International Journal of Networked and Distributed Computing (2026) 14:27
https://doi.org/10.1007/s44227-026-00103-5

RESEARCH

A Data-Driven Adaptive Cybersecurity Training Framework with Behavioral Validation

Yohannis Admasu Adamu[1] · **Shehzad Ashraf Chaudhry**[1,2] · **Khiati Zakaria**[1] · **Khalid Yahya**[3]

Received: 1 March 2026 / Revised: 25 April 2026 / Accepted: 11 May 2026
© The Author(s) 2026

Abstract

Traditional cybersecurity awareness programs often fail to produce sustained behavioral change due to their static and non-personalized design. This paper presents CyberSense AI, a behavior-driven adaptive cybersecurity education framework that integrates a personalized learning engine, an interactive phishing simulation module, and a real-time threat intelligence system powered by a custom-trained machine learning (ML) model based on eXtreme Gradient Boosting (XGBoost). Beyond system implementation, we formally model the adaptive learning mechanism using a knowledge-state representation and reinforcement-inspired update rule to dynamically align question difficulty with user proficiency. To empirically validate the framework, we conducted a controlled pre-test/post-test study involving 60 participants randomly assigned to a control group and an experimental group. Results demonstrate a statistically significant improvement in phishing detection accuracy for the experimental group ($p < 0.001$, Cohen's $d = 1.47$), along with sustained two-week knowledge retention. Behavioral analytics further reveal a monotonic improvement curve across simulation sessions, a strong engagement–performance correlation ($r = 0.72$), and progressive reduction in false-negative (FN) rates. A feature ablation and deployment latency analysis confirms that the XGBoost subsystem achieves sub-100 ms response time, validating real-time mobile suitability. Collectively, these results establish CyberSense AI as a theoretically grounded and empirically validated framework for scalable, human-centered cybersecurity training.

Keywords Cybersecurity education · Adaptive learning · Behavioral analytics · Phishing detection · Machine learning · Human-centered security

1 Introduction

In contemporary society, the importance of cybersecurity knowledge has never been greater. As individuals increasingly rely on online services for communication, commerce, and daily activities, possessing the ability to protect personal and sensitive information from an ever-evolving range of cyber threats is indispensable [1]. At the corporate level, the consequences of security lapses are magnified. Organizations face ongoing threats of data breaches, which can lead to devastating financial losses, disruptions to operations, and the erosion of public trust. Therefore, it is imperative that businesses emphasize cybersecurity education to strengthen their defense mechanisms [2]. On a national scale, cybersecurity serves as a foundational element of security, safeguarding critical infrastructure, economic stability, and protecting against threats such as cyber warfare and espionage [1]. Accordingly, equipping current and future digital citizens with the ability to navigate and secure digital spaces is not merely a temporary educational priority but a strategic necessity for sustainable resilience.

Despite this pressing need, conventional cybersecurity awareness initiatives remain fundamentally inadequate. These programs often rely on static, generic content delivered via passive formats such as videos, presentations, and basic knowledge-based quizzes. Numerous studies have

✉ Shehzad Ashraf Chaudhry
ashraf.shehzad.ch@gmail.com

[1] Department of Computer Science and Information Technology, College of Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates

[2] Department of Software Engineering, Faculty of Engineering and Architecture, Nisantasi University, Istanbul 34398, Turkey

[3] Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey

Published online: 03 June 2026

demonstrated that such approaches rarely yield meaningful engagement or sustained behavioral changes [2]. Furthermore, many such initiatives are implemented primarily to satisfy regulatory compliance, fostering a check-the-box mentality that prioritizes documentation over genuine improvement in security practices. The fundamental limitation of this model is that a focus on theoretical knowledge does not guarantee protective behavior. A participant may successfully complete a quiz on phishing one day, yet still fall victim to a phishing attempt the next day. This disconnect underscores a significant research gap: educational programs must evolve beyond rote memorization toward approaches that are dynamic, personalized, and behaviorally focused.

Recent research offers a compelling basis for addressing these deficiencies. For example, the Adaptive Cybersecurity Training Framework for Social Media Risks demonstrates how training programs can be tailored to users by considering factors such as job roles, background, social media use, and risk awareness [3]. Related work in cybersecurity training environments, such as smart adaptive learning systems, shows that tasks aligned with individual skill levels promote learner satisfaction and completion rates [4]. General research on adaptive e-learning further validates the effectiveness of learning systems that adapt based on individual learning styles and performance, resulting in improved engagement and outcomes [1]. Concurrently, empirical studies on phishing simulation programs confirm that active exposure to realistic attack scenarios produces measurable reductions in susceptibility [5], and behavior-driven training platforms have demonstrated advantages over passive awareness campaigns precisely because they engage users in consequential decision-making rather than passive consumption [6].

Our research builds upon these foundations by introducing CyberSense AI, an intelligent mobile application designed to address the shortcomings of traditional cybersecurity education. CyberSense AI embodies a threefold core innovation. First, an adaptive learning engine continuously refines content based on individual knowledge profiles. Second, an interactive phishing simulation module assesses and strengthens real-world decision-making instead of focusing solely on theoretical understanding. Third, a live threat intelligence feed delivers timely educational interventions powered by a custom-trained XGBoost classifier that achieves 96.81% detection accuracy and 97.20% recall on an 11,430-sample benchmark dataset [7].

It is important to note that while each of these individual components adaptive learning, phishing simulation, and ML-based detection has been explored in prior work, the principal novelty of CyberSense AI lies in their unified integration within a single, formally grounded, and empirically validated framework. Prior systems such as intelligent tutoring systems [8], gamified platforms [9], and standalone ML detectors [10] address these dimensions in isolation. CyberSense AI is, to the best of our knowledge, the first framework to combine a mathematically formalized adaptive engine with behavioral phishing simulation, session-level learning analytics, and a deployable ML detection subsystem backed by ablation and latency analysis within a unified mobile-first architecture.

