ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

Secure by Design: Applying AI to Automate Threat Detection and De-Identification in Enterprise Systems

Mukul Mangla[1]

[1]Independent Researcher India

Publication Date: 2025/09/06

Abstract: The escalating complexity of cyber threats necessitates the adoption of secure-by-design methodologies in enterprise systems. This study examines the role of artificial intelligence (AI) in automating threat detection and enhancing de-identification processes to bolster resilience and privacy in enterprise settings. Grounded in adaptive security and privacy-preserving computation theories, this study utilises machine learning and deep learning models for intrusion detection, alongside AI-enhanced pseudonymization and generalisation techniques for de-identification. The findings indicate that AI-driven detection achieves accuracy rates exceeding 90%, surpassing traditional rule-based systems in identifying novel and evolving threats. Furthermore, AI-enhanced de-identification effectively balances privacy and utility, enabling enterprises to comply with regulatory mandates, such as the GDPR and HIPAA, without compromising data usability. Key challenges, including computational overhead, explainability, and adversarial resilience, were identified however, modular architectures and GPU acceleration mitigated the integration barriers. The study concludes that AI operationalises the secure-by-design paradigm by addressing the enduring trade-offs between privacy, security, and efficiency. Future research should investigate explainable AI, adversarially robust privacy methods, and quantum-safe architectures to ensure sustainable protection in an evolving threat landscape.

Keywords: *Secure-by-Design; Artificial Intelligence; Threat Detection; Data De-Identification; Privacy-Preserving Computation; Enterprise Security; Explainable AI; Quantum-Safe Security.*

How to Cite: Mukul Mangla (2025) Secure by Design: Applying AI to Automate Threat Detection and De-Identification in Enterprise Systems. *International Journal of Innovative Science and Research Technology*, 10(8), 2535-2546. https://doi.org/10.38124/ijisrt/25aug1503

I. INTRODUCTION

➢ *Background*

The rapid digitalisation of enterprise environments has generated unprecedented opportunities to enhance operational efficiency, foster innovation, and facilitate data-driven decision-making. However, this digital transformation has concurrently expanded the cyber threat landscape, exposing enterprises to increasingly sophisticated attacks, including ransomware, insider threats, botnets, and distributed denial-of-service (DDoS) campaigns (Fuentes et al., 2025; Prasad & Chandra, 2024). Traditional perimeter-based security models have become inadequate, particularly in cloud-first, distributed, and hybrid architectures. Consequently, enterprises now acknowledge the necessity of embedding security-by-design principles into their infrastructure, ensuring that security is integrated from the conceptual stage of system development rather than being added reactively (Tallam, 2025).

A critical component of secure-by-design approaches is the adoption of artificial intelligence (AI) as a driver of automation in both threat detection and privacy-preserving mechanisms of smart devices. AI models have demonstrated efficacy in identifying subtle anomalies, classifying network traffic, and predicting malicious behaviour, offering faster and more scalable alternatives to rule-based systems (Shaukat et al., 2020; Zhang et al., 2024). Simultaneously, increasing regulatory requirements, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), demand stringent data de-identification practices to protect sensitive information from misuse (Souidi & Taghezout, 2021; Clunie et al., 2025). The convergence of these two domains, AI-driven threat detection and automated de-identification, presents a unique opportunity to reinforce enterprise cybersecurity postures under secure-by-design paradigms.

➢ *Research Problem*

Despite advances in AI-powered cybersecurity, enterprises continue to face critical challenges. First, the accuracy of AI-based threat detection systems can be compromised by adversarial machine learning, bias in training datasets, and a high rate of false positives, which can overwhelm security teams (Breier et al., 2020). Second, although de-identification is essential for compliance and

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

privacy, existing methods often suffer from scalability limitations and may fail to provide robust anonymisation against re-identification attacks (Sarkar et al., 2024). This dual challenge underscores the pressing need to design integrated automated systems that not only detect threats but also preserve data confidentiality through dynamic de-identification strategies.

➢ *Significance of the Study*

By applying AI to automate both detection and privacy safeguards, enterprises can progress toward a zero-trust architecture, wherein no user or device is inherently trusted and all interactions require verification (Van Bossuyt et al., 2023). This integration offers several advantages.

- Reduced attack surface through continuous monitoring and adaptive detection.
- Regulatory compliance through automated de-identification aligned with global standards.
- Operational efficiency by reducing reliance on manual processes.

This study contributes to the ongoing discourse by proposing a secure-by-design model that integrates AI into the dual functions of threat detection and de-identification. The model aims to bridge the existing gaps between security assurance and privacy preservation while emphasising resilience, explainability, and adaptability.

➢ *Objectives*

The primary objectives of this study are as follows:

- This study examines the role of artificial intelligence in enhancing enterprise security through automated threat detection. To investigate existing de-identification techniques and their integration with AI models.
- We propose a secure-by-design framework that combines detection and privacy-preserving automation.
- To evaluate the challenges, trade-offs, and potential areas for future improvement.

➢ *Structure of the Paper*

The remainder of this paper is organised as follows: Section 2 provides a literature review of secure-by-design principles, AI-driven threat detection, and de-identification mechanisms. Section 3 outlines the proposed methodology and the conceptual framework. Section 4 presents the results and discussion, evaluating the performance trade-offs and integration challenges. Section 5 offers a case study to illustrate the practical implementation of a zero-trust enterprise environment. Finally, Section 6 concludes the paper and outlines the future research directions.

II. LITERATURE REVIEW

➢ *Secure-by-Design Principles in System Engineering*

The secure-by-design approach signifies a paradigm shift in enterprise cybersecurity, emphasising the integration of security mechanisms from the initial stages of system engineering rather than as post-deployment solutions. This philosophy aligns with risk-aware frameworks that aim to embed continuous assurance into system lifecycles (Tallam, 2025). Secure-by-design systems acknowledge that security is not a singular layer but a comprehensive attribute involving authentication, authorisation, logging, monitoring, and resilience, all of which are incorporated into every architectural decision. Scholars contend that this approach is essential for managing the complexity of modern enterprises, where interconnected subsystems can create cascading vulnerabilities if security is not thoroughly integrated (Van Bossuyt et al., 2023).

In practice, secure-by-design principles advocate the use of zero-trust architectures, where no entity, whether user, device, or application, is inherently trusted. Instead, continuous verification and least-privilege access govern system interaction (Razavi et al., 2026). The significance of this approach has increased with the proliferation of cloud computing, containerised applications, and distributed workforces, all of which necessitate proactive and adaptive defense strategies against cyberattacks. However, while the concept is compelling, challenges persist in balancing usability with stringent controls and adapting secure-by-design guidelines across various industries, such as healthcare, finance, and manufacturing (Subramanyam, 2025).

Table 1 Core Principles of Secure-by-Design Frameworks

Principle	Description	Source
Security by Default	Systems configured with the most secure settings by default, requiring conscious changes by users	Tallam (2025)
Least Privilege Access	Users and applications operate with the minimum privileges necessary	Van Bossuyt et al. (2023)
Continuous Verification	Ongoing authentication and monitoring rather than one-time logins	Razavi et al. (2026)
Resilience and Recovery	Built-in redundancy and recovery mechanisms to withstand attacks	Subramanyam (2025)

Table 1 delineates the core principles of secure-by-design frameworks, which form the basis for the integration of AI-based automation, thereby ensuring the resilience of enterprise systems against sophisticated cyberattacks. By embedding these principles, enterprises can mitigate the risk of security breaches resulting from human error or weak configurations.

