, including adapted automation for continuous monitoring, such as continuous CCM and CCA.

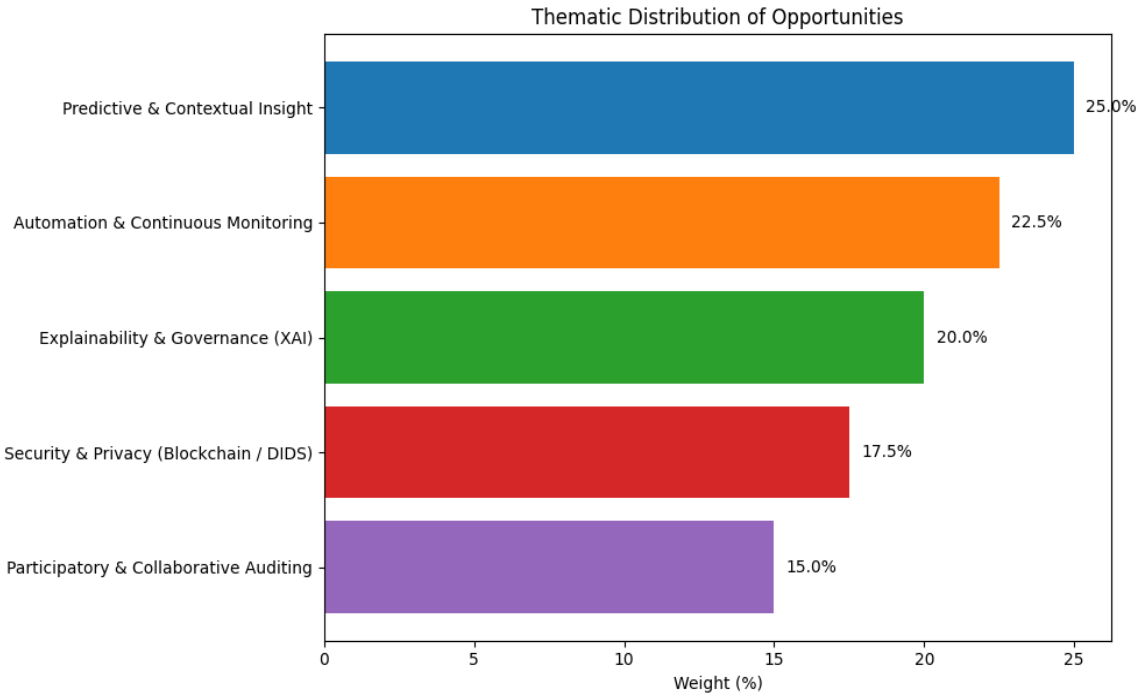

Figure 4 Opportunity themes derived from reviewed studies: Graphic representation

As depicted in

Figure **4**, the thematic analysis indicates that predictive and contextual insight (25.0%) is the most dominant opportunity, reflecting a strong emphasis on anticipatory and intelligence-driven cybersecurity approaches. This is key in the era of data-driven cybersecurity approaches [37]. The second dominant theme is automation and continuous monitoring (22.5%), indicating a shift toward continuous controls, auditing, and continuous monitoring to enable real-time assurance. Explainability and governance (XAI) account for 20.0%, indicating the importance of transparency, accountability, and trust in artificial intelligence-supported audit decisions. The issue of transparency and explainable AI has been emphasized in literature by studies such as Mutalib et al. [42]. From the themes presented in the figure, opportunities related to security and privacy, enabled by technologies such as blockchain and distributed learning, account for 17.5% and support the integrity and confidentiality of audit evidence. Participatory and collaborative auditing accounts for 15.0% and is emerging as an evolving research direction, highlighting the contributions of human-in-the-loop approaches and stakeholder engagement in modern cybersecurity auditing. Ndaedzo and Venter [1] indicate that technological advancements are associated with certain shortcomings. In the next section, the shortcomings that may hinder the adoption of GenAI and PredAI within the cybersecurity landscape are presented.

The dominance of predictive approaches and automation suggests that an enforcement-driven paradigm still underpins the application of AI in cybersecurity auditing. While these approaches enhance anticipatory capabilities, they often remain aligned with control verification rather than fully contextual risk interpretation. Collectively, these observations reinforce the Anti-Sheriff argument that existing auditing methodologies, rooted in binary compliance, are insufficient to capture the complexity of modern cybersecurity risks, thereby necessitating a more integrated model that combines intelligence-driven insights, continuous evaluation, and human-centric governance. Now that the opportunities of AI within the cybersecurity landscape have been presented, a thorough description of the identified gaps from the reviewed studies is presented next.

4.2. 2.Identified Gaps in Existing Approaches from the Literature

Artificial intelligence adoption in cybersecurity is not merely technical, but also institutional, requiring strong oversight structures and ethical safeguards [68], as

demonstrated by findings from the literature. The need to develop simpler, more adaptive AI algorithms is essential for dependable cybersecurity applications. Challenges identified in the literature confirm that the adoption of AI should be undertaken with due diligence. The shortcomings identified in the reviewed studies are presented in Table 5, with a detailed description following the table.

Table 5. Shortcoming themes derived from reviewed studies

Thematic grouping of Shortcomings	**Description**	**Representative references**
Algorithmic complexity (Black-Box Models)	Refers to the opacity and computational complexity of advanced AI models, which limit explainability, traceability, and an auditor's ability to justify findings. Black-box behavior undermines audit transparency, regulatory compliance, and stakeholder trust.	[46], [60], [19], [8], [69], [44], [47], [49], [51], [57], [56], [53], [52], [61]
Data quality and lineage issues	Highlights challenges related to biased, incomplete, synthetic, or poorly governed datasets, as well as weak data lineage and provenance. These issues directly affect the reliability, admissibility, and validity of AI-driven audit outcomes.	[70], [48], [19], [46], [45], [71], [2], [56], [50], [53]
Human Factors and Workforce Skill Gaps	Captures limitations arising from insufficient AI literacy, overreliance on automation, and limited human oversight. These gaps hinder effective interpretation, governance, and ethical use of AI outputs during cybersecurity audits.	[13], [66], [60], [9], [60], [62], [64], [50], [63]
Systemic latency and manual bottlenecks	Refers to delays introduced by manual data extraction, high computational demands, and resource-intensive processing, which constrain scalability and	[55], [19], [10], [54], [72], [61], [65]

Thematic grouping of Shortcomings	Description	Representative references
	undermine real-time or continuous auditing capabilities.	
Vulnerability to adversarial attacks	Addresses the susceptibility of AI models to adversarial manipulation, poisoning attacks, evasion techniques, and insider threats, which can compromise audit accuracy and model trustworthiness.	[59], [50], [52], [58]

As depicted in Table 5, the shortcomings of the reviewed studies highlight critical limitations that can hinder the effective adoption of AI in cybersecurity auditing. A dominant concern is algorithmic complexity, where black-box models reduce transparency, explainability, and the auditor's ability to justify findings, thereby weakening trust and regulatory compliance. The adoption of AI, if not done to cover the three core areas of technological advancement, which are people, process, and technology, runs into the risk of presenting more issues than solutions [1]. The data issue around cybersecurity and AI is a serious concern [24]. The other common theme from the reviewed studies is data quality and lineage issues, where biased, incomplete, or poorly governed data undermines the reliability and validity of AI-driven audit outcomes. Additionally, human factors and workforce skill gaps reveal challenges related to limited AI literacy, overreliance on automation, and insufficient human oversight, all of which affect governance and interpretation of results. Collectively, these shortcomings demonstrate that while AI introduces significant opportunities, its practical application in cybersecurity auditing remains constrained by technical, data, human, and security-related challenges.

Figure 5 illustrates the dominant shortcomings identified across the reviewed studies. Algorithmic complexity and black-box models emerge as the most significant limitation (31.8%), indicating persistent challenges related to explainability and audit defensibility. This is followed by data quality and lineage issues (22.7%), indicating the strong dependence of artificial intelligence-driven cybersecurity audits on reliable, traceable, and well-governed data. Human factors and workforce skill gaps account for 20.5%,

reflecting constraints in AI literacy, oversight, and governance capacity within organizations. Systemic latency and manual bottlenecks represent 15.9%, indicating structural limitations that hinder scalability and real-time or continuous auditing. Although vulnerability to adversarial attacks is the least cited shortcoming (9.1%), it remains a critical emerging risk as AI adoption in cybersecurity auditing continues to expand. Bharati et al. [5] indicate that, to respond to issues associated with AI adoption, governmental and commercial organizations should continuously release guidelines for sound adoption.

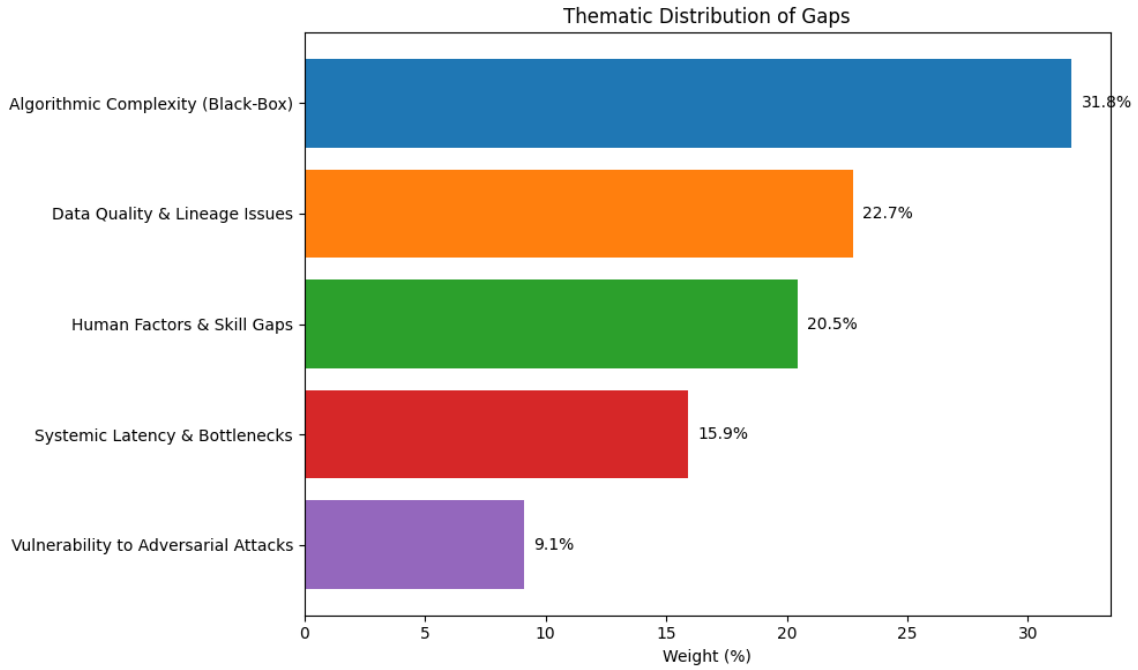

Figure 5. Shortcoming themes derived from reviewed studies: graphic representation

Despite the increasing availability of guidelines and best practices, the identified shortcomings remain fragmented and are not systematically integrated within existing cybersecurity auditing methodologies. This fragmentation highlights a critical need for a structured, risk-aware framework capable of coherently addressing these challenges, thereby enabling the effective, transparent, and accountable adoption of AI in cybersecurity auditing. To bridge the gap between the systematic review findings and the proposed conceptual model, the study is further extended through a comparative analysis of identified opportunities and shortcomings across key thematic dimensions.