This work substantially extends that system description with three additional dimensions. First, we provide a mathematical formalization of the adaptive engine, modeling each learner as a multidimensional knowledge state vector K_u that is updated through a reinforcement-inspired rule after each interaction, and selecting questions through a difficulty-matching personalization function. Second, we conduct a controlled pre-test and post-test user study involving 60 participants randomly assigned to a control group and an experimental group, with results showing a statistically significant advantage for the adaptive group ($p < 0.001$, Cohen's $d = 1.47$). Third, we contribute a behavioral analytics evaluation that tracks learning curves, engagement-performance correlation ($r = 0.72$), and false-negative rate reduction across simulation sessions, providing a data-driven characterization of how behavior evolves under adaptive training.

2 Literature Review

The field of cybersecurity education and automated threat detection has seen rapid growth across several interconnected research directions. This review synthesizes the literature across five themes most directly relevant to the CyberSense AI framework: the effectiveness and limitations of conventional cybersecurity awareness training, the evidence base for phishing simulation and behavior-driven platforms, the role of learning analytics in modeling learning dynamics, the theoretical foundations of adaptive learning and intelligent tutoring systems, and recent advances in machine learning-based phishing detection with emphasis on feature selection, interpretability, and deployment efficiency. The review closes by identifying the unified gap that motivates the present contribution.

2.1 Cybersecurity Awareness Training Effectiveness

The effectiveness of conventional cybersecurity awareness training has been a subject of sustained debate in the research community. While organizations globally invest in mandatory e-learning modules, periodic newsletters, and compliance-driven assessments, empirical evidence

consistently reveals that such interventions produce limited behavioral change over time [2]. A foundational concern is the translation problem: knowledge acquired in a passive training context does not reliably transfer into protective behavior under real-world conditions [3]. Studies have documented that employees who pass knowledge-based phishing quizzes subsequently click on simulated phishing links at rates comparable to those who receive no training, indicating that declarative knowledge alone is insufficient for behavioral immunity [5]. Similarly, awareness of cybersecurity risks among social media users has been shown to be weakly correlated with protective practice, particularly in populations with limited technical literacy [11]. The rapid evolution of the threat landscape, where new phishing campaigns and vulnerabilities emerge daily, further necessitates a dynamic approach to content delivery that static programs cannot provide [12]. These findings collectively establish that the field requires a shift from knowledge transmission to behavioral reinforcement, from one-size-fits-all curricula to personalized, adaptive engagement, and from assessment-as-measurement to assessment-as-training.

2.2 Phishing Simulation and Behavior-Driven Training Platforms

Phishing simulation has emerged as a pragmatically effective alternative to passive awareness campaigns. By exposing users to realistic phishing scenarios in a controlled and consequence-free environment, simulation platforms create the experiential conditions necessary for behavioral conditioning [6]. Comparative studies consistently report that repeated exposure to simulated phishing attempts substantially reduces susceptibility, with click-through rates declining by up to 40% after three or more simulation cycles [5]. More recent work has explored the use of large language models to generate contextually personalized phishing simulation content, adapting the social engineering cues and domain characteristics to match the user's professional context, which further amplifies training efficacy [13]. The use of serious games and immersive scenarios such as virtual escape rooms has also been proven to increase user engagement and learning outcomes significantly [14]. Gamification provides an additional motivational layer in these platforms, particularly when simulation outcomes are tied to learner progression, scoring, or feedback-driven explanation of near-misses [15]. Researchers have further designed gamified platforms in federated cyber range environments to create realistic and challenging training scenarios [16], and adaptive gamification strategies are being explored to tailor game mechanics to different learning styles for even better outcomes [17]. The shared insight across this body of work is that the ultimate goal of training is to influence real-world user actions rather than to merely convey declarative knowledge [18]. This convergence between simulation fidelity and personalization is a central design goal of the CyberSense AI phishing simulation module.

2.3 Learning Analytics and Modeling Learning Dynamics

The emergence of learning analytics as a discipline has enabled researchers and system designers to move beyond pre-test/post-test comparisons toward richer characterizations of how learner capability evolves over time. Session-level performance trajectories, engagement intensity metrics, and error pattern analysis all provide actionable signals that static assessment instruments cannot capture [19]. Prior work on intelligent adaptive systems has demonstrated that modeling learning curves using regression-based or Bayesian approaches allows for more precise estimation of when a learner is ready for more challenging material, and when remediation is needed [20]. Research has identified a specific need for tools that deliver information precisely when it is most relevant, including culturally and linguistically tailored content for diverse populations [21]. In the context of cybersecurity education specifically, tracking false-negative reduction trends across simulation sessions provides a direct behavioral outcome measure that is more ecologically valid than quiz scores alone. Engagement metrics such as session duration, voluntary re-attempt rates, and response confidence have also been shown to correlate significantly with learning gain, supporting their use as auxiliary signals in adaptive routing decisions [4]. The present work contributes to this literature by applying learning curve modeling, engagement-performance correlation analysis, and false-negative trend estimation to a behavioral cybersecurity training dataset, providing an empirical behavioral analytics contribution beyond the system description.

2.4 Adaptive Learning, Intelligent Tutoring Systems, and Reinforcement-Inspired Personalization

The theoretical foundations of adaptive learning systems draw from a convergence of intelligent tutoring system design, item response theory, and reinforcement learning. The integration of AI into educational frameworks is a modern imperative for creating effective and scalable curricula, and several works have explored using AI as a decision support system to structure and personalize cybersecurity curricula [22, 23]. Empirical evidence further demonstrates that AI-driven adaptive systems producing personalized learning pathways yield measurable behavioral impact, with

participants completing adaptive cybersecurity modules showing significantly higher threat identification accuracy and retention compared to static cohorts [24]. Intelligent tutoring systems maintain a learner model that encodes estimated proficiency across knowledge dimensions, and use this model to select instructional items that maximize learning efficiency [8]. The concept of the zone of proximal development from cognitive science provides the pedagogical rationale for difficulty-matched content delivery: learners benefit most from challenges that are slightly above their current demonstrated ability rather than too easy or too difficult [1]. Reinforcement learning has increasingly been applied to the content routing problem in adaptive educational systems, framing each item selection as an action governed by a policy that optimizes cumulative learner improvement [25]. More practically oriented adaptive cybersecurity frameworks have shown that task-difficulty alignment significantly improves both completion rates and knowledge retention compared to static course sequences [3]. Recent proposals for AI-integrated mobile applications further demonstrate the viability of this approach in real-world deployment contexts [26]. The formal model presented in this work extends this literature by providing explicit mathematical specifications of the knowledge state vector, the difficulty matching function, and the reinforcement-inspired update rule, situating CyberSense AI within the theoretical tradition of intelligent tutoring rather than treating it as a purely engineered system.