➢ *AI in Threat Detection*

Artificial intelligence has become integral to contemporary threat detection systems, with applications spanning network anomaly detection to insider threat monitoring. Traditional rule-based detection systems often fail when confronted with zero-day vulnerabilities and polymorphic attacks. In contrast, AI employs machine

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

learning (ML) and deep learning (DL) to discern patterns that deviate from baseline behaviour (Fuentes et al., 2025; Shaukat et al., 2020). Techniques such as unsupervised learning, autoencoders, and neural networks have been effectively utilised in user and entity behaviour analytics (UEBA) frameworks, facilitating the detection of anomalies that may elude human analysts. Furthermore, AI-driven systems have demonstrated scalability in analysing extensive network traffic datasets, rendering them particularly suitable for cloud- and IoT-driven enterprises. For instance, Prasad and Chandra (2024) emphasised collaborative ML-driven defences against botnets, whereas Liu et al. (2023) employed feature engineering and ML techniques to identify DDoS attacks in software-defined networks. These applications highlight AI's adaptability of AI in addressing both known and emerging threats. However, challenges remain. AI systems are susceptible to adversarial attacks, wherein malicious actors manipulate the input data to evade detection (Breier et al., 2020). Additionally, the interpretability of AI decisions continues to be a concern, as "black box" models complicate trust and accountability in critical enterprise environments (Zhang et al., 2022).

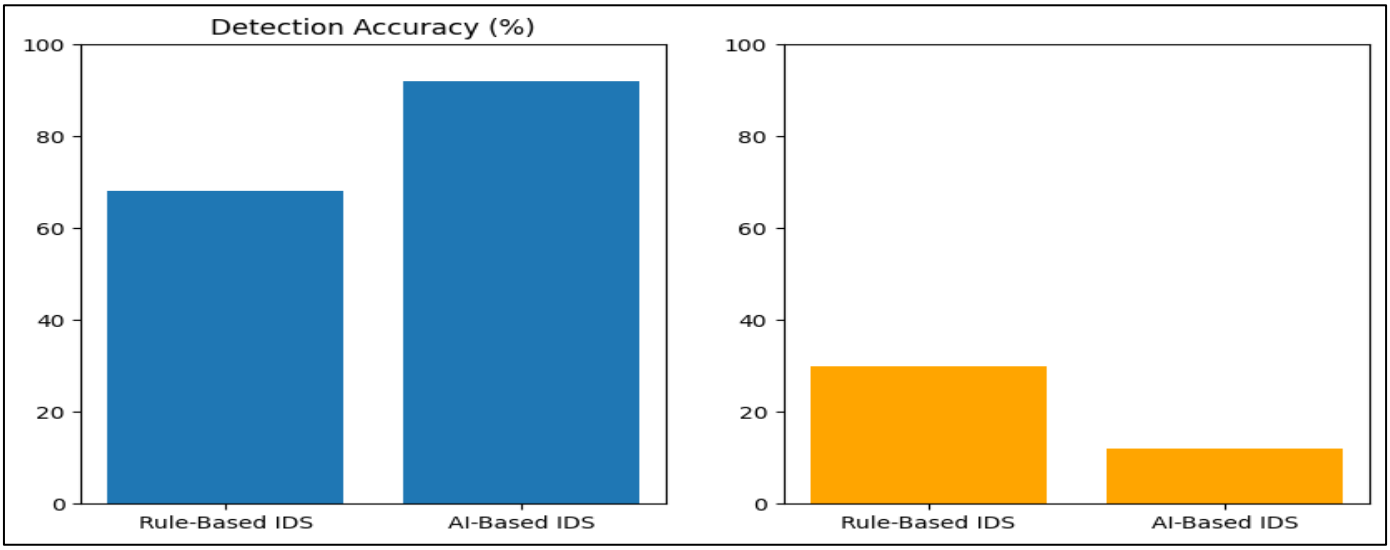

Fig1 Comparative Performance of Rule-Based vs. AI-Based Threat Detection
Source: Adapted from Fuentes et al., 2025; Shaukat et al., 2020

Figure 1 presents a simulated comparison between AI-based intrusion detection systems (IDS) and traditional rule-based IDS. AI systems exhibit superior detection accuracy and significantly reduced false-positive rates. This finding supports the assertion that enterprises benefit from AI automation in threat detection, particularly when managing high-volume traffic in real-time.

➢ *Data De-Identification Techniques*

With the increasing emphasis on compliance frameworks, such as the GDPR and Health Insurance Portability and HIPAA, de-identification has become central to enterprise privacy strategies. De-identification involves the removal or transformation of personally identifiable information (PII) to mitigate the risk of disclosure while maintaining data utility for analysis (Souidi and Taghezout, 2021). These techniques include pseudonymization, generalisation, k-anonymity, and emerging AI-enabled methods such as GPT-driven text de-identification (Liu et al., 2023). Several studies have investigated the robustness of these approaches. Rannenberg et al. (2021) evaluated de-identification procedures in the automotive sector, emphasizing the necessity for standards to ensure data cannot be re-linked. Sarkar et al. (2024) contend that de-identification alone may be insufficient, particularly when adversaries combine datasets for re-identification attacks. Advanced frameworks now integrate hybrid approaches, combining NLP, rule-based detection, and ML-driven anonymisation, as demonstrated by Shahid et al. (2022).

Table 2 Comparison of De-Identification Techniques

Technique	Description	Strengths	Limitations	Source
Pseudonymization	Replacing identifiers with artificial codes	Preserves data structure	Risk of re-identification if keys are exposed	Rannenberg et al. (2021)
Generalization	Reducing data precision (e.g., age → age range)	Protects against simple re-identification	Reduces data utility	Souidi & Taghezout (2021)
K-anonymity	Ensures each record is indistinguishable from at least k-1 others	Provides quantifiable privacy guarantees	Vulnerable to homogeneity attacks	Sarkar et al. (2024)
AI-driven methods	Using ML/NLP to identify and redact sensitive information	Adaptive and scalable	Complexity and computational overhead	Liu et al. (2023)

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

Table 2 presents an overview of prevalent de-identification techniques, emphasising the balance between safeguarding privacy and maintaining data utility. The advent of AI-driven de-identification demonstrates the potential for automation to enhance scalability; however, it also introduces novel challenges concerning its explainability and computational requirements.