4. 3. Comparative Analysis

Now that the opportunities and shortcomings from the systematic review have been presented in a compendium, the study is further expanded through a comparison of these opportunities and shortcomings across different themes. For example, as depicted in Figure 6, the literature indicates that, concerning predictive algorithms, there are more gaps than opportunities. A detailed description of the figure is provided following the graphical representation.

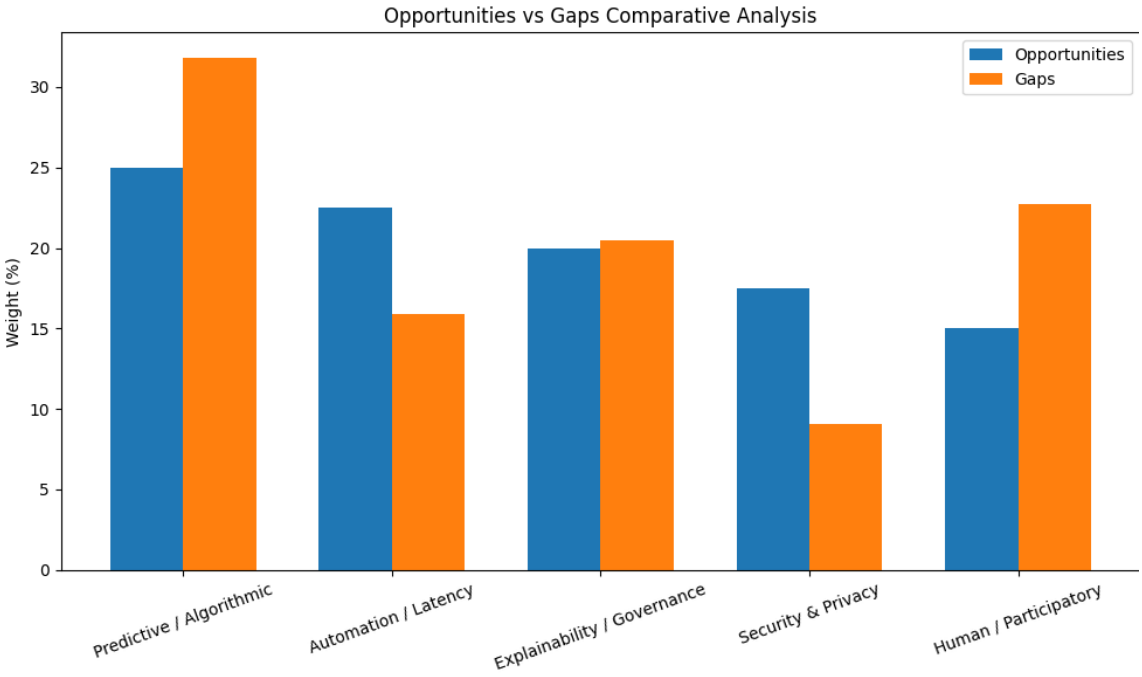

Figure 6. Opportunities versus Gaps Comparative analysis

While there is a clear appreciation of AI within the cybersecurity landscape [61], [51], the comparative analysis reveals a clear imbalance between the strengths and limitations of AI in cybersecurity auditing. Predictive and algorithmic capabilities show strong representation on both sides, with opportunities (25.0%) closely matched by gaps (31.8%), indicating that while AI excels in threat anticipation, it remains constrained by black-box limitations. Automation and continuous monitoring demonstrate a high opportunity (22.5%) but comparatively lower gaps (15.9%), suggesting relative maturity in this domain. This comes as no surprise, as most AI-driven models are designed to automate the prediction of potential cybersecurity threats [24]. Notably, explainability and governance remain nearly balanced (20.0% vs 20.5%), reinforcing the persistent challenge of making

AI outputs auditable and interpretable. On the other hand, security and privacy exhibit stronger opportunities (17.5%) than gaps (9.1%), indicating progress in integrity-preserving mechanisms such as blockchain. In contrast, human and participatory aspects reveal a significant gap (22.7%) compared to opportunities (15.0%), emphasizing the critical need for enhanced human oversight, accountability, and governance integration. Human-in-the-loop has proved to be a significant issue regarding the adoption of AI in some sensitive areas, such as health and legal practices [62]. Collectively, these findings demonstrate that while AI significantly enhances predictive capabilities and automation, it remains insufficient as a standalone auditing mechanism. The imbalance between technological advancement and governance readiness highlights the necessity of integrating structured human judgment within cybersecurity auditing processes. This directly supports the Anti-Sheriff premise that effective auditing must move beyond binary, compliance-driven approaches toward a more integrated model incorporating intelligence-driven analysis, continuous evaluation, and human-centric governance.

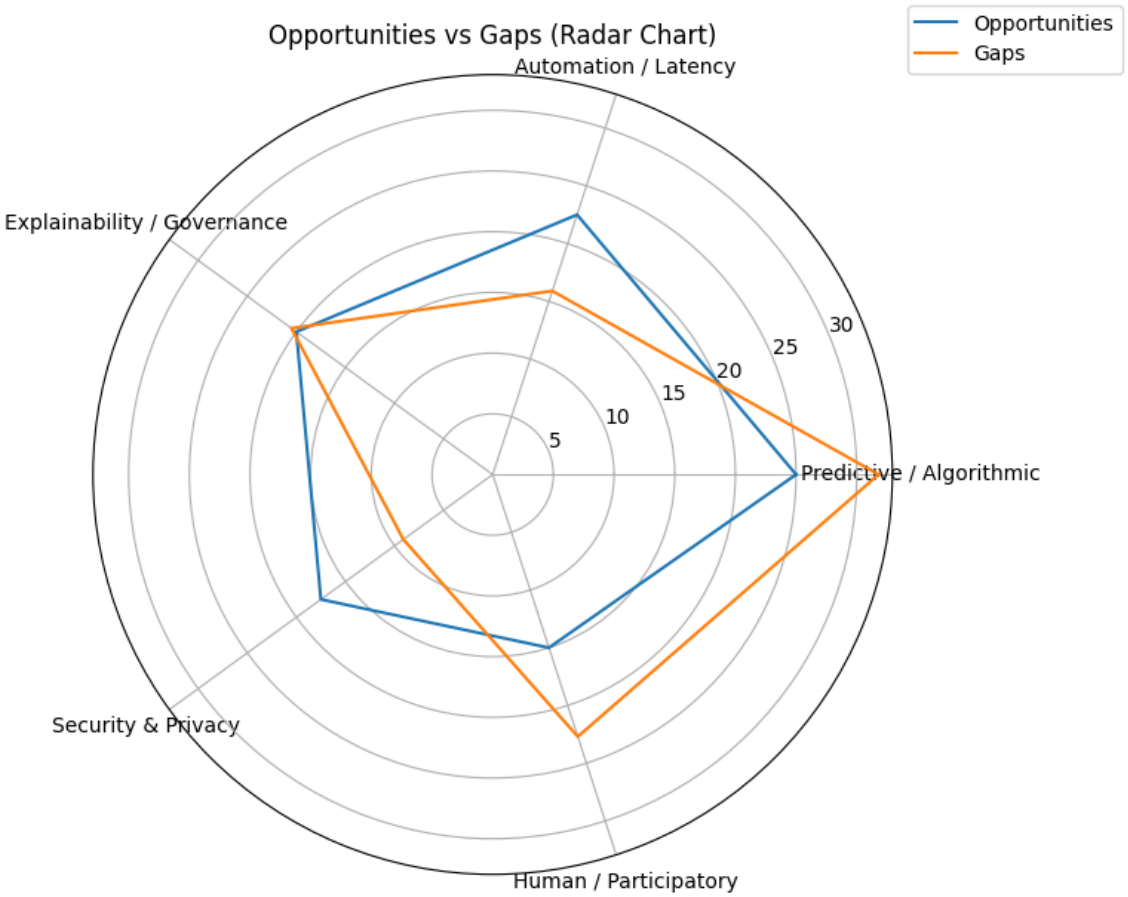

Figure 7. Radar Chart of Opportunities versus Gaps Comparative

Before presenting the proposed conceptual model aimed at addressing these imbalances, the analysis is further enriched through a radar-based visualization of opportunity themes, as shown in Figure 7. Using a radar chart in addition to a standard graph (e.g., Figure 6) is essential in this study because it provides a multi-dimensional perspective that cannot be fully captured by linear visualizations.

The digital ecosystem is becoming a highly complex and saturated environment, particularly due to the increasing dependence on third-party services [58]. As highlighted in Figure 7, there is a notable structural imbalance between AI-driven opportunities and auditing gaps across key thematic dimensions. Predictive and algorithmic capabilities exhibit the largest divergence, where high opportunity levels are overshadowed by even greater gaps, reinforcing concerns around black-box behavior. The black-box challenges are further influenced by the rapid adoption of AI as a response to modern technical challenges, which is often the case within Small and Medium Enterprises (SMEs) [1]. Automation demonstrates relatively stronger maturity, with opportunities exceeding gaps, suggesting practical readiness for continuous auditing. However, explainability and governance remain closely aligned, indicating that improvements in interpretability have not yet been translated into fully auditable systems. The question around "who audits the AI" is still a valid one. Security and privacy show a favorable balance, where opportunities outweigh gaps, reflecting progress in integrity-preserving mechanisms. While there is notable consideration regarding security within AI, the nature of cybersecurity threats is highly complex and requires continuous efforts to develop more robust approaches that can protect valuable assets from potential cyber threats [56]. In contrast, the human and participatory dimension reveals a pronounced deficit, emphasizing the lack of sufficient human oversight and governance integration. Overall, the radar chart visually reinforces the central argument that AI enhances technical capabilities but fails to adequately address auditability, thereby necessitating hybrid models such as the Anti-Sheriff framework.

Having presented the findings from the current literature and the associated opportunities, the study is further extended through a comprehensive discussion of research direction that serves as a foundational bridge to the second component of the research.