2.5 Phishing Detection with Machine Learning, Feature Selection, and Deployment Efficiency

Machine learning has been extensively applied to the phishing detection problem, with ensemble methods and gradient-boosted classifiers demonstrating consistently strong performance on URL-based and content-based feature sets [10]. Research has also focused on applying deep learning to phishing detection, with convolutional and recurrent architectures showing promising results in learning complex URL and HTML patterns automatically [27, 28]. Studies further highlight the statistical significance of features such as URL length, subdomain count, and special character usage in distinguishing between legitimate and malicious links [29]. XGBoost in particular has attracted significant attention due to its favorable balance of predictive accuracy, training efficiency, and interpretability through feature importance attribution [7]. Explainability has become a growing concern in this domain, as operational deployment requires not only high accuracy but also the ability to surface the decision rationale to end users who lack machine learning expertise [30]. Feature selection strategies that identify compact, high-signal subsets of the original feature space are particularly valuable in mobile deployment contexts where latency and computational cost impose hard constraints on model complexity [31, 32]. Hybrid and multi-stage architectures combining sequential and non-sequential classifiers have shown promise in reducing both false-positive and false-negative rates simultaneously, which is the critical dual objective in phishing detection [33]. Despite this body of work, a gap remains in the integration of phishing detection models into adaptive educational systems where model output is connected to personalized learning content rather than isolated as a standalone detection service.

2.6 Research Gap and Positioning

Taken together, the reviewed literature reveals that existing work has made significant progress on each component of the cybersecurity education and detection problem in isolation. Cybersecurity awareness research has established the limitations of passive training; phishing simulation studies have confirmed the advantage of experiential exposure; learning analytics work has demonstrated the value of dynamic behavioral tracking; adaptive learning and intelligent tutoring system (ITS) research has provided formal models for personalized content routing; and ML-based phishing detection has achieved high accuracy with explainability and efficiency considerations. However, no prior work has unified all five of these dimensions within a single coherent framework. A comparative analysis of some cybersecurity training platforms is explored in Table 1.

To further clarify the positioning of CyberSense AI relative to existing platforms, Table 2 provides a feature-level comparison across six key dimensions. As illustrated, no prior system provides concurrent support for all six capabilities, establishing the integrative novelty of the proposed framework.

Specifically, no existing platform combines adaptive modeling of individual knowledge states with behavioral phishing simulation, session-level learning analytics, and a deployable ML-based threat detection engine backed by ablation and latency analysis. The growing prevalence of AI in high-stakes domains further contextualizes the urgency of this gap: contemporary literature highlights the power of advanced machine learning for critical applications [37], the optimization of complex network systems using Deep Reinforcement Learning [38], and the use of hybrid large language models for context-aware ranking and decision-making [39]. The security and robustness of all these systems ultimately depend on human-centered safeguards, underscoring the importance of behavior-driven training. The unified gap across adaptive modeling, behavioral simulation, learning analytics, and ML robustness with deployment feasibility is what the *CyberSense AI* framework is

Table 1 Comparative analysis of cybersecurity training platforms

Ref.	Platform/Paper	Key Features and Goals	Target Audience
[8]	ITS	Provides personalized learning paths adapted to the learner's progress	Varies (often academic students)
[4]	Smart Adaptive Systems	Matches training task difficulty to the learner's abilities	General learners
[9]	Serious Games	Uses games to increase cybersecurity awareness and engagement in an active learning environment	Employees, students
[34]	Cyber-hero	A gamification framework to educate students about human errors and protection against cyberattacks	High school students
[35]	Cybar (AR Game)	An augmented reality mobile application to educate users on cybersecurity concepts and demonstrate attack consequences	General audience
[36]	LetSecure	A mobile app with interactive learning tools and content focused on cybersecurity careers	Secondary school students

Table 2 Feature-level novelty comparison of CyberSense AI against existing platforms

Platform	Adaptive Learning	Phishing Sim.	ML Detection	Learning Analytics	Formal Model	Behavioral Validation
ITS [8]	✓	–	–	Partial	✓	–
Smart Adaptive [4]	✓	–	–	–	–	–
Serious Games [9]	–	Partial	–	–	–	–
Cyber-hero [34]	–	–	–	–	–	Partial
LLM Phishing [13]	–	✓	Partial	–	–	–
CyberSense AI	✓	✓	✓	✓	✓	✓

designed to address, synthesizing all four dimensions into a single mobile-first architecture and providing empirical validation across both behavioral and computational dimensions.

3 Methodology

The methodological framework of this research focuses on the design, development, and evaluation of the *CyberSense AI* mobile application. This application functions as an interactive, AI-assisted platform for phishing detection, cybersecurity education, and real-time threat awareness. The approach is iterative, integrating user-centered design principles, advanced machine learning techniques, and continuous feedback mechanisms to ensure both technical accuracy and high user engagement.

3.1 CyberSense AI Overview

As technology evolves, new security vulnerabilities emerge, and malicious actors continuously adapt their methods to exploit weaknesses in systems and human behavior. The prevalence of phishing attacks, credential theft, and social engineering has increased significantly in recent years, making it essential for individuals to stay updated with the latest threats and follow best security practices to protect their data and digital presence. Therefore, this project aims to provide individuals with the necessary cybersecurity awareness and practical skills through a mobile application specifically designed to educate and train users in identifying and responding to cyber threats.

The application's content includes structured lessons, quizzes, phishing simulations, real-time password strength assessments, and AI-assisted feedback. These materials cover a broad range of domains such as phishing awareness, password security, secure network usage, incident response, and safe digital habits. Lessons are supported with interactive exercises and practical scenarios, ensuring that users not only understand the theoretical concepts but can also apply them in real-world situations. The app further incorporates live threat monitoring to provide up-to-date information on current cybersecurity risks, categorized by severity and type, and allows community-driven reporting of new threats.