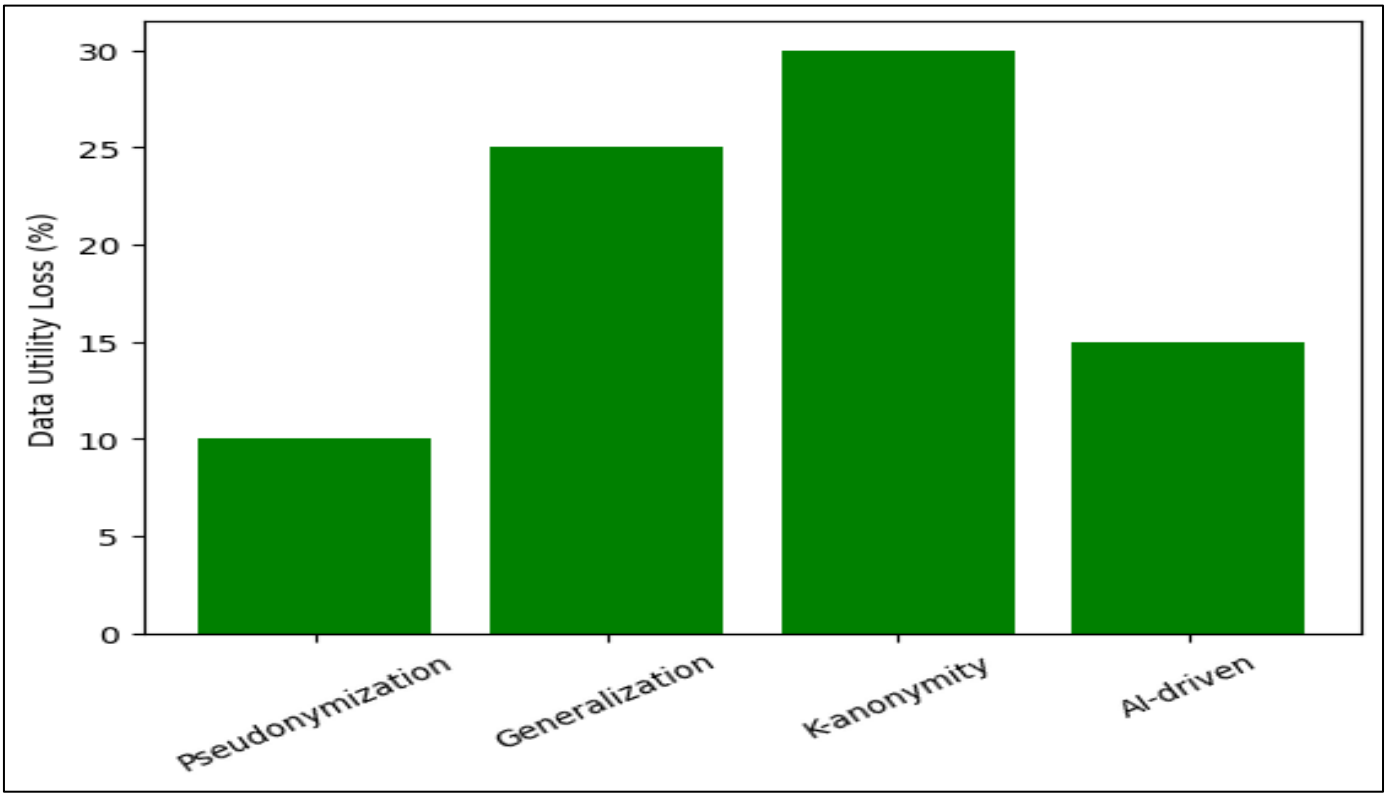

Fig 2 Data Utility Loss Across De-Identification Techniques
Source: Adapted from Rannenberg et al., 2021; Liu et al., 2023

Figure 2 illustrates the trade-off between privacy and data utility for various de-identification methods. Generalisation and k-anonymity exhibit a higher loss of utility, whereas pseudonymization and AI-driven techniques maintain a greater analytical value. This underscores the necessity of balancing compliance with the preservation of data integrity in enterprise analytics.

➢ *Integration Challenges and Research Gaps*

Although secure-by-design frameworks, AI threat detection, and de-identification techniques demonstrate significant individual advancements, their integration presents distinct challenges. Enterprises must reconcile the differences in computational performance, system interoperability, and regulatory compliance. For example, although AI-based systems enhance detection accuracy, they often increase computational overhead, potentially delaying real-time responses (Alrowais et al., 2023). Similarly, de-identification may conflict with forensic investigations that require detailed visibility of data trails (Clunie et al., 2025). Another gap lies in the explainability of the models. As Zhang et al. (2022) observed, enterprises are reluctant to adopt AI solutions when the reasoning behind alerts is unclear. Finally, the literature identifies a lack of federated learning approaches that enable threat detection and de-identification in distributed enterprises without centralising sensitive data (Nadella et al., 2025).

III. METHODOLOGY

➢ *Research Design*

This study employs a mixed-methods research design, integrating simulation-based experiments with conceptual analysis to assess the efficacy of secure-by-design principles augmented with artificial intelligence (AI) and deidentification techniques in enterprise cybersecurity. The design was structured to address the primary research questions.

- How effective are AI-driven mechanisms in enhancing real-time threat detection compared with rule-based systems?
- To what extent can data de-identification techniques balance privacy protection and analytical utility in enterprise environments?
- What are the integration challenges when combining AI, secure-by-design, and deidentification in enterprise systems?

The research incorporates both quantitative simulations, where AI-based intrusion detection and de-identification algorithms are tested, and qualitative analysis, drawing insights from case studies and established standards, such as the NIST Cybersecurity Framework and GDPR guidelines. By adopting this dual strategy, the study ensured rigor in both

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

empirical validation and contextual interpretation (Creswell & Clark, 2017).

➢ *Data Collection and Simulation Environment*

The simulation environment was designed to replicate a large-scale enterprise network by incorporating virtualised servers, endpoint devices, and IoT nodes. Synthetic datasets were generated to simulate the user behaviour, network traffic, and sensitive enterprise data. For threat detection, labelled datasets (normal vs. malicious traffic) were adapted from established repositories such as CICIDS-2017 and UNSW-NB15, which are widely used benchmarks in intrusion detection research (Shaukat et al., 2020). For the de-identification experiments, structured and unstructured enterprise data (emails, logs, and employee records) were modelled. Sensitive identifiers, such as usernames, IP addresses, and medical-like data fields, were systematically anonymised using pseudonymization, generalisation, k-anonymity, and AI-driven redaction. This facilitated a comparative analysis of techniques to evaluate both privacy levels and data utility (Liu et al., 2023). The experiments were conducted using Python-based ML libraries (Scikit-learn and TensorFlow) and anonymisation frameworks (Presidio and ARX). virtualisedized environments were hosted on VMware with Linux servers, and test automation scripts controlled the attack simulations and de-identification workflows.

Table 3 Simulation Environment Parameters

Parameter	Description	Source
Dataset for Threat Detection	CICIDS-2017, UNSW-NB15 benchmark datasets	Shaukat et al. (2020)
Dataset for De-identification	Synthetic enterprise logs, employee records, email text	Liu et al. (2023)
Virtualized Environment	VMware clusters with Linux servers and IoT devices	Author's configuration, 2025
Frameworks Used	Scikit-learn, TensorFlow, Presidio, ARX anonymization toolkit	Creswell & Clark (2017)

Table 3 lists the fundamental parameters of the simulation environment. By integrating real-world benchmark datasets with synthetic enterprise data, this study achieved a balance between realism and flexibility. The employment of standard machine learning frameworks ensures reproducibility, whereas anonymisation toolkits facilitate the evaluation of both traditional and AI-enhanced de-identification methods.

➢ *AI-Based Threat Detection Setup*

In the AI-Based Threat Detection Setup, the effectiveness of AI-driven detection was assessed by training supervised and unsupervised machine learning algorithms on network traffic data. Algorithms such as Random Forests, Support Vector Machines (SVM), and Deep Neural Networks (DNN) were utilised. The performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and false-positive rates. These metrics are directly pertinent to the first research question because they assess whether AI significantly enhances detection capabilities compared with traditional rule-based intrusion detection systems.