4. 4. Discussion

It is undeniable that modern technological advancements, particularly AI, have influenced virtually all domains [62]. The adoption of both Generative AI (GenAI) and Predictive AI (PredAI) is well established in the literature, with studies leveraging these approaches based on specific analytical objectives. Notably, the findings of the systematic review reveal a strong dominance of hybrid AI approaches, accounting for approximately 60% of the reviewed studies. This dominance can be attributed to the inherent complexity of cybersecurity environments, where isolated predictive models are often insufficient to capture dynamic and evolving threat patterns. As highlighted by Rananga and Venter [1], no single approach, whether AI-based or otherwise, can be considered a complete solution to the cybersecurity threats faced in the modern AI-driven era. It is for that reason that many researchers are exploring the combination of both GenAI and PredAI to combat cybersecurity threats. Hybrid models integrate multiple techniques, including machine learning, deep learning, and rule-based systems, enabling a more adaptive and context-aware approach to threat detection and risk assessment. These approaches are in order to reasonably protect valuable assets from stealth and sophisticated cybersecurity threats.

Despite the growing adoption of hybrid approaches, the literature also indicates the presence of standalone PredAI models. However, predictive approaches in isolation remain limited, as they primarily focus on forecasting risks without incorporating contextual interpretation, explainability, or alignment with compliance verification processes. Their effectiveness is further constrained by challenges related to data quality, completeness, and representativeness, which remain persistent issues in cybersecurity datasets [24]. Similarly, while GenAI demonstrates potential in areas such as automated reporting and simulation, its integration within cybersecurity auditing remains limited, particularly in relation to governance, control validation, and audit assurance processes.

A critical observation emerging from the reviewed studies is that the integration of GenAI and PredAI is predominantly oriented toward enhancing detection and prevention capabilities, rather than enabling risk-aware, governance-aligned cybersecurity auditing. As a result, existing approaches remain fragmented, lacking a unified structure that integrates compliance verification, intelligence-driven analysis, and human judgment. If

unaddressed, this fragmentation may reinforce the perception of cybersecurity auditing as a policing mechanism rather than a value-driven governance function.

This limitation directly motivates the development of the proposed Anti-Sheriff cybersecurity auditing framework, which seeks to bridge the divide between technical detection capabilities and strategic risk governance. The framework integrates three core components: binary control verification, AI-driven intelligence, and human governance. Unlike existing approaches that operate in isolation, the proposed model establishes a structured and unified cybersecurity auditing lifecycle, ensuring that the strengths of hybrid AI approaches are preserved while their limitations are mitigated through governance oversight and contextual interpretation.

A key methodological contribution of this study is the structural integration of PRISMA and DSR. Rather than being applied sequentially, these methodologies are interdependent. The systematic literature review (PRISMA) provides an empirical foundation for identifying gaps in AI-driven cybersecurity auditing, while DSR translates these gaps into formalized design requirements that guide the development of the proposed artefact. This ensures that the Anti-Sheriff framework is not conceptually abstract but systematically derived from evidence-based deficiencies identified in the literature.

4. 5. Linking Literature Findings to the Anti-sheriff Model

Table 6 operationalizes the linkage between PRISMA-derived gaps and the corresponding design requirements, demonstrating how each identified limitation theme is systematically translated into components of the conceptual Anti-Sheriff model. As already alluded to in the previous section, the shortcomings from the literature necessitate specific design responses, which are explicitly addressed within the proposed model. The reviewed literature reflected a dominant theme, which is algorithmic complexity, commonly referred to as the black-box problem. The black box problem aligns with the explainability that is still an issue within the adoption of AI, as also alluded to by Islam et al [68], Kumar et al [61], and Xu et al [73]. The black box algorithm limitation undermines audit transparency and necessitates explainability, traceability, and justification in cybersecurity auditing processes. In response to these themes, the Anti-Sheriff model integrates intelligence-driven insights augmented by human judgment,

ensuring that AI outputs remain interpretable, auditable, and aligned with governance expectations.

The other theme dominating the literature is data quality and lineage challenges, which highlight the limitations of relying solely on static or internally generated audit evidence. In essence, data quality and data scarcity are broader challenges in AI and its application in cybersecurity [24]. Many ML models for cybersecurity are still to be tested on a larger scale, due to the shortage of data and the quality thereof. From the systematic reviews, it is evident that there is a need for contextual and intelligence-driven inputs to assess risk in dynamic environments accurately. The proposed model addresses this by incorporating external intelligence sources, such as threat intelligence, EPSS, and OSINT, alongside validated audit evidence, thereby enhancing data reliability, contextual relevance, and audit accuracy. The proposed model makes use of what is already available to build a dataset.

There is no doubt that the other influential factor in regard to technological advancement is skill set and access to computational capabilities [11]. From the literature, the human factors and skill gaps theme further reinforces the necessity of governance-driven auditing. The reviewed studies consistently indicate that AI systems cannot independently ensure sound decision-making due to limitations in contextual understanding and ethical reasoning. This directly informs the inclusion of a dedicated Human Judgment (HJ) layer, which incorporates expert interpretation, governance oversight, and maturity-based assessment mechanisms to ensure balanced and accountable audit outcomes. The process of auditing is a complex, process-oriented exercise that is prone to various challenges, including the involvement of different stakeholders with diverse backgrounds. As presented in the model, the human-in-the-loop is essential for ensuring alignment with the business context. For example, what is considered ethical in cases such as cyberbullying in South Africa may require contextual interpretation, and this is where human judgment becomes critical within the proposed model.

Modern cybersecurity threats are surging at a rapid pace. Of particular concern is that it is becoming increasingly easy to initiate complex cybersecurity threats due to the availability of readily accessible tools. This exposure is likely to exhaust cybersecurity

monitoring capabilities and create bottlenecks. It was evident from the thematic challenges that systemic latency and operational bottlenecks are a serious concern. These challenges identified in the literature underscore the limitations of periodic and retrospective auditing practices. These inefficiencies necessitate a transition toward continuous and automated auditing processes. The Anti-Sheriff model addresses this through an AI-enhanced layer that enables continuous monitoring, automated data ingestion, and real-time risk evaluation, thereby supporting dynamic and adaptive assurance.

As already alluded to, AI is increasingly being adopted even in sensitive domains involving human lives. In such contexts, AI-assisted outcomes have little room for error. Vulnerabilities associated with adversarial attacks further highlight the risks of relying on unvalidated AI outputs. The literature emphasizes the need for resilient and adaptive auditing mechanisms. In response, the proposed model integrates continuous AI-driven monitoring with human oversight to detect anomalies, validate outputs, and ensure robustness against adversarial manipulation. To further expand on the link between the shortcomings from the literature, Table 6 is presented in detail as shown next.

Table 6. PRISMA to DSR Mapping

PRISMA Finding (Gap)	Design Requirement	Anti-Sheriff Model
Algorithmic Complexity (Black box)	Need for explainability, traceability, and audit justification	Intelligence-driven insights that explicitly integrate human judgment, rather than relying on AI to replace human decision-making.
Data Quality & Lineage Issues	Need for reliable, contextual, intelligence-driven inputs.	Integration of external intelligence sources (e.g., threat intelligence, EPSS, OSINT) combined with validated audit evidence to enhance data reliability, contextual relevance, and audit accuracy.
Human Factors & Skill Gaps	Need for human-in-the-loop governance.	Human Judgment (HJ) layer that incorporates expert interpretation, governance oversight, and maturity-

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

PRISMA Finding (Gap)	Design Requirement	Anti-Sheriff Model
		based assessment (e.g., CMM), ensuring balanced decision-making.
Systemic Latency & Bottlenecks	Need for automation and continuous auditing.	AI-Enhanced layer enabling continuous monitoring, automated data ingestion, and real-time risk evaluation to support continuous auditing practices.
Vulnerability to Adversarial Attacks	Need for automation and continuous auditing.	Combination of continuous monitoring (IAE) and human oversight (HJ) to detect anomalies, validate AI outputs, and ensure resilience against adversarial manipulation.

As depicted in Table 6, the PRISMA findings provide a systematic identification of key gaps in AI-driven cybersecurity auditing, which directly inform the design requirements of the study. These empirically derived shortcomings, ranging from algorithmic opacity and data reliability issues to human skill limitations and operational inefficiencies, are translated into structured design objectives through the DSR process. In this context, DSR functions as the bridging mechanism that transforms literature-driven gaps into a coherent artefact. This approach as also influenced by Muyambo et al [25], who, after performing a systematic review, also proposed a conceptual approach as a feasible response to the identified thematic shortcomings from the literature. Rananga et al. [1] also highlighted that the war between cybersecurity threats and defense is far from over, and there is a need for more conceptual approaches to address these modern challenges.

The proposed conceptual model operationalizes these requirements (as depicted in Table 5) through layered architecture comprising Binary Controls (baseline assurance), AI-Enhanced Intelligence (contextual and continuous risk evaluation), and Human Judgment (governance and interpretability). Importantly, this three-layer structure is not arbitrarily defined but reflects the minimum set of complementary dimensions required to address the fragmentation identified in existing approaches. The binary layer preserves the

foundational role of compliance verification, the AI-enhanced layer introduces intelligence-driven risk evaluation, and the human judgment layer ensures governance, explainability, and accountability. Any omission of these dimensions would perpetuate the same structural deficiencies observed in the literature and continue to render cybersecurity auditing as a "sheriff" exercise.

Through this structured integration from PRISMA results to DSR design, the Anti-Sheriff model emerges as a direct and traceable outcome of the systematic review. By aligning empirical findings with design requirements and translating them into a unified framework, the model provides a coherent, methodologically grounded, and practically relevant approach to cybersecurity auditing. This integration supports a transition from checklist-driven compliance toward intelligence-driven, risk-aware, and governance-oriented cybersecurity assurance. The details of the proposed model are presented next.

4.6. Model Derivation

Traditionally, the motivation for a study is presented at the beginning; however, because the present study is twofold (i.e., to provide a systematic review and to propose a model), the motivation arising from the first component (i.e., the systematic literature review) is restated. The second component of the present study (i.e., the model presentation) is motivated by the structural limitations identified in the systematic review, including traditional binary audit approaches and the fragmented nature of current artificial intelligence-driven cybersecurity research.

Although the adoption of AI in cybersecurity is evident from the reviewed literature, the evolving nature of cyber threats necessitates continuous improvement of existing approaches. Despite the progress made, emerging attack vectors, adaptive adversarial techniques, and increasing system complexity present persistent challenges. The next section presents the problem statement and the underlying motivation for the present study, narrowing the research focus toward the development of the proposed model.

1) Model motivation and methodology adopted for model development

Most existing studies on the incorporation of AI within cybersecurity emphasize enhanced threat detection, monitoring, and prevention; however, they seldom integrate compliance verification, intelligence-driven risk assessment, and human governance

within a unified assurance lifecycle. As established in the preceding discussion, existing AI approaches primarily enhance detection capabilities but do not transform cybersecurity auditing assurance structures. Alqahtani and Kumar et al. [74] indicate that most current applications of AI within cybersecurity focus on enhancement rather than process-level transformation. Existing studies predominantly focus on AI-driven threat detection, anomaly classification, fraud identification, and predictive intrusion modeling [6], [75]. While these contributions enhance operational cybersecurity capabilities, they do not transform the assurance paradigm underpinning cybersecurity auditing. Literature advances detection and prevention mechanisms but does not advance audit assurance architectures within cybersecurity ecosystems.