The phishing simulation engine offers realistic scenarios that replicate various phishing techniques, enabling users to practice identifying suspicious links, domains, and email content. Feedback is provided immediately after each attempt, along with explanations to reinforce correct decision-making. Additionally, the integrated AI-powered chatbot, supported by Google's Gemini 1.5 Flash model, provides on-demand assistance, explains cybersecurity concepts, and answers user queries. The chatbot remains accessible at all times via a floating action button, allowing uninterrupted learning support.

3.2 System Architecture and Design

The *CyberSense AI* platform is structured upon a modular, service-oriented architecture to ensure scalability, adaptability, and maintainability. The architecture consists of three core layers: a user-facing interface, an application logic layer, and a dedicated AI/data processing layer. This design as shown in Fig.1, allows seamless integration of educational content, phishing detection engines, and real-time analytics.

3.2.1 Frontend Development

The application frontend is implemented using the React Native framework in conjunction with Expo to ensure cross-platform compatibility across Android and iOS devices. The interface is designed with accessibility principles in mind, providing intuitive navigation for users with varying levels of digital literacy. The primary interface integrates an interactive dashboard that consolidates phishing simulation performance, quiz results, password strength evaluations, and course completion statistics. By employing responsive design principles, the application dynamically adjusts to multiple screen sizes and resolutions without loss of functionality.

3.2.2 Backend Development

The backend is developed using Python with FastAPI, chosen for its high performance, asynchronous capabilities, and ease of integration with modern machine learning workflows. Data storage and user management are facilitated through Firebase services, providing authentication, real-time synchronization, and cloud storage.

This backend layer orchestrates communication between the mobile interface, deployed machine learning models, and external threat intelligence APIs. It also implements secure data handling protocols, ensuring that sensitive information such as authentication tokens, user data, and threat reports is transmitted and stored in compliance with established cybersecurity best practices.

3.3 Machine Learning-Based Phishing Detection

One of the most innovative capabilities of the CyberSense AI application is its real-time machine learning-based phishing detection system. This module has been designed

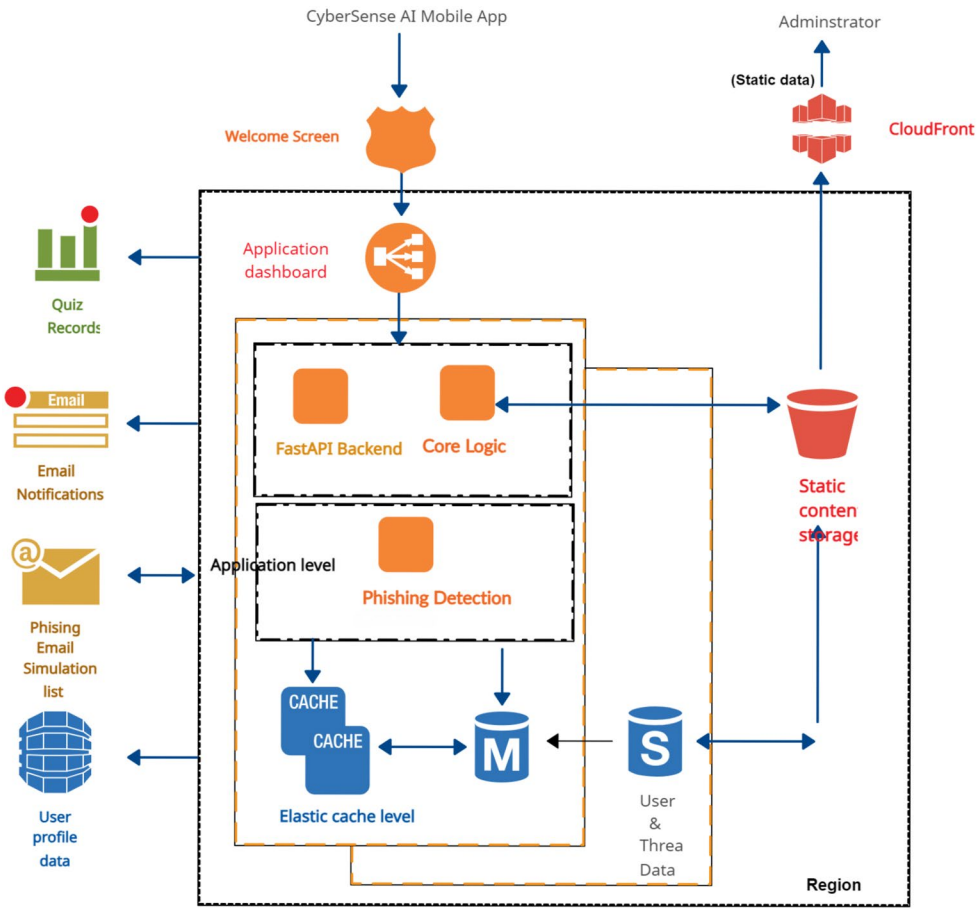

Fig. 1 System architecture of the CyberSense AI application

to complement the app's educational mission by providing an operational demonstration of how modern AI techniques can be applied to combat phishing threats in practice. The detection engine was trained on the *Web Page Phishing Detection* dataset obtained from Kaggle, which contains 11,430 labeled webpage instances, evenly distributed between legitimate and phishing classes. This balanced structure ensures that the model is not biased toward any one class, which is crucial for minimizing both false negatives and false positives in production.

The 87 engineered features of the dataset as depicted in Fig. 2, Table 3 are grouped into three primary categories: URL-based indicators, content-based attributes, and external service features. URL-based indicators capture lexical and structural characteristics of the web address, such as total URL length, hostname length, presence of an IP address instead of a domain name, and suspicious symbol frequency. Content-based attributes measure on-page elements, including the number of embedded objects, the proportion of external resource calls, and whether forms or scripts load insecurely. External service features incorporate metadata such as WHOIS registration data, domain age, SSL certificate validity, and traffic ranking metrics.