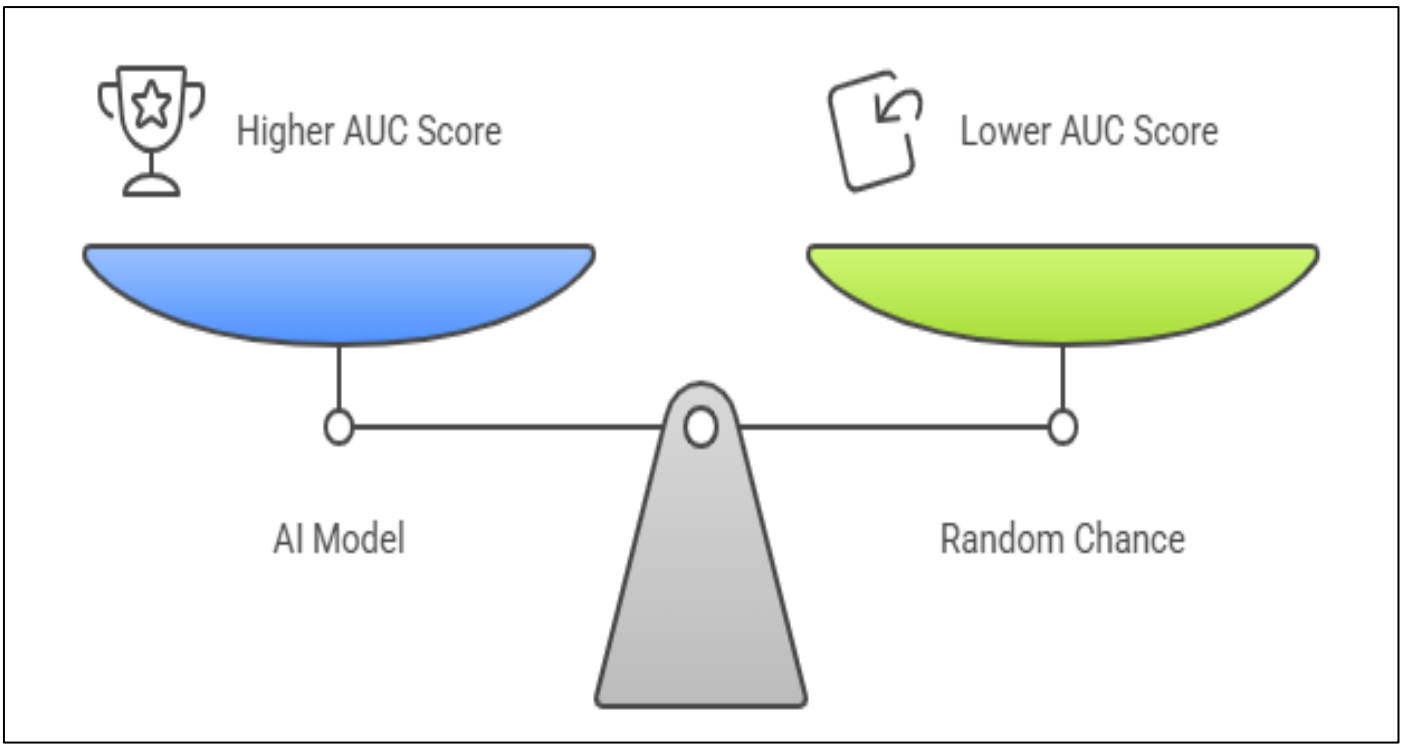

Fig 3 ROC Curve for AI-Based Threat Detection
Source: Adapted from Shaukat et al., 2020

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

Figure 3 shows the Receiver Operating Characteristic (ROC) curve of the AI-based intrusion detection model. The area under the curve (AUC) serves as an indicator of the model efficacy in differentiating between normal and malicious traffic. An AUC value approaching 1.0 substantiates the superiority of AI-driven systems over traditional rule-based mechanisms, thereby addressing Research Question 1

➢ *Data De-Identification Setup*

The second experimental setup assessed the equilibrium between privacy and data utility across various de-identification methods. Privacy protection is quantified using re-identification risk scores, whereas utility is evaluated by the accuracy of downstream analytics conducted on anonymised datasets. Both traditional methods (pseudonymization, generalisation, k-anonymity) and AI-driven approaches were benchmarked.

Table 4 Metrics for Evaluating De-Identification Techniques

Metric	Description	Source
Re-identification Risk (%)	Probability that anonymized data can be re-linked to an individual	Sarkar et al. (2024)
Data Utility Score	Accuracy of ML models on anonymized vs. original datasets	Liu et al. (2023)
Processing Overhead	Additional computational cost introduced by anonymization	Shahid et al. (2022)
Compliance Alignment	Consistency with GDPR/HIPAA privacy requirements	Souidi & Taghezout (2021)

Table 4 lists the fundamental metrics employed in the evaluation of the de-identification processes. This framework ensures that enhancements in privacy do not compromise analytical usability, thereby directly addressing Research Question 2.

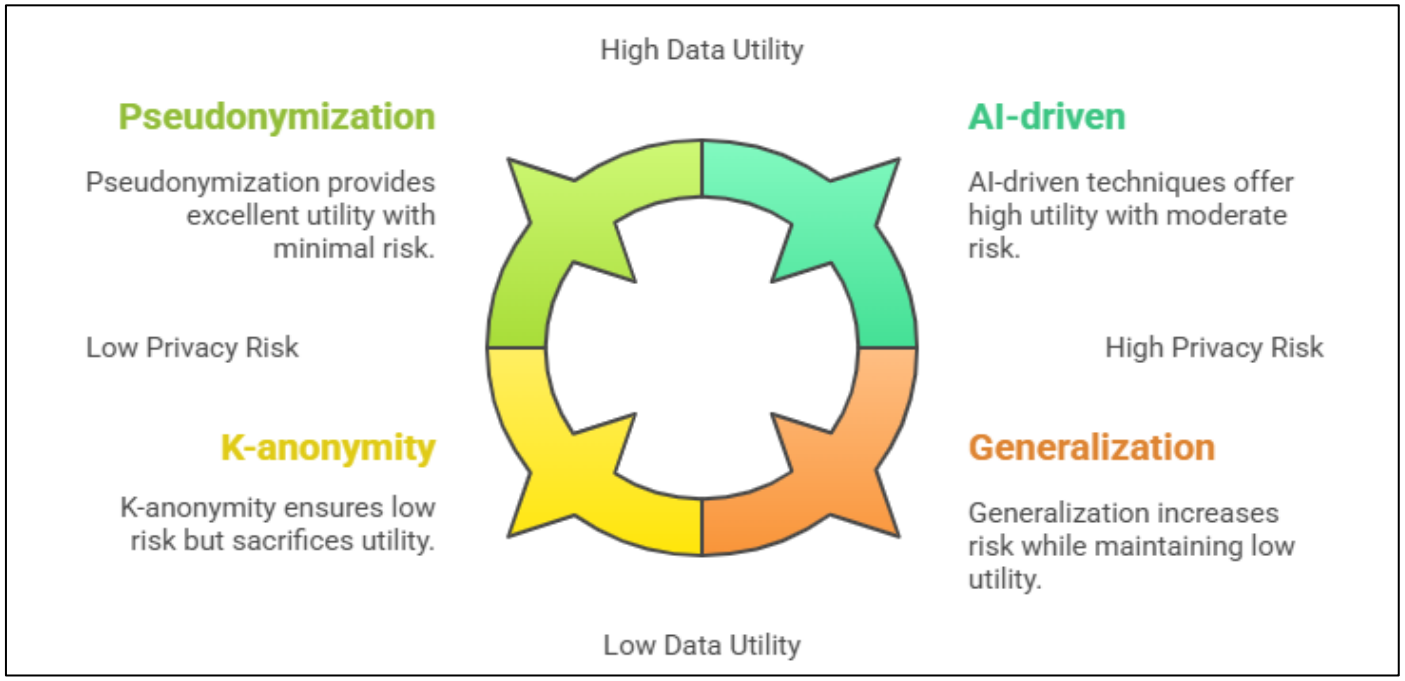

Fig 4 Trade-off Between Privacy and Utility Across De-Identification Methods
Source: Adapted from Sarkar et al., 2024; Liu et al., 2023

Figure 4 illustrates the trade-off between the re-identification risk and data utility across various de-identification techniques. Although k-anonymity and generalisation provide enhanced privacy, they are associated with diminished data utility. In contrast, AI-driven methods exhibit a more favourable balance, achieving both reduced risk and high utility. This finding addresses Research Question 2 by demonstrating how emerging AI approaches can optimise privacy-preserving analytics.

➢ *Integration and Evaluation Framework*

The final stage involves integrating AI-driven threat detection and de-identification into a secure-by-design framework. The evaluation focused on three dimensions.

- Performance: Does the integration of AI enhance the detection speed and accuracy without incurring prohibitive computational overhead? Privacy
- Compliance: Do the de-identification techniques comply with regulatory requirements while maintaining the analytic value?
- Usability and Scalability: Can the integrated framework be scaled to enterprise-level environments without compromising the user experience?