The persistent question surrounding modern AI, namely, "Who audits the auditing tools?", underscores the need to establish a balance between technical capability, explainability, and accountability within the AI ecosystem. If these gaps remain unresolved, cybersecurity auditing will continue to be perceived as a policing exercise centered on control enforcement rather than risk assessment aimed at business advancement. The incorporation of AI should not be perceived to intensify policing mechanisms or to catch stakeholders off guard. It should instead function as an enabler of balanced, risk-based assurance mechanisms that integrate operational, technical, and managerial dimensions within the broader governance framework of an organization.

The reviewed studies indicate a need to formally define the parameters required to facilitate the transition from a binary compliance exercise to a structured, risk-based approach within the cybersecurity landscape. The literature reflects ongoing efforts to respond to modern cybersecurity threats; however, current research lacks a formally structured model that integrates three essential audit dimensions into a unified lifecycle framework, such as:

1) Binary control verification (B) – Confirms the existence of controls and compliance with defined international standards, configuration baselines, or organizational policies.
2) Artificial intelligence-driven intelligence risk indicators (I) – Represents dynamic, data-informed risk signals derived from predictive analytics,

behavioral modeling, or external intelligence sources. This includes enhancing other resources, such as open-source intelligence, to develop risk indicators.

3) Human governance and judgment layer (H) – Involves expert interpretation, policy alignment, risk acceptance evaluation, and business contextualization.

The existing cybersecurity audit paradigm remains largely binary, with audit outcomes simplified as shown in Equation 2.

$$Audit_{Traditional}=f;B \in\{0,1\} \quad (2)$$

Where: B represents the presence or absence of control.

As depicted in Equation (2), the traditional binary exercise inherently assumes that control existence equates to adequate assurance; however, in complex, adaptive environments characterized by modern, sophisticated cyber threats, such an assumption is increasingly insufficient. Control existence does not imply control effectiveness under evolving exploit conditions, nor does it account for contextual intelligence or governance maturity.

Conversely, as demonstrated from the literature review, AI-focused (GenAI and PredAI) capabilities within the cybersecurity application typically operate within the detection and predictive domains. This can be denoted as:

$$Cybersecurity_{AI}=f(I) \quad (3)$$

Where: I represent predictive probabilities, anomaly scores, or model-derived threat likelihoods.

The theme that emerges strongly in the literature is the lack of a hybrid approach that unifies the binary checks (B), AI (I), and human intervention (H). This integration can be denoted as:

$$Audit_{Integrated}=f(B, I, H) \quad (4)$$

Where:

B = Binary compliance dimension

I = Intelligence-driven risk dimension

H = Human governance and interpretive dimension

The absence of such integration, as presented in Equation (4), results in a fragmented assurance model in which compliance checks remain static, AI outputs remain operationally siloed, and governance interpretation remains reactive rather than structurally embedded. The literature does not provide a formalized paradigm that transitions cybersecurity auditing from an enforcement-centric "Sheriff" model to an intelligence-augmented, governance-aligned assurance framework. The "Sheriff" approach is characterized by checklist-based validation and post-facto compliance confirmation, with limited integration of dynamic risk intelligence or strategic business alignment.

Although prior research, including the work of Mohawesh et al. [71], acknowledges the limitations of traditional auditing, no structured Anti-Sheriff paradigm has been academically conceptualized and articulated as a coherent model. Most models in the literature assume that control existence equates to adequate assurance, which can be misleading in the context of complex and sophisticated cyber threats. Control existence does not imply control effectiveness under evolving exploitation conditions, nor does it account for contextual intelligence or governance maturity relative to business objectives. In that regard, there is a need to present a more risk-centered cybersecurity audit model as presented next.

2) Conceptual Proposed Model

The need for a rethink within modern cybersecurity has been emphasized by several researchers, such as Mohawesh et al. [71], Dambe et al. [47], and Al-Hashimi et al [69]. Authors indicate that while there is growing appreciation of AI in the broader cybersecurity landscape, its integration for audit purposes, particularly in cybersecurity auditing, remains limited. As a result, there is a pressing need to propose a model that ensures the adoption of AI within cybersecurity auditing is undertaken with due diligence, appropriate governance, and accountability.

The proposed model appreciates GenAI and PredAI within the cybersecurity auditing community while ensuring that cybersecurity audits do not become tick-box exercises, referred to herein as the "sheriff approach". The sheriff approach is a binary check exercise that relies heavily on confirming the existence of cybersecurity controls and places limited emphasis on risk-aware processes that inform audit insights.

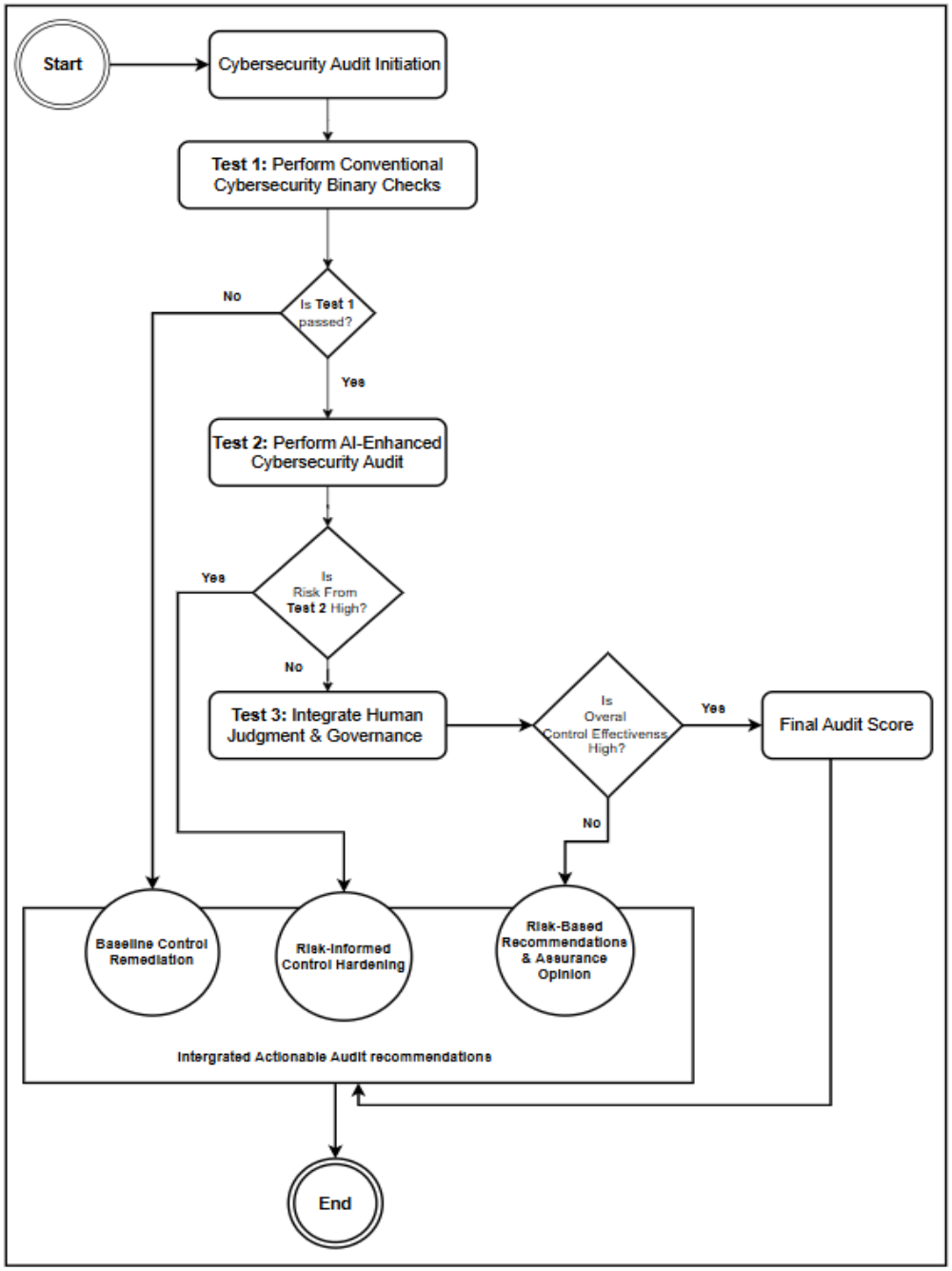

Figure 8. Conceptual high-level model representation (Anti-Sheriff artificial intelligence-driven cybersecurity model)

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

To counter this challenge, the study proposes a conceptual Anti-Sheriff cybersecurity approach. As depicted in Figure 8, the conceptual logic of the model shifts cybersecurity auditing from a tick-box exercise to an intelligence-driven, risk-based approach through the application of AI. The conceptual model does not replace traditional cybersecurity auditing practices. Rather, it augments and integrates conventional cybersecurity control assessments with a more dynamic, risk-based approach, enhanced by generative and PredAI. By embedding intelligent analytics into the audit lifecycle, the model strengthens proactive risk identification, contextual threat analysis, and adaptive decision-making. The explanation of the model components and their interactions follow the graphical representation.

The proposed model at this stage of the study functions as a conceptual framework rather than a detailed technical implementation. Muyambo et al. [25] and Ramazhamba et al. [76] indicate that, to demonstrate the relevance of novel approaches, researchers should first present a high-level representation of the model before elaborating on its finer details and operational aspects. The objective of the model, in its current state, is to illustrate the relationships among the key components required to support artificial intelligence-enhanced cybersecurity auditing.

In practice, the model identifies relevant stakeholders, defines the audit scope, determines key systems, identifies outdated systems to be excluded owing to stability issues, and maps processes within the audited environment. These elements are evaluated through risk-calibrated cybersecurity assessment mechanisms that consider governance structures, technological controls, and operational practices. In the context of the study, the model produces a structured evaluation of cybersecurity readiness and audit preparedness, enabling organizations to balance the adoption of artificial intelligence-driven technologies with the maintenance of adequate cybersecurity safeguards.

The literature assumes that the adoption of AI in the cybersecurity landscape is symbiotic with improvements in cybersecurity controls aimed at countering evolving threats. This assumption does not consistently hold within the cybersecurity auditing landscape. As the adoption of AI within cybersecurity auditing becomes more evident, this integration must be undertaken with due diligence to strengthen risk management rather than

focusing solely on detection and prevention. As illustrated in Figure 8, the model introduces a progressive, multi-layered cybersecurity auditing process that moves beyond traditional binary compliance verification toward a risk-calibrated, intelligence-supported audit approach that incorporates human judgment.

a) Cybersecurity audit initiation

As reflected in the background section, the process of conducting cybersecurity auditing is rooted in IS auditing processes. Rananga and Venter [1] indicate that the nature of modern cybersecurity threats requires continuous scrutiny, and implemented cyber controls should undergo continuous testing and improvement. The cybersecurity auditing process includes planning, execution, reporting, and debriefing.