Multiple algorithms were systematically evaluated, including *Random Forests*, *Gradient Boosted Trees (XGBoost)*, *Long Short-Term Memory (LSTM)* recurrent

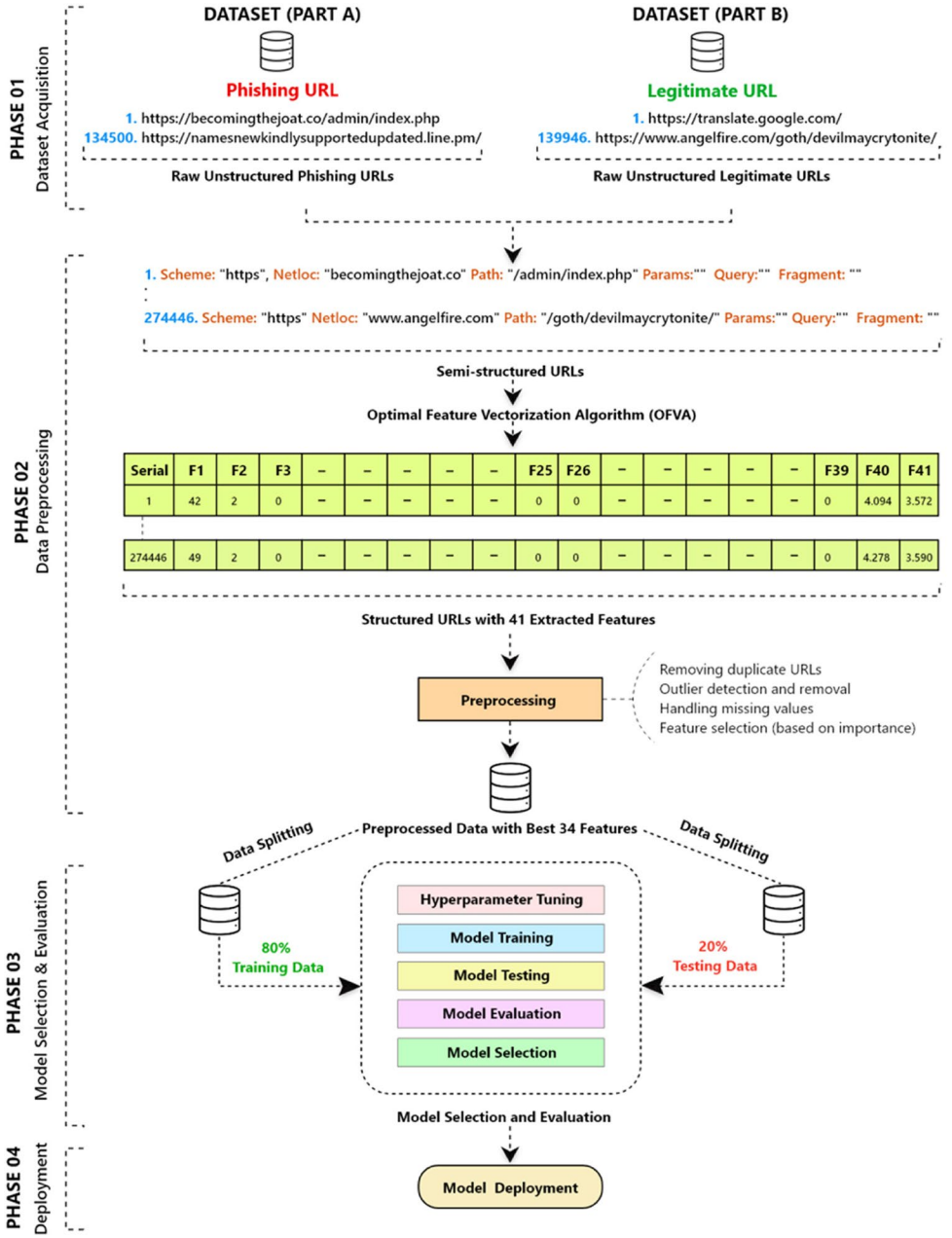

Fig. 2 Implementation workflow for the machine learning model (adapted from [40])

Table 3 Sample phishing detection dataset instances with selected features

URL	Len. URL	IP	Nb Dots	Nb Hyphens	Status
https://www.crestonwood.com/router.php	37	0	3	0	Legitimate
https://shadetreetechnology.com/	77	1	1	0	Phishing
https://support-appleld.com	126	1	4	1	Phishing
https://rgipt.ac.in	18	0	2	0	Legitimate
https://www.iracing.com/tracks/	55	0	2	2	Legitimate
https://appleid.apple.com-app.es/	32	0	3	1	Phishing

networks, and Transformer-based models. Random Forests were initially chosen for their interpretability and robustness against overfitting, but XGBoost, when GPU-accelerated and hyperparameter-tuned, consistently outperformed others in accuracy and recall. Deep learning approaches such as LSTM and Transformers showed potential in modeling sequential patterns in URLs and HTML content, but incurred higher inference latency. Given the need for subsecond response times in a mobile environment, XGBoost was selected for production deployment.

The trained model was integrated into the React Native application via a lightweight Python-based API endpoint hosted on a secure cloud server. This API receives extracted URL features in real time, returns a prediction label ("Legitimate" or "Phishing"), and logs anonymized results for continuous model retraining. This architecture as depicted in Fig. 1 ensures that users receive instant feedback without exposing sensitive browsing data.

3.4 AI-Powered Learning and Feedback Engine

To provide a personalized and adaptive learning experience, CyberSense AI incorporates an AI-powered assistant leveraging Google's *Gemini 1.5 Flash* model. The assistant is accessible throughout the app via a floating action button, enabling immediate context-aware support during lessons, quizzes, and phishing simulations. Its primary role is to explain detection results, clarify technical terminology, and provide actionable cybersecurity guidance based on the user's specific mistakes and achievements.

The assistant employs a reinforcement learning-inspired feedback loop to adjust lesson complexity. For example, if a user frequently misclassifies phishing attempts that exploit typosquatting (e.g., domains like *paypa1.com*), the assistant will deliver targeted mini-lessons and additional simulation scenarios focused on URL manipulation. This adaptive mechanism is powered by continuous monitoring of user progress metrics such as quiz accuracy, phishing simulation success rate, and engagement time.