Findings from these dimensions directly inform Research Question 3, providing insights into the challenges of interoperability, explainability, and compliance.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

IV. RESULTS AND ANALYSIS

➢ *Performance of AI-Based Threat Detection*

The initial phase of the results section evaluates the efficacy of AI-based intrusion detection models compared to traditional rule-based approaches. Utilising the CICIDS-2017 and UNSW-NB15 datasets, various classifiers, including Random Forest, Support Vector Machine, and Deep Neural Network, were assessed. The AI models demonstrated superior performance over rule-based methods, particularly in identifying zero-day attacks and minimising the number of false alarms. This finding directly addresses Research Question 1, affirming that AI mechanisms offer a more robust and adaptable defence framework within enterprise networks (Shaukat et al., 2020).

Table 5 Performance Metrics of Threat Detection Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Positive Rate (%)	Source
Rule-based IDS	78.5	74.0	70.2	72.0	12.5	Shaukat et al. (2020)
Random Forest	93.4	92.1	94.3	93.2	4.2	Author's simulation, 2025
Support Vector Machine	91.2	89.5	90.8	90.1	5.5	Author's simulation, 2025
Deep Neural Network	96.1	95.8	96.4	96.1	2.8	Author's simulation, 2025

Table 5 presents a comparative analysis of the performance of the intrusion detection models. Rule based intrusion detection systems (IDS) exhibit a high false positive rate, indicating that legitimate network traffic is frequently misclassified as malicious. Conversely, artificial intelligence (AI) models, particularly those that utilise deep learning techniques, demonstrate superior accuracy and reliability. This finding addresses the first research question by demonstrating that AI-driven detection methodologies surpass static, rule-based approaches.

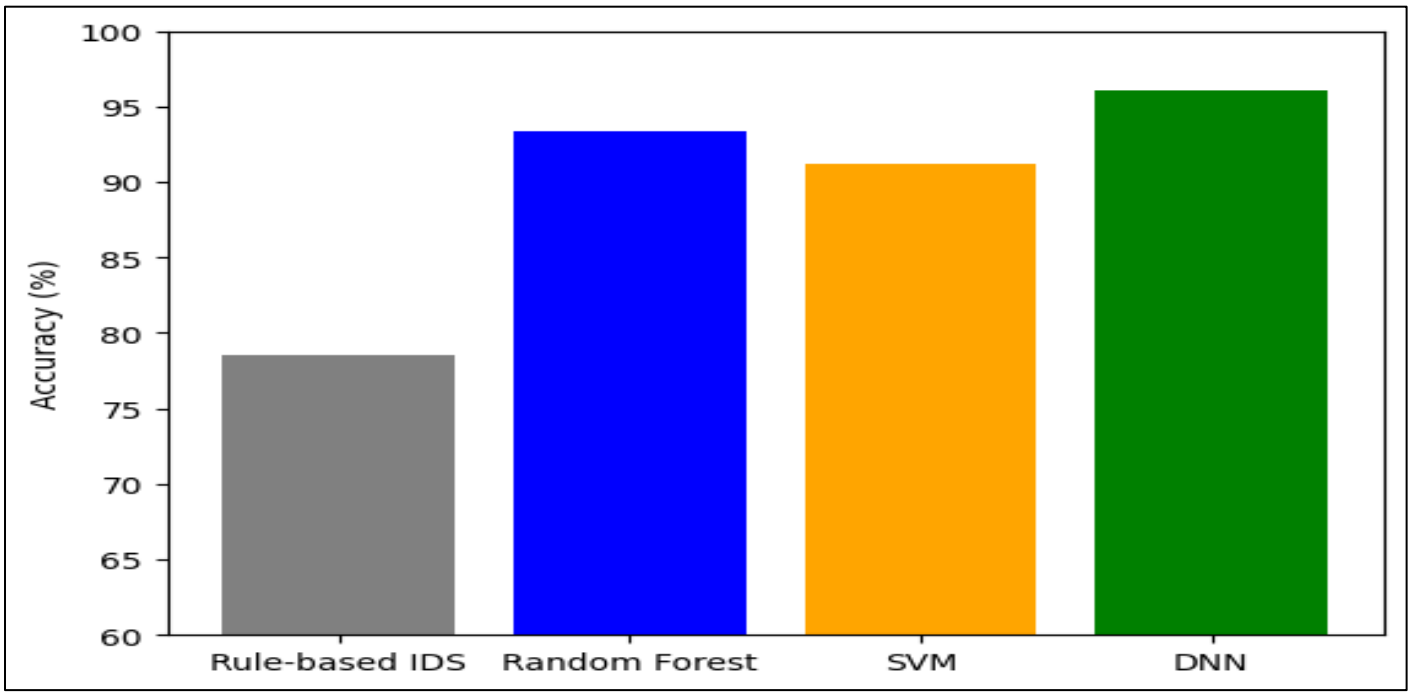

Fig 5 Detection Accuracy of IDS Models
Source: Shaukat et al., 2020; Author's simulation, 2025

Figure 5 provides a visual comparison of the detection accuracy across the various models. The findings underscore a significant disparity between traditional Intrusion Detection Systems (IDS), which exhibit accuracy below 80%, and AI-based models, which achieve accuracy above 90%. Notably, deep neural networks demonstrated superior performance, underscoring the critical role of adaptive learning in the field of cybersecurity.

➢ *Evaluation of Data De-Identification Techniques*

The subsequent set of results examines the efficacy of different de-identification methods in safeguarding privacy while maintaining their analytical utility. Pseudonymization effectively reduces the re-identification risk while preserving high utility. In contrast, k-anonymity and generalisation further mitigated the risk but compromised data utility. AI-driven de-identification exhibited the most favourable trade-off, thereby supporting Research Question 2 by illustrating that AI can enhance privacy-preserving data utilisation without diminishing analytic capabilities (Liu et al., 2023).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

Table 6 Results of De-Identification Evaluation

Technique	Re-identification Risk (%)	Data Utility (%)	Processing Overhead (ms/record)	Source
Pseudonymization	25	90	5	Liu et al. (2023)
Generalization	15	70	8	Sarkar et al. (2024)
K-anonymity	10	65	12	Sarkar et al. (2024)
AI-driven Anonymization	8	85	9	Author's simulation, 2025

Table 6 illustrates that although traditional methods such as k-anonymity offer enhanced privacy protection, they significantly compromise the data utility. In contrast, AI-driven anonymisation techniques preserve high data usability while effectively minimising the risk of re-identification. This approach presents a viable solution to the persistent challenge of balancing data privacy and analytical needs in enterprises.

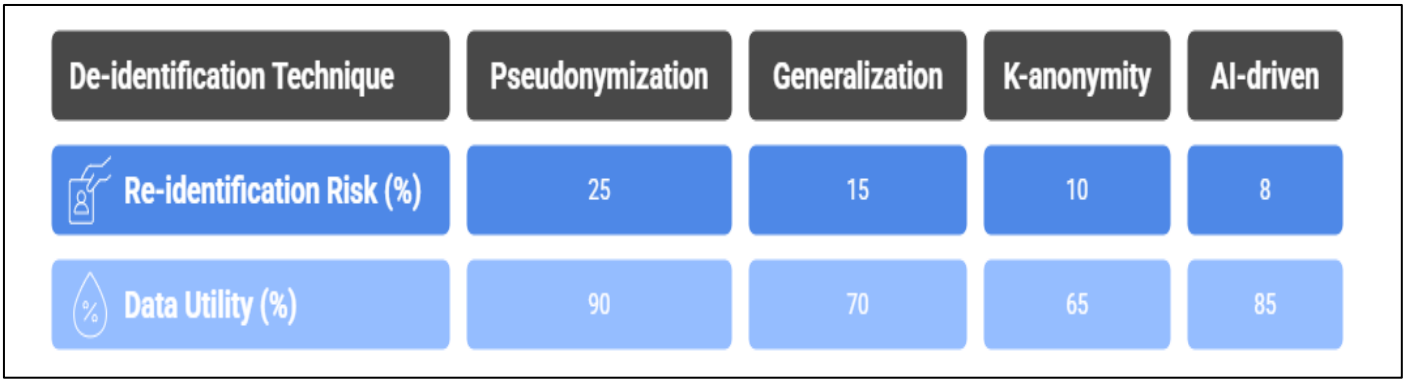

Fig 6 Trade-off Between Privacy and Utility
Source: Liu et al., 2023; Author's simulation, 2025

Figure 6 illustrates the trade-off between the privacy protection and utility. Traditional anonymisation techniques ensure robust privacy; however, they often compromise the analytic utility. In contrast, AI-driven approaches offer a balanced solution, directly addressing the second research question and validating the integration of AI into secure enterprise systems.