The detailed activities within each stage are not central to the proposed model or the objective of the study; however, the audit process must be initiated, as shown in Figure 8. Following the initiation of the audit, the steps proposed in the conceptual model are applied sequentially within the cybersecurity audit process, as illustrated in the subsequent sections.

b) Conventional cybersecurity binary check exercise (Test 1)

Advancements in technology present several opportunities within the cybersecurity landscape. More complex cyber threats can, however, be anticipated [35]. To test the effectiveness of cybersecurity controls continuously, organizations scrutinize these controls using assessment approaches such as vulnerability assessments, firewall reviews, security configuration reviews, and penetration testing. Recently, cybersecurity auditing has been incorporated into conventional IS auditing to extend the assessment of cybersecurity controls.

While the cybersecurity audit process may be regarded as a feasible mechanism to ensure that implemented security controls remain effective and intact, a key concern is that such audits are predominantly conducted as binary compliance-check exercises. In this approach, controls are assessed as either "implemented" or "not implemented", with limited consideration of contextual risk, control maturity, or evolving threat dynamics. As depicted in Figure 8, the binary cybersecurity check approach is largely limited to

verifying the presence or absence of controls, which does not fully support a comprehensive and risk-informed cybersecurity audit.

The primary objective of the binary check exercise is to confirm the existence of a control for compliance with specific standards, internally defined ICT policies, or prescribed configuration guidelines, such as standard operating procedures. For example, when undertaking a binary check, an auditor verifies whether a security information and event management system is deployed within the environment. This exercise does not extend to determining whether the system is correctly configured to support critical business requirements or evolving threat conditions. Although in certain instances within the scope of penetration testing, the cybersecurity auditing process extends beyond confirmation to attempt exploitation of identified vulnerabilities, this approach remains largely technical. As a result, the process often fails to consider other influential factors, including information available about the entity from dark web sources, alignment of vulnerabilities with business requirements, the contribution of expert judgment, and residual risks associated with open-source intelligence.

Another practical example of a binary check exercise arises in firewall rule management. While a binary check may confirm that a firewall is implemented and active, it may fail to identify overly permissive "any–any" rules, unused legacy rules supporting decommissioned systems, or the absence of segmentation between critical business systems and less-trusted network zones. In such cases, the control formally exists and passes the audit, yet it introduces significant exposure to lateral movement and data exfiltration risks that threaten business operations.

Following this discussion, the binary check exercise is expanded to include relevant control types and testing procedures applicable within the process. Conventional auditing methodology employs operational verbs such as verify, compare, extract, review, test, discover, validate, analyze, assess, monitor, and inspect. These verbs articulate the procedures by which audit evidence is collected and evaluated, and guide the auditor in remaining aligned with auditing standards [77]. As the audit process is process-oriented, a Risk and Control Matrix document is often used to guide the audit process, as defined in the Certified IS Auditor manual [77].

As presented in Table 7, the selection of these verbs serves as guidance grounded in control assessment methodologies defined in established governance frameworks, including international standards such as NIST, CIS, Control Objectives for Information and Related Technologies (COBIT), and ISO. For example, when testing binary control (BC1), a cybersecurity auditor applies the NIST SP framework to assess access control by verifying that cybersecurity controls, such as access restrictions, are adequately implemented. The same principle applies to all other controls presented in Table 7.

Table 7. Conventional cybersecurity binary check exercise standards alignments

Control ID	Framework	Relevant controls	Example of test procedure approaches
BC1	NIST SP 800-53 Rev 5	Access control Configuration management) System logs and integrity Identification and authentication System audit and accountability	Verify: Confirm that cybersecurity controls such as access restrictions, authentication mechanisms, and logging are implemented. Compare: Review system configurations against NIST baseline configurations, internal standard operating procedures (SOPs), security baselines, and security policies. Extract: Retrieve database, system logs, configuration files, and user account data to evaluate compliance and detect anomalies.
BC2	ISO/IEC 27001:2022	Annex A controls, such as access control, logging, network security, and patch management.	Verify: Confirm the existence of documented approved security policies and implemented controls. Review: Examine evidence such as access logs, change management records, version controls, and patch deployment reports. Test: Evaluate control adequacy and effectiveness through sampling of system configurations and access permissions.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Control ID	Framework	Relevant controls	Example of test procedure approaches
BC3	CIS Critical Security Controls v8	Controls 1–8, such as inventory, access control, vulnerability management, and logging.	Discover: Identify and enumerate assets and software using manual or automated tools. Validate: Confirm secure configuration baselines and patch levels as per the defined procedures and processes. Analyze: Review vulnerability scan outputs and logging mechanisms to assess exposure and monitoring effectiveness.
BC4	NIST cybersecurity framework (CSF)	Identify Protect Detect	Assess: Evaluate asset management and risk identification processes. Test: Examine protective technologies such as access control, encryption, and network segmentation. Monitor: Review security monitoring capabilities, including intrusion detection and log analysis.
BC5	COBIT 2019	DSS05 (Manage security services), BAI10 (Configuration management).	Inspect: Evaluate governance structures and security service management processes. Review: Analyze configuration management procedures and change management records. Compare: Compare operational security practices with COBIT governance objectives.

Table 7 presents how binary control (BC) assessment within the proposed framework is systematically aligned with widely recognized cybersecurity standards and frameworks, including NIST, ISO/IEC 27001, CIS Critical Security Controls, and COBIT. Each control domain (BC1–BC5) is mapped to relevant control areas such as access control,

configuration management, logging, vulnerability management, and governance. This alignment ensures that the assessment of cybersecurity controls is not performed in isolation but is anchored in globally accepted best practices and regulatory expectations. In some instances, organizations might choose to use the internally developed policies, but in essence, the cybersecurity control testing should be grounded on sound standards or frameworks, as demonstrated in Table 7.

In addition, the table highlights the practical audit procedures used to evaluate these controls, including verification, comparison, extraction, review, testing, discovery, validation, analysis, monitoring, and inspection. These procedures emphasize an evidence-based approach to auditing, where system configurations, logs, policies, and operational processes are critically examined to assess both compliance and effectiveness. Collectively, this demonstrates that binary control assessment forms a structured and foundational layer of cybersecurity auditing, providing a baseline upon which more advanced intelligence-driven and governance-focused analyses can be built.

The importance of integrating intelligence within the broader cybersecurity landscape has been emphasized by several authors, including Rananga and Venter [1] and Guntuka and Shakshuki [10]. As indicated in the problem statement, there is little doubt that intelligence-driven cybersecurity represents the future of modern cybersecurity auditing. In the present study, the proposed cybersecurity model is enhanced using AI capabilities.

Each cybersecurity binary control entails important testing procedures, as presented in Table 7. The evolution of cybersecurity threats demonstrates that relying on a binary check-and-balance exercise alone can adversely affect business continuity and operational stability, as latent control weaknesses, including zero-day cyberattacks and lateral movement within networks, remain undetected. This may lead to security incidents that disrupt core business processes and affect revenue generation, even when the organization appears compliant from a traditional audit perspective. In the proposed model, the limitations of the binary check exercise are enhanced through the integration of GenAI and PredAI, as discussed next.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

c) Artificial intelligence-enhanced cybersecurity audit (Test 2)

In the proposed conceptual cybersecurity model, the authors incorporate AI as a second layer to support sound cybersecurity auditing outcomes. AI applications in cybersecurity include modern threat-hunting techniques enabled through behavioral analytics [53]. With IS auditing expanding through the integration of specialized audits such as cybersecurity audits, opportunities arise to integrate AI within these processes. GenAI and PredAI have significant potential to shift cybersecurity auditing from a conventional control checklist to an intelligent, risk-based assessment approach.

PredAI enhances cybersecurity auditing by analyzing historical security events, configuration data, and operational telemetry to predict the likelihood of control failure, misconfiguration, or exploitation. This approach moves cybersecurity auditing beyond pass-or-fail outcomes by assessing the probability and consequences of control weaknesses, providing a more realistic perspective on residual risk and business exposure.

GenAI capabilities enhance multiple aspects of cybersecurity [78]. Rather than relying on static checklists, Gena synthesizes large volumes of logs, policies, and configuration artifacts to generate meaningful audit insights. The literature, including Dambe et al. [78], identifies several opportunities, such as security orchestration, automation, and response, within AI and cybersecurity. These capabilities support the auditing process by enabling the assessment of cybersecurity control effectiveness, identifying undocumented risk patterns, and simulating plausible attack paths that may not yet have materialized.

AI also strengthens data-driven cybersecurity risk assessment, as demonstrated in Mohawesh et al. [71]. These insights are critical when organizations develop resilient cybersecurity strategies. This capability transforms the audit from a retrospective compliance exercise into a forward-looking assurance activity that evaluates how controls perform under evolving threat conditions. The integration of GenAI and PredAI establishes cybersecurity auditing as a continuous, intelligence-driven function that reflects technical control maturity and business-aligned risk, thereby resolving limitations associated with binary audit approaches. As indicated in Muhammad et al. [35], explainability and accountability remain essential when incorporating these technologies within cybersecurity auditing.

Building on the binary check exercise described earlier, the proposed conceptual model introduces artificial intelligence-enhanced procedures, as presented in Table 8. The testing procedures in Table 8 extend traditional auditing by augmenting standard control verification activities with artificial intelligence-driven analytical capabilities. While traditional binary checks rely on operational audit verbs such as verify, compare, extract, review, and test, the enhanced procedures introduce analytical verbs such as predict, correlate, forecast, prioritize, generate insights, and simulate. These verbs reflect the integration of advanced analytical techniques that enable the auditor not only to confirm the existence or effectiveness of cybersecurity controls but also to evaluate patterns, trends, and potential future risks within the audited environment.

In practice, the AI-enhanced procedures follow the same Risk and Control Matrix structure used in traditional auditing methodologies, as adopted in the Certified IS Auditor manual [77]. This guide is widely referenced within the IS auditing landscape. As illustrated in Table 8, each binary check is complemented by an artificial intelligence-enhanced check that expands the audit procedure from a static compliance validation exercise to an intelligent, risk-aware assurance process capable of identifying emerging threats, forecasting control failures, and prioritizing cybersecurity risks.

For instance, in BC1, the auditor verifies that access control mechanisms such as authentication and access restrictions are properly implemented. In the corresponding AIE1, the procedure is extended through artificial intelligence-driven techniques. Predictive analytics are applied to authentication logs and configuration changes to estimate the likelihood of unauthorized access. Anomaly detection algorithms identify abnormal behavior patterns in audit logs, and correlation analysis combines system telemetry, configuration states, and historical incidents to forecast potential control failures.