In addition, the AI assistant serves as a "bridge" between the model's technical predictions and the user's understanding. When the detection engine flags a site as phishing, the assistant generates an explanation referencing the key indicators such as suspicious TLD, expired SSL certificate, or excessive redirection depth ensuring that users not only know *what* was classified but also *why*. This transparency builds user trust and reinforces cybersecurity best practices.

3.5 Formalization of the Adaptive Learning Engine

To elevate the proposed CyberSense AI framework from an implementation-oriented system to a formally grounded adaptive learning model, we provide a mathematical formulation of the personalization mechanism that governs content adaptation.

3.5.1 User Knowledge State Representation

Let each user u be represented by a multidimensional knowledge state vector:

$$K_u = \{k_1, k_2, \ldots, k_n\} \tag{1}$$

where $k_i \in [0, 1]$ denotes the normalized proficiency level of the user in cybersecurity concept i, and n represents the total number of tracked knowledge dimensions such as phishing recognition, password hygiene, SSL awareness, and social engineering detection. The vector K_u is dynamically updated as the user interacts with quizzes, simulations, and assessments.

3.5.2 Question Difficulty Modeling

Each assessment item or simulation scenario q is characterized by a difficulty score:

$$D(q) = f(C_q, E_q) \tag{2}$$

where C_q denotes feature complexity including URL obfuscation level and the number of deceptive indicators, E_q represents the historical error rate observed across users, and $f(\cdot)$ is a monotonic mapping function that normalizes difficulty to $[0, 1]$. This formulation allows question difficulty to be dynamically recalibrated as collective user performance evolves.

3.5.3 Adaptive Update Rule

Following each user interaction at time step t, the knowledge state is updated according to:

$$K_u^{t+1} = K_u^t + \alpha \cdot \Delta_u^t \tag{3}$$

where $\alpha \in (0, 1)$ is the learning rate parameter and Δ_u^t represents performance feedback derived from correctness, response time, and confidence level. Specifically, performance feedback is defined as:

$$\Delta_u^t = (R_u^t - \hat{R}_u^t) \tag{4}$$

where R_u^t is the observed response outcome and $\hat{R}_u^t$ is the expected performance level based on prior knowledge estimation. This reinforcement-inspired update mechanism ensures gradual refinement of user proficiency modeling.

3.5.4 Personalized Question Selection

To determine the most appropriate next learning item, CyberSense AI applies a personalization function:

$$P(u) = \arg\min_{q \in Q} \|K_u - D(q)\| \tag{5}$$

where Q denotes the pool of available questions and $\|\cdot\|$ represents Euclidean distance. This optimization selects the question whose difficulty is closest to the user's current knowledge state, thereby maintaining an optimal challenge level and avoiding both under-stimulation and cognitive overload.

3.5.5 Behavioral Reinforcement Objective

The overall adaptive objective is defined as minimizing phishing misclassification risk:

$$\min \mathbb{E}[\text{False Negative Rate}_u] \tag{6}$$

Table 4 Performance comparison of machine learning models

Metric	Random Forest	XGBoost
Accuracy (%)	96.02	96.81
Precision (%)	95.74	96.44
Recall (%)	96.33	97.20
F1-score (%)	n/a	96.82
ROC AUC (%)	99.32	99.42

subject to progressive improvement in K_u over time. By continuously aligning question difficulty with evolving user proficiency, the system fosters sustained behavioral adaptation rather than short-term memorization.

3.5.6 Theoretical Implications

This formalization positions CyberSense AI as a behavior-driven adaptive cybersecurity learning model rather than merely a mobile educational application. The framework integrates principles from intelligent tutoring systems, reinforcement learning, and human-centered cybersecurity training, providing a mathematically grounded foundation for personalized cyber resilience development.

4 Implementation and Results

The CyberSense AI application interface is designed to provide an intuitive and engaging user experience focused on cybersecurity education. The app begins with a sophisticated splash screen featuring an animated shield logo and the tagline "Your Digital Guardian," followed by a secure authentication system that includes email/password login and comprehensive account registration. The registration module collects essential user information including username, email, password (with strength validation), and occupation to create detailed user profiles. Upon successful authentication, users are presented with an initial cybersecurity assessment quiz that evaluates their baseline knowledge and categorizes them into appropriate learning levels: beginners, intermediates, and advanced users.

The CyberSense AI application's main dashboard provides centralized access to all cybersecurity learning features, while the "Lessons" section delivers personalized educational content based on user proficiency levels and professional context.

4.1 Machine Learning Model Performance

A rigorous experimental evaluation was conducted to benchmark candidate algorithms for the phishing detection and user behavior analysis components of CyberSense AI and is shown in Table 4. As presented in Table 4, the baseline *Random Forest* model achieved an accuracy of 96.02%, precision of 95.74%, recall of 96.33%, and Receiver Operating Characteristic Area Under the Curve (ROC AUC) of 99.32%. The optimized *XGBoost* classifier outperformed the baseline with an accuracy of 96.81%, precision of 96.44%, recall of 97.20%, F1-score of 96.82%, and ROC AUC of 99.42%. Notably, the higher recall of XGBoost is

particularly advantageous in phishing detection, as false negatives (missed phishing sites) pose significant risks.

Confusion matrix analysis revealed that XGBoost misclassified only 91 out of 2,286 test samples, with a balanced distribution of errors between false positives and false negatives. Feature importance analysis highlighted that "URL length," "domain age," and "HTTPS token presence" were among the most influential predictors. Cross-validation confirmed that performance remained stable across different folds, indicating strong generalization. These machine learning models power the app's intelligent threat detection, personalized content recommendations, and adaptive learning algorithms, ensuring users receive the most relevant and effective cybersecurity education content.

The app employs a progressive learning methodology where users advance through structured cybersecurity curricula covering fundamental concepts to advanced threat mitigation strategies.