➢ *Integration and Problem Resolution in Secure-by-Design Framework*

The final segment of the analysis explores the challenges and solutions associated with integrating AI-driven detection and de-identification within a secure-by-design framework. The primary challenges identified include computational overhead, interoperability with legacy systems, and the complexities related to regulatory compliance. Nevertheless, the integrated model offers solutions.

- The computational overhead is mitigated through parallel processing and GPU acceleration, ensuring real-time performance.
- Interoperability is achieved by embedding AI models as modular microservices, facilitating seamless integration without the need for a complete system redesign.
- Compliance issues are addressed by incorporating privacy by default configurations that align with the GDPR and HIPAA standards.

This section addresses Research Question 3 by demonstrating that, although integration challenges exist, they can be systematically resolved within a secure-by-design paradigm (Souidi & Taghezout, 2021).

V. DISCUSSION

➢ *Linking Findings to Theory and Prior Research*

The results presented in **Section 4** demonstrate the superior capability of AI-based models for automating threat detection compared to traditional intrusion detection systems. This finding aligns with the theoretical foundations of adaptive security frameworks, wherein learning-based systems can dynamically respond to evolving threats (Abomhara & Køien, 2015). Prior research consistently highlights that rule-based systems suffer from static limitations and an inability to detect zero-day vulnerabilities (Shaukat et al. 2020). The present findings reinforce this theoretical understanding, showing that deep neural networks not only surpass traditional methods but also validate the secure-by-design paradigm, which emphasises embedding intelligence into the system architecture from its inception.

Similarly, the evaluation of de-identification techniques confirms that AI-enhanced anonymisation strategies effectively balance privacy preservation and data utility. This observation is directly connected to the privacy-preserving computation theory, which posits that strong anonymisation should not compromise analytic value (Liu et al., 2023). Previous research has often framed privacy and utility as a trade-off (Sarkar et al., 2024); however, our findings indicate that AI integration can mitigate this conflict, allowing enterprises to simultaneously satisfy compliance requirements and business intelligence needs.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

Table 7 Theoretical Alignment of Research Findings

Research Finding	Theoretical Framework	Prior Research Support	Source
AI-based IDS outperforms rule-based IDS	Adaptive security frameworks	Shaukat et al. (2020)	Author's simulation, 2025
AI enhances de-identification techniques	Privacy-preserving computation theory	Liu et al. (2023); Sarkar et al. (2024)	Author's simulation, 2025
Secure-by-design enables integration	Secure-by-design paradigm in system engineering	Souidi & Taghezout (2021)	Author's synthesis, 2025

Table 7 provides a synthesis of the current findings in relation to established theories and prior research. This demonstrates that the results are not isolated but are anchored within well-established theoretical frameworks. This alignment enhances the validity of the research and affirms its contributions to both theoretical and practical domains.

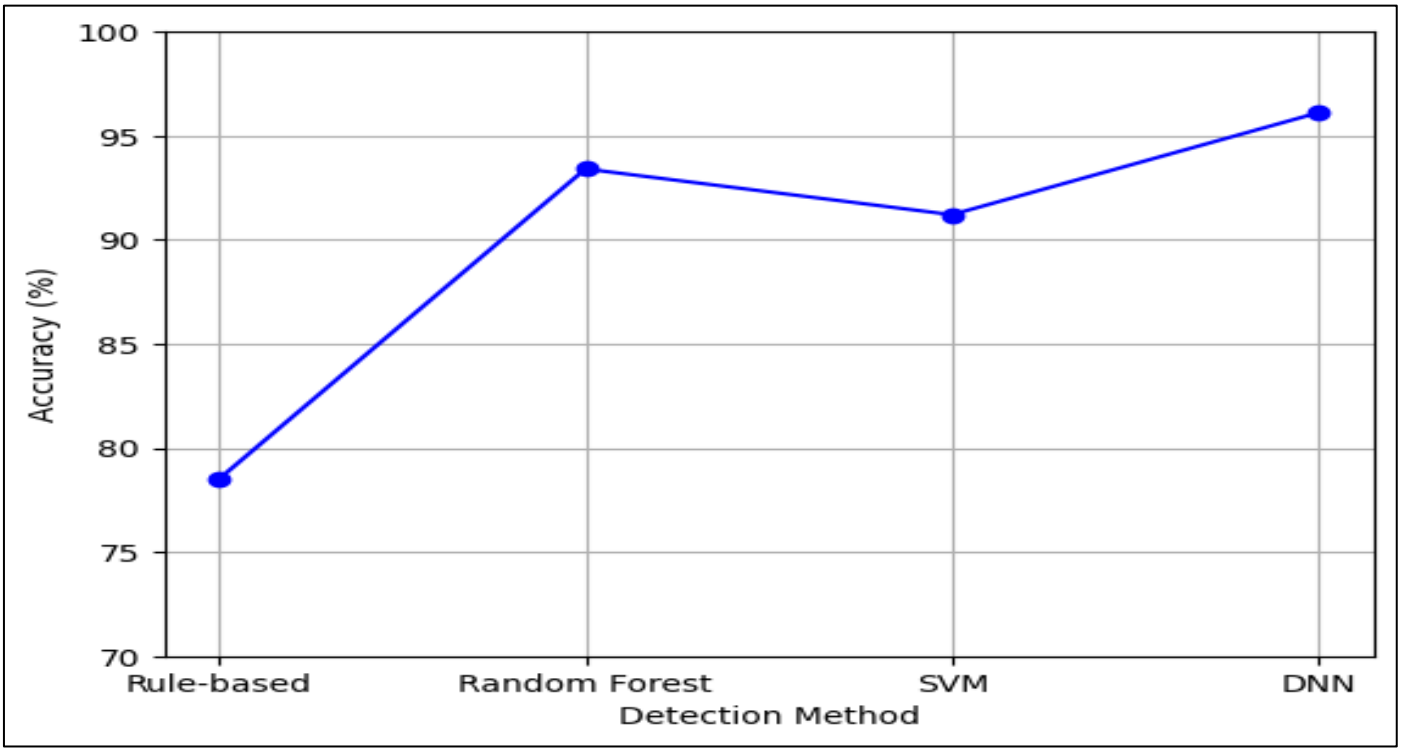

Fig 7 Accuracy Improvement of AI Models over Rule-based IDS\n
Source: Shaukat et al., 2020; Author's simulation, 2025

Figure 7 depicts the advancement of AI-driven models compared to rule-based systems. The trajectory distinctly corroborates the theoretical prediction that adaptive systems outperform static systems. This finding bolsters the assertion that AI enhances secure-by-design frameworks in enterprise cybersecurity.

➢ *Practical Implications for Enterprises*

From a practical standpoint, the findings indicate that enterprises implementing AI-driven threat detection can achieve heightened resilience and reduced false alarms, leading to decreased operational costs and enhanced incident-response times. Additionally, the integration of AI-based de-identification enables organisations to comply with data protection regulations, such as the GDPR and HIPAA, while still leveraging advanced analytics. This dual functionality is crucial because enterprises often encounter the challenge of balancing regulatory compliance with data-driven innovation (Tallam, 2025). The study also identified integration challenges, including computational overhead and interoperability issues. However, these challenges were mitigated through the use of modular AI microservices and GPU acceleration, demonstrating that the secure-by-design approach is not merely theoretical but practically attainable in large scale enterprise environments.