Table 8. AI-enhanced cybersecurity testing procedures aligned with the binary check from Test 1

Binary check control ID	AI-enhanced check control ID	Example of test procedure approaches
BC1	AIE1	Predict: Apply predictive analytics on authentication logs and configuration changes to estimate the

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Binary check control ID	AI-enhanced check control ID	Example of test procedure approaches
		likelihood of unauthorized access or misconfiguration. Detect: Use anomaly detection algorithms on audit logs to identify abnormal behavior patterns. Correlate: Combine system telemetry, configuration states, and historical incidents to forecast potential control failures.
BC2	AIE2	Analyze: Apply ML to log repositories to identify abnormal access patterns. Forecast: Use predictive models to estimate the probability that delayed patching will lead to exploitation. Generate Insights: Use AI and any other intelligence to summarize large log datasets and configuration artifacts into interpretable audit findings.
BC3	AIE3	Discover: Apply AI-assisted asset discovery to detect unmanaged or shadow IT assets. Prioritize: Use PredAI to rank vulnerabilities based on the likelihood of exploitation. Monitor: Apply behavioral analytics to detect deviations from baseline system activity.
BC4	AIE4	Identify: Use AI models to analyze threat intelligence feeds and open-source intelligence (OSINT) sources to identify emerging threats. Protect: Evaluate the effectiveness of security controls using predictive models trained on attack patterns. Detect: Apply anomaly detection to network traffic and endpoint telemetry to identify early indicators of compromise.
BC5	AIE5	Assess: Use AI-driven analytics to evaluate the effectiveness of security operations processes. Predict: Analyze configuration change history to identify patterns associated with system failures or vulnerabilities.

Binary check control ID	AI-enhanced check control ID	Example of test procedure approaches
		Simulate: Use AI-based attack simulation models to assess potential exploitation scenarios.

While AI enhancement strengthens cybersecurity technical controls, human-in-the-loop remains an important factor influencing the adoption of AI within the cybersecurity ecosystem [79]. The proposed conceptual model is further expanded through the integration of human judgment as a critical layer.

d) Human Judgment (governance and interpretation) (Test 3)

The main objective of conducting a cybersecurity audit is to scrutinize how cybersecurity controls respond to different cybersecurity threats. While the binary exercise and intelligence-driven risk indicators are essential, human judgment remains critical. Rananga and Venter [11] indicate that technical, operational, and managerial aspects are equally important. In the context of cybersecurity auditing, managerial aspects include expert judgment from the auditor and other stakeholders involved in the audit process.

The human judgment components in the proposed model include the assessment of policy enforcement, process maturity, risk acceptance, business alignment, decision-making, and the admissibility of cybersecurity outcomes. For example, a binary check may confirm that cybersecurity controls are implemented; the artificial intelligence-enhanced layer further scrutinizes these controls using intelligence from sources such as open-source intelligence and public domain repositories; and human judgment aligns the findings with the audited organization's context and business objectives. The final component of the proposed conceptual cybersecurity model is the audit outcome, as presented next.

Table 9 extends traditional binary control testing and artificial intelligence-enhanced procedures through the introduction of a human judgment layer that complements both binary check and artificial intelligence-enhanced procedures. The human judgment controls reflect the contribution of the cybersecurity auditor in interpreting and contextualizing technical and analytical findings. For example, in BC1, the auditor verifies that access control mechanisms are implemented, while AIE1 applies predictive analytics

and anomaly detection to identify potential unauthorized access or configuration weaknesses. The corresponding HJ1 step requires the auditor to interpret, evaluate, assess, decide, prioritize, recommend, contextualize, advise, govern, review, and guide the analytical outputs by determining whether the identified issues align with the organization's risk tolerance, regulatory obligations, and governance objectives.

This layer ensures that automated and analytical findings are translated into context-aware governance decisions, thereby integrating technical verification, artificial intelligence-driven analysis, and expert judgment into a comprehensive cybersecurity auditing process.

Table 9. Human Judgment Alignment with Binary Check from Test 1, and AI-enhanced from Test 2

Control ID	AI-enhanced check Control ID	Human judgment Control ID	Example of test procedure approaches
BC1	AIE1	HJ1	Interpret: Evaluate whether implemented controls align with organizational risk tolerance and regulatory obligations. Assess: Determine whether detected anomalies or vulnerabilities represent acceptable residual risk. Decide: Recommend whether to accept, mitigate, or escalate the risk based on governance priorities.
BC2	AIE2	HJ2	Evaluation: Assess the maturity of information security management processes. Review: Determine whether control implementation aligns with organizational policies and compliance requirements. Align: Interpret findings in relation to business objectives and risk appetite defined within the ISMS.
BC3	AIE3	HJ3	Prioritize: Determine which vulnerabilities or control weaknesses require immediate remediation based on business criticality. Validate:

Control ID	AI-enhanced check Control ID	Human judgment Control ID	Example of test procedure approaches
			Confirm whether automated findings are operationally relevant or false positives. Recommend: Provide remediation strategies aligned with operational capabilities.
BC4	AIE4	HJ14	Contextualize: Interpret technical findings within the organization's cybersecurity risk management strategy. Assess whether the detected risks threaten critical business services. Advise: Provide governance-level recommendations to strengthen the cybersecurity posture.
BC5	AIE5	HJ5	Govern: Evaluate whether cybersecurity processes align with enterprise governance structures. Review: Assess whether risk management decisions reflect organizational objectives. Guide: Recommend governance improvements and strategic cybersecurity investments.

Having presented the diverse testing procedures adopted in the proposed conceptual model, the final aspect of the audit process concerns the effective communication of audit outcomes. These outcomes must be conveyed in a manner that is understandable to the various stakeholders involved in the audit process, as discussed in the subsequent section.

e) Enhanced audit insights (audit outcome)

The main objective of conducting an audit exercise aligns with that of conventional financial statement auditing and specialized IS auditing, including cybersecurity auditing and data analytics. Audits are conducted to generate reliable insights into the state of business processes in relation to potential risks. This enables organizations to reduce risks to acceptable levels and minimize adverse effects on business operations should these risks materialize.

The rapid advancement of AI within information technology auditing, as indicated by Pycka and Zastempowski [44], has the potential to reshape audit practices toward a more risk-based approach. The auditing community increasingly recognizes the influence of AI and technological advancement on how audits are performed [80]. As shown in Figure 3, the proposed model uses AI to evaluate the effectiveness of cybersecurity controls and to enhance insights derived from the audit process.

These AI-enhanced insights support effective communication by ensuring that audit findings are accessible to both technical and non-technical stakeholders. For example, a cybersecurity auditor may conclude that a firewall effectively detects and prevents malicious network traffic. Non-technical stakeholders may not readily understand the business implications of this control. The final component of the conceptual model introduces a capability that accounts for both the intended audience and the communication of findings, ensuring the model effectively reaches relevant stakeholders.

Although several studies acknowledge the technical potential of AI, they largely retain a conventional orientation that verifies whether cybersecurity control exists rather than evaluating how effectively that control responds to evolving threat conditions or business-critical risks. The predominant focus remains on technical implementation, with comparatively limited emphasis on aligning AI outputs with business objectives, governance maturity, and intelligence-driven risk indicators, such as using empirical threat data to dynamically define audit scope. This study synthesizes the contributions and limitations identified in the literature to motivate the development of a model that balances technical cybersecurity controls with strategic business imperatives and broader human considerations. While the proposed model remains conceptual, its central aim is to advance cybersecurity auditing beyond a binary compliance exercise to a hybrid, adaptive, and intelligence-supported assurance approach that incorporates modern AI capabilities.

In the corporate environment, cybersecurity cannot be treated purely as a technical function; it must integrate with operational performance, managerial oversight, and value realization. Executive leadership and key business stakeholders increasingly require visibility into the return on investment associated with cybersecurity initiatives, including audit engagements. Such visibility becomes attainable when auditors are perceived not

as enforcement-oriented "sheriffs" but as independent, objective, and strategically aligned partners contributing to enterprise resilience and sustainable business objectives.

The proposed model establishes a foundation for a balanced and forward-looking cybersecurity auditing paradigm. Given the dynamic and complex nature of modern cyber threats, no single AI-driven mechanism can comprehensively address all dimensions of risk. Accordingly, while the model provides a structured and integrative approach, certain methodological and practical limitations remain. These limitations are acknowledged transparently and are discussed in the subsequent section as the study is concluded.

5. CONCLUSION

There is little doubt regarding the evolving nature of cybersecurity threats and the rapid adoption of AI, which has lowered the barrier to launching sophisticated cyberattacks. While cybersecurity auditing has increasingly been integrated into conventional information systems auditing to enhance resilience, a fundamental limitation persists: current practices remain largely rooted in binary control verification, referred to in this study as the “sheriff approach.” This approach often frames auditing as an enforcement-driven activity, limiting stakeholder engagement and reducing the effectiveness of risk-informed assurance. Through a PRISMA-based systematic literature review, this study identified two central patterns. First, the growing adoption of AI, particularly hybrid approaches, demonstrates the increasing importance of predictive and contextual intelligence in cybersecurity. Second, despite these advancements, there remains a critical gap in structured cybersecurity auditing models that integrate compliance verification, intelligence-driven risk assessment, and human governance within a unified framework. In response to this gap, this study employed DSR to propose the Anti-Sheriff cybersecurity auditing framework. The model conceptualizes cybersecurity auditing as an integrated lifecycle combining binary control verification, AI-enhanced intelligence, and human judgment. This structure directly addresses the fragmentation identified in existing approaches and repositions auditing toward a more risk-aware and governance-aligned assurance paradigm. However, the proposed framework is conceptual in nature and has not yet been empirically validated. As such, it should be interpreted as a foundational model that requires further testing and refinement. Future research should

focus on empirical validation, the development of quantitative evaluation mechanisms, and practical implementation across diverse organizational contexts.

REFERENCES

[1] N. Rananga and H. S. Venter, "Beyond the Hype : Readiness Assessment of SMEs for AI-Driven Cybersecurity Audits," *2025 Int. Conf. Intell. Comput. Next Gener. Networks*, pp. 1–6, 2025, doi: 10.1109/ICNGN67480.2025.11413704.

[2] A. Calvo, S. Escuder, N. Ortiz, J. Escrig, and M. Compastié, "RBD24 : A labelled dataset with risk activities using log application data," *Comput. Secur.*, vol. 150, no. December 2024, p. 104290, 2025, doi: 10.1016/j.cose.2024.104290.

[3] L. Guan, H. Shi, H. Chen, and Y. Wang, "Dynamic Scheduling for Security Protection Re-2 Sources in Cloud–Edge Collaboration Scenarios Using Deep Reinforcement Learning," *Mathematics*, vol. 13, no. 19, pp. 1–23, 2025, doi: 10.3390/math13193055.

[4] B. Faturoti, "Online learning during COVID19 and beyond: a human right based approach to internet access in Africa," *Int. Rev. Law, Comput. Technol.*, vol. 36, no. 1, pp. 68–90, 2022, doi: 10.1080/13600869.2022.2030027.