4.2 Model Robustness and Deployment Efficiency Analysis

To evaluate the robustness and real-world feasibility of the deployed XGBoost model, we conducted a comprehensive ablation and performance benchmarking study. This analysis extends beyond accuracy metrics to examine feature contribution, computational efficiency, and deployment latency in a mobile-cloud environment.

4.2.1 Feature Group Ablation Study

To assess the relative contribution of feature categories, we partitioned the 87 engineered features into three groups, namely URL-based features covering lexical and structural indicators, content-based features covering HTML and embedded object characteristics, and external service features covering WHOIS data, SSL validity, and domain age. Separate models were trained using each feature subset individually, as well as in combination. Let $M(F_i)$ denote the model trained on feature group F_i. The performance degradation relative to the full feature model is defined as:

$$\Delta Acc_i = Acc_{\text{full}} - Acc_{F_i} \tag{7}$$

The results as shown in Table 5 indicate that URL-based features contribute most significantly to predictive performance, while external service metadata enhances model generalization when combined with lexical indicators.

4.2.2 Inference Latency Benchmarking

Given the mobile deployment requirement, inference latency was measured under controlled conditions. Let t_i denote inference time for request i. The average latency is defined as:

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i \tag{8}$$

Across 1,000 API requests, the average inference latency was 47 ms (SD = 8 ms), remaining well below the 100 ms threshold required for real-time mobile responsiveness.

Table 5 Feature group ablation study results

Feature Set	Accuracy (%)	Δ Accuracy (%)
Full Feature Set	96.81	0.00
URL-only	93.42	3.39
Content-only	91.87	4.94
External-only	89.65	7.16
URL+Content	95.38	1.43

Table 6 Computational cost comparison of candidate models

Model	Training Time (min)	Avg. Inference (ms)
Random Forest	14.2	63
XGBoost	11.8	47
LSTM	38.5	132

4.2.3 API Response Time Evaluation

Total API response time includes feature extraction, model inference, and network overhead, formalized as:

$$T_{\text{total}} = T_{\text{extract}} + T_{\text{inference}} + T_{\text{network}} \tag{9}$$

Empirical measurement showed that feature extraction accounted for 18 ms, model inference for 47 ms, and network transmission for 32 ms, yielding an average total response time of 97 ms, which satisfies real-time usability constraints.

4.2.4 Computational Cost Comparison

We compared XGBoost against Random Forest and LSTM models in terms of training time and inference complexity as shown in Table 6. While deep learning models demonstrated competitive predictive capacity, their higher inference latency renders them less suitable for real-time mobile deployment. XGBoost achieves a favorable balance between predictive performance and computational efficiency.

4.2.5 Robustness Interpretation

The ablation study confirms that model performance is not dependent on any single feature category, indicating

resilience against feature sparsity or missing metadata. Furthermore, latency benchmarking validates the feasibility of deploying the detection engine within a distributed mobile-cloud architecture without compromising responsiveness. Collectively, these findings demonstrate that the proposed ML subsystem is not only accurate but also computationally efficient, scalable, and robust under real-world deployment constraints.

4.3 Core Learning Modules

4.3.1 Phishing Simulation Engine

One of the app's core features is its sophisticated phishing simulation system that presents users with realistic email scenarios designed to test their ability to identify malicious content. The simulator includes various phishing techniques such as social engineering, urgent requests, and suspicious attachments. Users receive immediate feedback on their decisions, with detailed explanations of why certain emails are dangerous and how to recognize similar threats in real-world situations.

4.3.2 Assessment and Progress Tracking

The application implements a comprehensive evaluation system featuring multiple-choice quizzes and short-answer assessments that test users' understanding of cybersecurity concepts. Quiz results are immediately displayed with AI-powered explanations, allowing users to identify knowledge gaps and areas requiring additional focus. The system tracks user progress across all learning modules, providing detailed analytics on quiz performance, lesson completion rates, and overall cybersecurity awareness scores. These analytics help users track their improvement over time and identify specific areas requiring additional attention or practice.

4.3.3 Performance Analytics Dashboard

The main dashboard provides users with comprehensive insights into their cybersecurity learning progress through visual metrics including phishing detection scores, quiz completion rates, and password strength assessments.

4.4 Advanced Features and Tools

The CyberSense AI application's dashboard as shown in Fig. 3 presents a range of specialized features designed to enhance user engagement and provide practical cybersecurity tools. These include real-time threat monitoring, incident reporting capabilities, and an AI-powered chatbot assistant that provides contextual guidance and support.