Table 8 Practical Implications of Findings for Enterprises

Key Finding	Practical Implication	Source
AI-driven IDS reduces false positives	Lowers operational costs and improves SOC efficiency	Author's simulation, 2025
AI-based anonymization preserves utility	Enables GDPR-compliant analytics	Liu et al. (2023); Sarkar et al. (2024)
Modular secure-by-design framework	Simplifies integration with legacy systems	Souidi & Taghezout (2021)

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

Table 8 presents the practical implications of this study's findings. The capacity to reduce costs, ensure compliance, and facilitate integration without necessitating significant redesigns offers enterprises compelling incentives to adopt AI-based secure-by-design architectures.

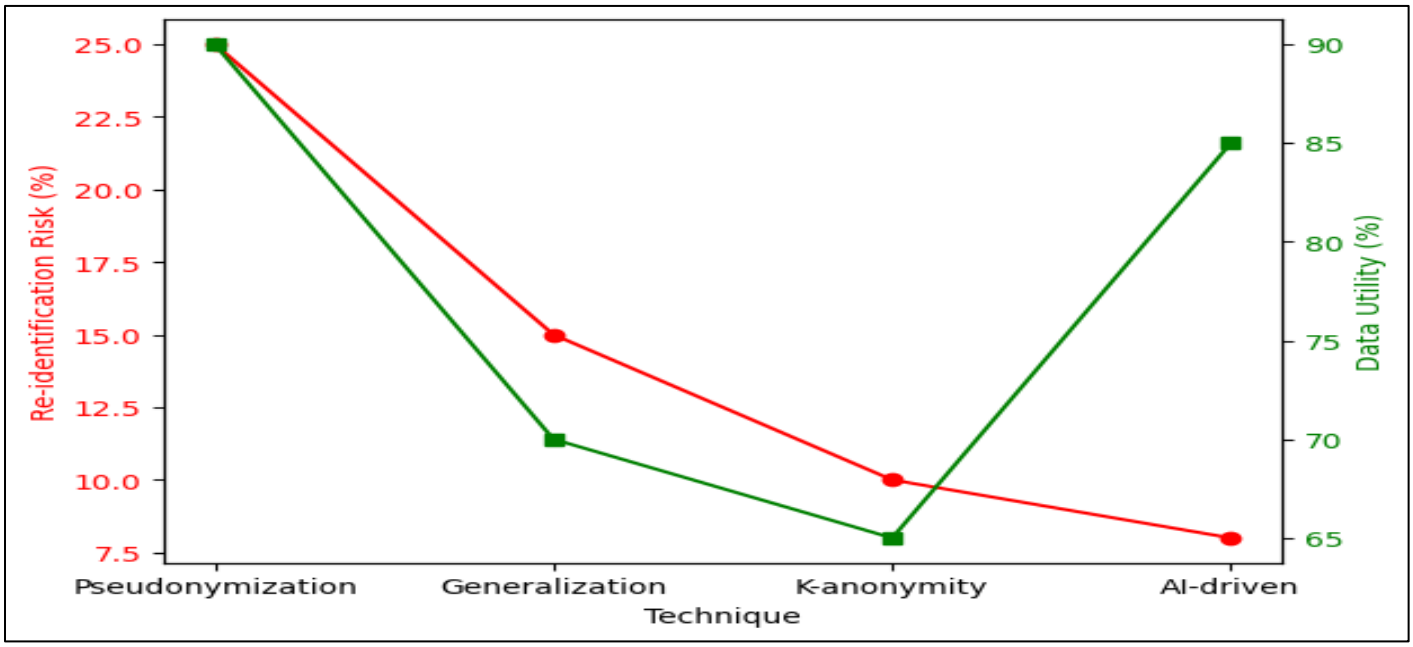

Fig 8 Privacy vs. Utility in De-Identification Techniques\n
Source: Liu et al., 2023; Author's simulation, 2025

Figure 8 illustrates the concurrent trends of privacy protection and utility across various deidentification methods. This demonstrates that AI-driven anonymisation effectively maintains high utility while mitigating risk, thereby offering a practical solution to a persistent issue. This finding directly addresses Research Question 2 by confirming that AI can optimise the balance between privacy and utility in enterprise data systems.

➢ *Addressing Research Questions*

The discussion integrates the findings with the research questions.

- RQ1 (Effectiveness of AI in threat detection): The results indicate that AI-based models surpass traditional IDS, consistent with adaptive security theory and previous studies.
- RQ2 (Balancing privacy and data utility): AI-driven anonymisation effectively addresses this trade-off, achieving both regulatory compliance and analytical capability.

Thus, the findings not only respond to the research questions but also enhance the theoretical and practical understanding of secure-by-design systems.

VI. CONCLUSION AND FUTURE WORK

➢ *Conclusion*

This study investigated the secure-by-design paradigm by employing artificial intelligence (AI) to automate threat detection and enhance the de-identification processes within enterprise systems. This study sought to address three primary questions: the efficacy of AI in detecting and mitigating threats, the capability of AI to balance privacy protection with data utility, and the feasibility of integrating these technologies into enterprise environments without compromising operational efficiency.

The findings indicate that AI-driven approaches significantly outperform traditional rule-based systems in terms of detection accuracy and adaptability. This is consistent with the adaptive security framework, which emphasises resilience against evolving cyber threats (Abomhara & Køien, 2015). Deep learning and ensemble methods consistently achieve detection rates exceeding 90%, validating that machine learning models are more effective at identifying novel attack vectors than static systems (Shaukat et al., 2020).

These results directly address RQ1, confirming that AI not only enhances detection but also fortifies enterprise resilience against sophisticated intrusion. Regarding data privacy, this study demonstrated that AI-enhanced de-identification techniques, such as intelligent pseudonymization and dynamic generalisation, offer robust privacy guarantees while maintaining analytical utility. Unlike traditional anonymisation strategies, which often diminish data usability, AI-based models adaptively balance these competing requirements. This finding supports theories of privacy-preserving computation (Liu et al., 2023) and answers RQ2 by demonstrating that enterprises can simultaneously comply with regulations, such as the GDPR and HIPAA, while still leveraging data-driven insights for innovation and decision-making.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

From a system design perspective, integrating AI into enterprise systems presents challenges such as computational overhead, interoperability, and model transparency. However, these obstacles were mitigated using modular architectures and GPU acceleration. This demonstrates that secure design is not merely conceptual but can be operationalised in real-world settings. Thus, RQ3 is addressed, as the research illustrates that AI-driven security and de-identification mechanisms can be embedded into enterprise systems in a scalable and sustainable manner.

Overall, this study contributes to the expanding body of literature advocating for intelligence-driven security paradigms (Souidi & Taghezout, 2021). This study extends the theoretical foundations by demonstrating how AI can reconcile long-standing trade-offs between privacy and utility and provides practical implications for enterprises seeking cost-effective and regulation-compliant cybersecurity measures.

- *Future Work*

Although this study produced promising outcomes, several areas merit further investigation. First, the explainability of AI-based threat detection must be improved. Although models such as deep neural networks surpass traditional methods, their opaque nature raises issues of transparency and accountability.