[5] S. Bharati, M. R. H. Mondal, and P. Podder, "A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When?," *IEEE Trans. Artif. Intell.*, vol. 5, no. 4, pp. 1429–1442, 2024, doi: 10.1109/TAI.2023.3266418.

[6] H. Hartono, "A Study on Competitiveness of ICT Adoption and Entrepreneurship Orientation on SMEs in Indonesia," *2019 Int. Conf. Inf. Manag. Technol.*, vol. 1, no. August, pp. 53–57, 2019.

[7] H. Mo and S. Ouyang, "(Generative) AI in Financial Economics (Generative) AI in Financial Economics," *J. Chinese Econ. Bus. Stud.*, vol. 23, no. 4, pp. 509–588, 2025, doi: 10.1080/14765284.2025.2569006.

[8] C. Rout, S. Sethi, J. C. Badajena, and R. K. Sahoo, "FSEGM: feature selection and ensemble generative model for adaptive cloud security," *J. Cloud Comput.*, vol. 15, no. 1, 2026, doi: 10.1186/s13677-025-00818-w.

[9] S. Biswal, "SCOUT: Surveillance and Cyber harassment Observation of Unseen Threats," *Proc. - 2024 Int. Conf. Artif. Intell. Metaverse Cybersecurity, ICAMAC 2024*, pp. 1–6, 2024, doi: 10.1109/ICAMAC62387.2024.10828792.

[10] S. Guntuka and E. Shakshuki, "Application of Generative Artificial Intelligence in Minimizing Cyber Attacks on Vehicular Networks," *Procedia Comput. Sci.*, vol. 251, pp. 140–149, 2024, doi: 10.1016/j.procs.2024.11.094.

[11] N. Rananga and H. S. Venter, "Mobile Cloud Computing Adoption Model as a Feasible Response to Countries' Lockdown as a Result of the COVID-19 Outbreak and beyond," *2020 IEEE Conf. e-Learning, e-Management e-Services, IC3e 2020*, no. Mcc, pp. 61–66, 2020, doi: 10.1109/IC3e50159.2020.9288402.

[12] S. Al-Eidi, O. Darwish, Y. Chen, M. Maabreh, and Y. Tashtoush, "A deep learning approach for detecting covert timing channel attacks using sequential data," *Cluster Comput.*, vol. 5, 2023, doi: 10.1007/s10586-023-04035-5.

[13] L. Morales-Navarro, M. Gan, E. Yu, L. Vogelstein, Y. B. Kafai, and D. Metaxa, "Learning AI Auditing: A Case Study of Teenagers Auditing a Generative AI Model," *Proc. ACM Human-Computer Interact.*, vol. 9, no. 7, pp. 0–29, 2025, doi: 10.1145/3757620.

[14] E. Hermann and S. Puntoni, "Artificial intelligence and consumer behavior: From predictive to generative AI," *J. Bus. Res.*, vol. 180, no. April, p. 114720, 2024, doi: 10.1016/j.jbusres.2024.114720.

[15] R. M. Skrynkovskyy, "An IT Audit as a Tool for Strategic Enterprise Management," *Probl. Ekon.*, vol. 1, no. 35, pp. 231–236, 2018.

[16] O. Ilori, C. I. Lawal, S. C. Friday, N. J. Isibor, and E. C. C.- Eke, "Cybersecurity Auditing in the Digital Age: A Review of Methodologies and Regulatory Implications," *J. Front. Multidiscip. Res.*, vol. 3, no. 1, pp. 174–187, 2022, doi: 10.54660/.ijfmr.2022.3.1.174-187.

[17] S. K. Jah Rizvi, K. F. Javed, and M. Moazam, "CAS - Attention based ISO/IEC 15408-2 Compliant Continuous Audit System for Insider Threat Detection," *3rd IEEE Int. Conf. Artif. Intell. ICAI 2023*, no. February, pp. 153–157, 2023, doi: 10.1109/ICAI58407.2023.10136657.

[18] R. Barton, P. W. C. Prasad, I. Seher, and A. Elchouemi, "Artificial Intelligence (AI) in Cybersecurity and Inhibitors to AI Adoption," *IEIR 2024 - Proc. 3rd Int. Conf. Intell. Educ. Intell. Res.*, pp. 1–10, 2024, doi: 10.1109/IEIR62538.2024.10959777.

[19] M. Nachouki, "AI-Augmented Financial Ratio Analysis for Early Warning , Monitoring , and Assurance," *2025 8th Int. Conf. Signal Process. Inf. Secur.*, pp. 1–5, 2025, doi: 10.1109/ICSPIS67605.2025.11318382.

[20] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy," *IEEE Access*, vol. 11, no. August, pp. 80218–80245, 2023, doi: 10.1109/ACCESS.2023.3300381.

[21] M. Wessel *et al.*, "Generative AI and its Transformative Value for Digital Platforms Generative AI and its Transformative Value for Digital Platforms," *J. Manag. Inf. Syst.*, vol. 42, no. 2, pp. 346–369, 2025, doi: 10.1080/07421222.2025.2487315.

[22] H. D. and E. G. Joe Devanny, *Generative AI and Intelligence Assessment*. 2023.

[23] S. Metta1 and & A. F. E. , Isaac Chang2, Jack Parker3, Michael P. Roman1, "Generative AI in Cybersecurity Shivani."

[24] M. Roopesh, N. Nishat, I. Arif, and A. E. Bajwa, "a Comprehensive Review of Machine Learning and Deep Learning Applications in Cybersecurity: an Interdisciplinary Approach," *Acad. J. Sci. Technol. Eng. Math. Educ.*, vol. 4, no. 04, pp. 37–53, 2024, doi: 10.69593/ajsteme.v4i04.118.

[25] D. F. Voting, M. Edmore, S. Baror, and S. Makura, "The Indonesian Journal of Computer Science," vol. 14, no. 6, pp. 10219–10237, 2025.

[26] J. Prümmer, T. van Steen, and B. van den Berg, "A systematic review of current cybersecurity training methods," *Comput. Secur.*, vol. 136, no. October 2023, p. 103585, 2024, doi: 10.1016/j.cose.2023.103585.

[27] L. S. Goecks, M. De Souza, T. P. Librelato, and L. R. Trento, "Design Science Research in practice: Review of applications in Industrial Engineering," *Gest. e Prod.*, vol. 28, no. 4, pp. 1–19, 2021, doi: 10.1590/1806-9649-2021v28e5811.

[28] R. Al Nafea and M. Amin Almaiah, "Cyber Security Threats in Cloud: Literature Review," *2021 Int. Conf. Inf. Technol. ICIT 2021 - Proc.*, pp. 779–786, 2021, doi: 10.1109/ICIT52682.2021.9491638.

[29] S. Rananga, B. Isong, A. Modupe, and V. Marivate, "Misinformation Detection: A Review for High and Low-Resource Languages," *J. Inf. Syst. Informatics*, vol. 6, no. 4, pp. 2892–2922, 2024, doi: 10.51519/journalisi.v6i4.931.

[30] M. Kamruzzaman, M. K. Bhuyan, R. Hasan, S. F. Farabi, S. I. Nilima, and M. A. Hossain, "Exploring the Landscape: A Systematic Review of Artificial Intelligence Techniques in Cybersecurity," *Proc. 2024 IEEE Int. Conf. Commun. Comput. Cybersecurity Informatics, CCCI 2024*, pp. 1–6, 2024, doi: 10.1109/CCCI61916.2024.10736474.

[31] L. Ofusori, T. Bokaba, and S. Mhlongo, "Artificial Intelligence in Cybersecurity: A Comprehensive Review and Future Direction," *Appl. Artif. Intell.*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2024.2439609.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[32] A. Valente, M. Holanda, A. M. Mariano, R. Furuta, and D. Da Silva, "Analysis of Academic Databases for Literature Review in the Computer Science Education Field," *Proc. - Front. Educ. Conf. FIE*, vol. 2022-Octob, pp. 1–7, 2022, doi: 10.1109/FIE56618.2022.9962393.

[33] P. Vareta, H. Muzenda, T. Nyamupaguma, and B. Ndlovu, "The Indonesian Journal of Computer Science," vol. 14, no. 6, pp. 10072–10102, 2025.

[34] F. Carcassi and G. Sbardolini, "Assertion, denial, and the evolution of Boolean operators," *Mind Lang.*, vol. 38, no. 5, pp. 1187–1207, 2023, doi: 10.1111/mila.12448.

[35] A. A. Khan *et al.*, "BDLT-IoMT—a novel architecture: SVM machine learning for robust and secure data processing in Internet of Medical Things with blockchain cybersecurity," *J. Supercomput.*, vol. 81, no. 1, pp. 1–22, 2025, doi: 10.1007/s11227-024-06782-7.

[36] V. Bolbot, K. Kulkarni, P. Brunou, O. V. Banda, and M. Musharraf, "Developments and research directions in maritime cybersecurity: A systematic literature review and bibliometric analysis," *Int. J. Crit. Infrastruct. Prot.*, vol. 39, no. October, 2022, doi: 10.1016/j.ijcip.2022.100571.

[37] A. Al-Dhaqm, S. A. Razak, R. A. Ikuesan, V. R. Kebande, and K. Siddique, "A review of mobile forensic investigation process models," *IEEE Access*, vol. 8, pp. 173359–173375, 2020, doi: 10.1109/ACCESS.2020.3014615.

[38] J. Venable, J. Pries-Heje, and R. Baskerville, "A comprehensive framework for evaluation in design science research," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7286 LNCS, no. 2012, pp. 423–438, 2012, doi: 10.1007/978-3-642-29863-9_31.

[39] J. vom Brocke, A. Hevner, and A. Maedche, "Introduction to Design Science Research," no. November, pp. 1–13, 2020, doi: 10.1007/978-3-030-46781-4_1.

[40] N. H. A. Mutalib, A. Q. M. Sabri, A. W. A. Wahab, E. R. M. F. Abdullah, and N. AlDahoul, "Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: a review," *Artif. Intell. Rev.*, vol. 57, no. 11, 2024, doi: 10.1007/s10462-024-10890-4.

[41] H. Suryotrisongko and Y. Musashi, "Evaluating hybrid quantum-classical deep learning for cybersecurity botnet DGA detection," *Procedia Comput. Sci.*, vol. 197, no. 2021, pp. 223–229, 2021, doi: 10.1016/j.procs.2021.12.135.

[42] H. Shi, "Network Security Situation Awareness Model and Empirical Re-search Based on Artificial Intelligence," *Proc. 2025 4th Int. Conf. Intell. Syst. Commun. Comput. Networks, ISCCN 2025*, no. February 2025, pp. 77–84, 2025, doi: 10.1145/3732945.3732956.