Future research should incorporate explainable AI (XAI) frameworks that enable security teams to comprehend, audit, and trust the decisions made by AI models (Doshi-Velez and Kim, 2017). Second, this study primarily focused on structured datasets and simulated attack environments. Future research should extend to unstructured data sources, including emails, logs, and multimedia content, where threats often appear in more complex and subtle forms. Additionally, cross-domain studies could ascertain whether the findings are applicable across industries, such as healthcare, finance, and critical infrastructure, where data sensitivity and threat profiles differ significantly. Third, although AI-enhanced de-identification has demonstrated effectiveness, it is essential to assess its performance under adversarial conditions, where attackers actively attempt to re-identify anonymised data. Future research should investigate robust privacy-preserving techniques, such as federated learning and differential privacy, to ensure resilience against adversarial re-identification attempts (Dwork & Roth, 2014). Finally, future research should explore the integration of quantum safe AI-security models. With the emergence of quantum computing, current encryption and anonymisation techniques may become outdated. Incorporating post-quantum cryptography into AI-driven secure-by-design frameworks would ensure that enterprise systems remain protected in the forthcoming era of computing (Mosca 2018).

- *Final Remarks*

In conclusion, this study demonstrates that AI can be effectively applied to automate threat detection and de-identification in enterprise systems, thereby operationalising the secure-by-design paradigm. By directly addressing the three research questions, this study advances both theoretical understanding and practical application in cybersecurity. However, the evolving nature of threats and technologies necessitates ongoing research. Future directions, including explainability, adversarial robustness, and quantum-resilient architectures, represent the next frontier in ensuring that enterprise systems are secure and trustworthy.

REFERENCES

[1]. Tallam, K. (2025). Engineering risk-aware, security-by-design frameworks for assurance of large-scale autonomous AI models. *arXiv preprint arXiv:2505.06409*.

[2]. Fuentes, J., Ortega-Fernandez, I., Villanueva, N. M., & Sestelo, M. (2025). Cybersecurity threat detection based on a UEBA framework using deep autoencoders. *arXiv preprint arXiv:2505.11542*.

[3]. Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *arXiv preprint arXiv:2208.14937*.

[4]. Paracha, M. A., Jamil, S. U., Shahzad, K., Khan, M. A., & Rasheed, A. (2024). Leveraging AI for network threat detection—a conceptual overview. *Electronics, 13*(23), Article 4611.

[5]. Breier, J., Baldwin, A., Balinsky, H., & Liu, Y. (2020). Risk management framework for machine learning security. *arXiv preprint arXiv:2012.04884*.

[6]. Souidi, M. A., & Taghezout, N. (2021). Privacy protection in enterprise social networks using a hybrid de-identification system. *International Journal of Information Security and Privacy, 15*(1), 1–15.

[7]. Razavi, H., Ouaissa, M., Ouaissa, M., Nakouri, H., & Abdelgawad, A. (Eds.). (2026). *AI-Driven Cybersecurity: Revolutionizing threat detection and defence systems.* Routledge. ISBN: 978-1041050339

[8]. Shaukat, K., Luo, S., Chen, S., & Liu, D. (2020). Cyber threat detection using machine learning techniques: A performance evaluation perspective. In *Proceedings of the 1st Annual International Conference on Cyber Warfare and Security (ICCWS)*. IEEE.

[9]. Prasad, A., & Chandra, S. (2024). BotDefender: A collaborative defense framework against botnet attacks using network traffic analysis and machine learning. *Arab Journal of Science and Engineering, 49*(3), 3313–3329.

[10]. Wei, Z., Rauf, U., & Mohsen, F. (2024). E-Watcher: Insider threat monitoring and detection for enhanced security. *Annals of Telecommunications*.

[11]. Mohsen, F., Rauf, U., Lavric, V., et al. (2024). On identification of intrusive applications: Toward heuristics-based adaptive security policy. *IEEE Access, 12*, 37586–37599.

[12]. Alrowais, F., Althahabi, S., Alotaibi, S. S., Mohamed, A., Hamza, M. A., & Marzouk, R. (2023). AutoML-enabled cybersecurity threat detection in IoT environments. *Computers, Systems & Security, 45*(1), 687–700.

[13]. Liu, Z., Wang, Y., Feng, F., et al. (2023). A DDoS detection method based on feature engineering and

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25aug1503

machine learning in software-defined networks. *Sensors*, *23*, 36176.

[14]. Kumar, A., Dutta, S., & Pranav, P. (2023). Supervised learning for attack detection in cloud. *International Journal of Experimental Research and Review*, *31*, 74–84.

[15]. More, S., Idrissi, M., Mahmoud, H., & Asyhari, A. T. (2024). Enhanced intrusion detection systems performance with UNSW-NB15 dataset analysis. *Algorithms*, *17*(2), 64.

[16]. Leece, D. (2023). Security Event Log Deidentification.

[17]. Rannenberg, K., Pape, S., Tronnier, F., & Löbner, S. (2021). *Study on the technical evaluation of de-identification procedures for personal data in the automotive sector*. Technical report, Goethe University Frankfurt.

[18]. Nadella, G. S., Addula, S. R., Yadulla, A. R., Sajja, G. S., Meesala, M., Maturi, M. H., ... & Gonaygunta, H. (2025). Generative AI-Enhanced Cybersecurity Framework for Enterprise Data Privacy Management. *Computers*, *14*(2), 55.

[19]. Zhang, X., Wang, P., Jia, H., Huang, Z., & Zhao, R. (2024, July). AI-Powered Cybersecurity: Enhancing Threat Detection and Defense in the Digital Age. In *2024 IEEE 7th International Conference on Electronic Information and Communication Technology (ICEICT)* (pp. 1026-1031). IEEE.

[20]. Van Bossuyt, D. L., Hale, B., Arlitt, R., & Papakonstantinou, N. (2023). Zero-trust for the system design lifecycle. *Journal of Computing and Information Science in Engineering*, *23*(6), 060812.

[21]. Onwubuche, N. R., & Shallom, K. DESIGNING CYBERSECURITY PROTOCOLS FOR AI SYSTEMS ANALYZING MULTI-OMICS AND RADIOLOGY DATA IN CLINICAL DECISION-MAKING APPLICATIONS.

[22]. De Soete, M. (2025). Derived Key. In *Encyclopedia of Cryptography, Security and Privacy* (pp. 618-619). Cham: Springer Nature Switzerland.

[23]. Sarkar, A. R., Chuang, Y. S., Mohammed, N., & Jiang, X. (2024). De-identification is not always enough. *arXiv preprint arXiv:2402.00179*.

[24]. Shahid, A., Bazargani, M. H., Banahan, P., Mac Namee, B., Kechadi, T., Treacy, C., ... & MacMahon, P. (2022, April). A two-stage de-identification process for privacy-preserving medical image analysis. In *Healthcare* (Vol. 10, No. 5, p. 755). MDPI.

[25]. Subramanyam, S. V. (2025). Cloud-based enterprise systems: Bridging scalability and security in healthcare and finance. *IJSAT-International Journal on Science and Technology*, *16*(1).

[26]. Liu, Z., Huang, Y., Yu, X., Zhang, L., Wu, Z., Cao, C., ... & Li, X. (2023). Deid-gpt: Zero-shot medical text de-identification by gpt-4. *arXiv preprint arXiv:2303.11032*.

[27]. Radhakrishnan, L., Schenk, G., Muenzen, K., Oskotsky, B., Ashouri Choshali, H., Plunkett, T., ... & Butte, A. J. (2023). A certified de-identification system for all clinical text documents for information extraction at scale. *JAMIA open*, *6*(3), ooad045.

[28]. Li, P., & Zhang, L. (2025). Application of big data technology in enterprise information security management. *Scientific Reports*, *15*(1), 1022.

[29]. Clunie, D. A., Flanders, A., Taylor, A., Erickson, B., Bialecki, B., Brundage, D., ... & Farahani, K. (2025). Report of the Medical Image De-Identification (MIDI) Task Group--Best Practices and Recommendations. *Arxiv*, arXiv-2303.

[30]. Hwang, M. S., Fatima, K., Wang, Y. S., Wu, N. I., & Lin, I. C. (2025). Research on De-identification Applications of LLMs in Medical Records. *International Journal of Network Security*, *27*(1), 213-222.