[43] T. Bishtawi and R. Alzubi, "Cyber Security of Mobile Applications Using Artificial Intelligence," *1st Int. Eng. Conf. Electr. Energy, Artif. Intell. EICEEAI 2022*, pp. 1–6, 2022, doi: 10.1109/EICEEAI56378.2022.10050484.

[44] M. Pycka and M. Zastempowski, "Machine learning and artificial intelligence techniques adopted for IT audit," *Management*, vol. 29, no. 1, pp. 65–87, 2025, doi: 10.58691/man/200768.

[45] M. J. Hossain, K. Alam, M. F. Monir, M. M. Hoque, and T. Ahmed, "Explainable AI Meets Synthetic Data: A Deep Learning Framework for Detecting Network Intrusion in NextG Network Infrastructure," *IEEE Access*, vol. 13, no. July, pp. 114979–115001, 2025, doi: 10.1109/ACCESS.2025.3585783.

[46] S. M. Darwish, A. I. Salama, A. A. Elzoghabi, and N. A. El-Shoafy, "An intelligent memetic approach to detect online fraud for distributed fintech environments," *Electron. Commer. Res.*, 2025, doi: 10.1007/s10660-025-10050-y.

[47] S. Dambe, S. Gochhait, and S. Ray, "The Role of Artificial Intelligence in Enhancing Cybersecurity and Internal Audit," *2023 3rd Int. Conf. Adv. Electron. Commun. Eng. AECE 2023*, pp. 88–93, 2023, doi: 10.1109/AECE59614.2023.10428353.

[48] M. Y. Alzahrani and A. M. Bamhdi, "Hybrid deep-learning model to detect botnet attacks over internet of things environments," *Soft Comput.*, vol. 26, no. 16, pp. 7721–7735, 2022, doi: 10.1007/s00500-022-06750-4.

[49] M. Kumari, M. Gaikwad, and S. A. Chavan, "A secure IoT-edge architecture with data-driven AI techniques for early detection of cyber threats in healthcare," *Discov. Internet Things*, vol. 5, no. 1, 2025, doi: 10.1007/s43926-025-00147-z.

[50] D. Chiba, H. Nakano, and T. Koide, "DomainDynamics: Advancing lifecycle-based risk assessment of domain names," *Comput. Secur.*, vol. 153, no. December 2024, p. 104366, 2025, doi: 10.1016/j.cose.2025.104366.

[51] T. Xin, Y. He, E. D. Zamani, M. Evans, and C. Luo, "A cyber risk economics model for organization-wide risk management (CYREM-ORM)," *Comput. Secur.*, vol. 165, no. July 2025, 2026, doi: 10.1016/j.cose.2026.104873.

[52] L. Yuan *et al.*, "FusionITD: enhanced cross-modal insider threat perception framework via behavior-semantic fusion," *Cybersecurity*, vol. 9, no. 1, 2026, doi: 10.1186/s42400-026-00555-w.

[53] A. Reddy and A. Pradesh, "Innovative Approaches to Dark Web Monitoring for Cybersecurity Intelligence," pp. 1–5, 2025.

[54] M. Shahin, F. F. Chen, A. Hosseinzadeh, H. Bouzary, and R. Rashidifar, "A deep hybrid learning model for detection of cyber attacks in industrial IoT devices," *Int. J. Adv. Manuf. Technol.*, vol. 123, no. 5–6, pp. 1973–1983, 2022, doi: 10.1007/s00170-022-10329-6.

[55] R. R. Irshad *et al.*, "Enhancing Cloud-Based Inventory Management: A Hybrid Blockchain Approach With Generative Adversarial Network and Elliptic Curve Diffie Helman Techniques," *IEEE Access*, vol. 12, no. January, pp. 25917–25932, 2024, doi: 10.1109/ACCESS.2024.3367445.

[56] S. Sathyakala and E. Anbalagan, "Comparative Analysis of Cyber Security Threat Detection Based on Artificial Intelligence Approaches," *2024 Asian Conf. Intell. Technol. ACOIT 2024*, pp. 1–8, 2024, doi: 10.1109/ACOIT62457.2024.10940025.

[57] Y. C. Tung, E. C. Liou, P. C. Hu, and C. H. Yu, "VWA-6G AI assisted continuous security monitoring over open RAN service management orchestration," *Comput. Secur.*, vol. 157, no. June, p. 104566, 2025, doi: 10.1016/j.cose.2025.104566.

[58] M. Repetto, "Cybersecurity Digital Twins: Concept, blueprint, and challenges for multi-ownership digital service chains," *J. Inf. Secur. Appl.*, vol. 96, no. November 2025, p. 104299, 2026, doi: 10.1016/j.jisa.2025.104299.

[59] W. Wei and L. Liu, "Trustworthy Distributed AI Systems: Robustness, Privacy, and Governance," *ACM Comput. Surv.*, vol. 57, no. 6, 2025, doi: 10.1145/3645102.

[60] A. E. Muhammad, K. C. Yow, N. Bačanin-Džakula, and M. A. Khan, *L-xaids: A LIME-based eXplainable AI framework for intrusion detection systems*, vol. 28, no. 10. 2025.

[61] R. Kumar, A. Aljuhani, D. Javeed, P. Kumar, S. Islam, and A. K. M. N. Islam, "Digital Twins-enabled Zero Touch Network: A smart contract and explainable AI integrated cybersecurity framework," *Futur. Gener. Comput. Syst.*, vol. 156, no. February, pp. 191–205, 2024, doi: 10.1016/j.future.2024.02.015.

[62] M. Malatji, "Augmented Intelligence Framework for Human–Artificial Intelligence Teaming in Cybersecurity," *Human-Centric Intell. Syst.*, vol. 5, no. 2, pp. 151–180, 2025, doi: 10.1007/s44230-025-00103-8.

[63] L. Alevizos, "Automated cybersecurity compliance and threat response using AI, blockchain and smart contracts," *Int. J. Inf. Technol.*, vol. 17, no. 2, pp. 767–781, 2025, doi: 10.1007/s41870-024-02324-9.

[64] W. Ibrar, D. Mahmood, A. S. Al-Shamayleh, G. Ahmed, S. Z. Alharthi, and A. Akhunzada, "Generative AI: a double-edged sword in the cyber threat landscape," *Artif. Intell. Rev.*, vol. 58, no. 9, 2025, doi: 10.1007/s10462-025-11285-9.

[65] N. Goel, "Federated Learning Framework for Risk diagnosis in Enterprise Information Security model," *2025 Tenth Int. Conf. Sci. Technol. Eng. Math.*, pp. 1–6, doi: 10.1109/ICONSTEM65670.2025.11374875.

[66] X. Qiu and S. Zhang, "Application analysis of generative artificial intelligence in basic education," no. December 2024, pp. 41–46, 2024, doi: 10.1145/3723420.3723428.

[67] J. Kim *et al.*, "Anomaly detection based on traffic monitoring for secure blockchain networking," *IEEE Int. Conf. Blockchain Cryptocurrency, ICBC 2021*, vol. 19, no. 3, pp. 3619–3632, 2021, doi: 10.1109/ICBC51069.2021.9461119.

[68] S. Islam, N. Basheer, S. Papastergiou, M. Ciampi, and S. Silvestri, "Intelligent dynamic cybersecurity risk management framework with explainability and interpretability of AI models for enhancing security and resilience of digital infrastructure," *J. Reliab. Intell. Environ.*, vol. 11, no. 3, pp. 1–25, 2025, doi: 10.1007/s40860-025-00253-3.

[69] H. A. Al-Hashimi, R. A. Khan, H. S. Alwageed, A. M. Algarni, S. Ayouni, and A. O. Almagrabi, "Exploring the role of generative AI in enhancing cybersecurity in software development life cycle," *Array*, vol. 28, no. March, p. 100509, 2025, doi: 10.1016/j.array.2025.100509.

[70] M. Chen, F. Liu, D. Liang, S. Zhong, and Y. Li, "Entity Recognition for Power Equipment Data Based on Optional Word Vectors and Feature Fusion," *IEEE Access*, vol. 13, no. June, pp. 143767–143780, 2025, doi: 10.1109/ACCESS.2025.3598316.

[71] R. Mohawesh, M. A. Ottom, and H. B. Salameh, "A data-driven risk assessment of cybersecurity challenges posed by generative AI," *Decis. Anal. J.*, vol. 15, no. January, p. 100580, 2025, doi: 10.1016/j.dajour.2025.100580.

[72] K. S. Kumavat and J. Gomes, "Multi-layer DDoS attacks detection with mitigation in IoT-enabled sensor network," *Cybersecurity*, vol. 8, no. 1, pp. 1–24, 2025, doi: 10.1186/s42400-025-00378-1.

[73] R. Xu, D. Liao, H. Meng, and Z. Feng, "Research on Application of Artificial Intelligence Technology in Industrial Information Security," *Proc. 2021 IEEE 3rd Int. Conf. Civ. Aviat. Saf. Inf. Technol. ICCASIT 2021*, pp. 683–686, 2021, doi: 10.1109/ICCASIT53235.2021.9633669.

[74] H. Alqahtani and G. Kumar, "A comprehensive review of generative AI techniques and their impact on cybersecurity," *Soft Comput.*, vol. 29, no. 13–14, pp. 4945–4982, 2025, doi: 10.1007/s00500-025-10702-z.

[75] V. Božić, "AI and Predictive Analytics," *J. Sport. Ind. Blockchain Technol.*, vol. 1, no. 1, p. 1, 2023.

[76] P. T. Ramazhamba and H. Venter, "Blockchain Forensics and Regulatory Technology for Crypto Tax Compliance: A State-of-the-Art Review and Emerging Directions in the South African Context," *Appl. Sci.*, vol. 16, no. 2, p. 799, 2026, doi: 10.3390/app16020799.

[77] B. P. David L. Cannon Timothy, Timothy S. Bergmann, *CISA Certified Information Systems Auditor™ Study Guide*, vol. 1, no. April. 2015.

[78] Z. L. Teo, C. W. N. Quek, J. L. Y. Wong, and D. S. W. Ting, "Cybersecurity in the generative artificial intelligence era," *Asia-Pacific J. Ophthalmol.* , vol. 13, no. 4, p. 100091, 2024, doi: 10.1016/j.apjo.2024.100091.

[79] R. Roberts and C. Mcdermott, "Integrating Human Factors into Insider Threat Detection – A Systematic Review Integrating Human Factors into Insider Threat Detection – A Systematic Review," no. February, pp. 0–37, 2026, doi: 10.1145/3798089.

[80] A. A. Vărzaru, "An Empirical Framework for Assessment of the Effects of Digital Technologies on Sustainability Accounting and Reporting in the European Union," *Electron.*, vol. 11, no. 22, 2022, doi: 10.3390/electronics11223812.

[81] S. Das Guptta, K. T. Shahriar, H. Alqahtani, D. Alsalman, and I. H. Sarker, "Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques," *Ann. Data Sci.*, 2022, doi: 10.1007/s40745-022-00379-8.