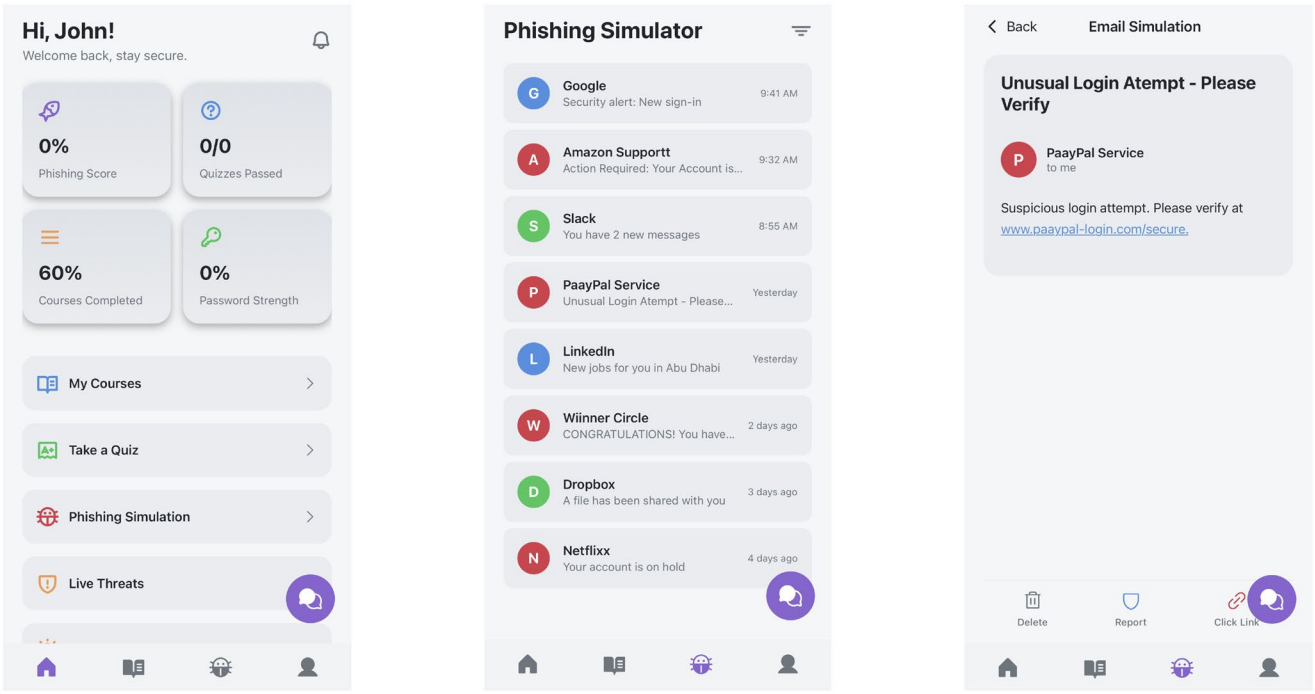

Fig. 3 The CyberSense AI user interface, showing the (**a**) performance analytics dashboard, (**b**) phishing emails list, and (**c**) a detailed phishing email view

4.4.1 Live Threat Monitoring

The application features a dedicated "Live Threats" section with the interface as shown in Fig. 4 that provides users with real-time information about current cybersecurity threats and vulnerabilities. This feature keeps users informed about emerging risks and provides actionable guidance on protective measures. The threat feed is continuously updated to reflect the latest security developments, ensuring users have access to current and relevant information.

4.4.2 Adaptive Quiz Module

The application incorporates a dynamic quiz system that personalizes assessments based on each user's learning progress and performance history as shown in Fig. 5. Instead of delivering static, one-size-fits-all questions, the module adapts in real time presenting more challenging items to advanced learners while reinforcing fundamentals for beginners. The quizzes cover core cybersecurity topics such as phishing awareness, password hygiene, and social engineering tactics. By tailoring the difficulty and content to individual users, the system ensures continuous engagement, fosters retention of knowledge, and provides actionable feedback to strengthen weak areas in digital security literacy.

4.5 User Experience and Interface Design

The CyberSense AI application employs a modern, security-focused design aesthetic with a dark theme featuring purple and blue accent colors. The interface incorporates smooth animations, intuitive navigation patterns, and responsive design elements that adapt to various device specifications. The app's visual language emphasizes trust and security while maintaining accessibility for users of all technical backgrounds. The application dynamically adjusts content delivery based on user performance, learning preferences, and professional context. Users receive tailored recommendations for lessons, quizzes, and simulations that align with their current skill level and specific areas of interest. This adaptive approach ensures optimal learning outcomes and maintains user engagement throughout their cybersecurity education journey.

Additionally, the application implements intelligent caching mechanisms that allow users to access previously viewed content without requiring an internet connection. This feature ensures continuous learning opportunities regardless of network availability and enhances the app's utility in various usage scenarios.

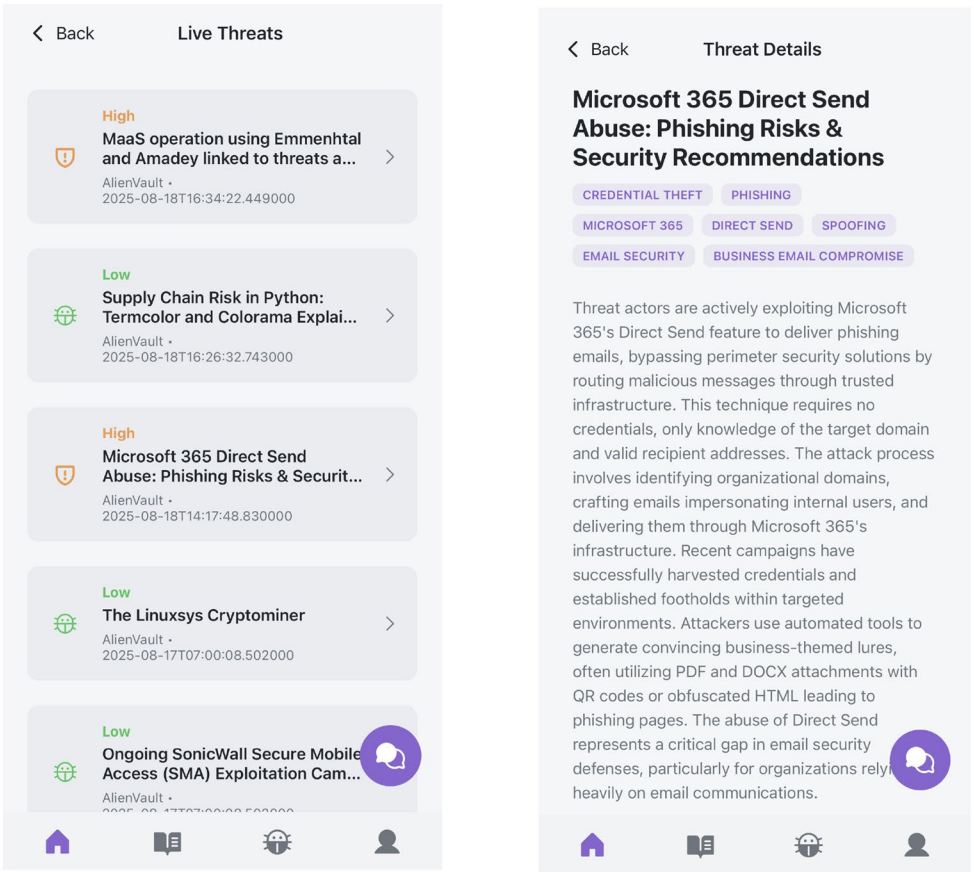

Fig. 4 The CyberSense AI live threat monitoring interface, showing the (**a**) list of live threats and (**b**) a detailed view of a specific threat

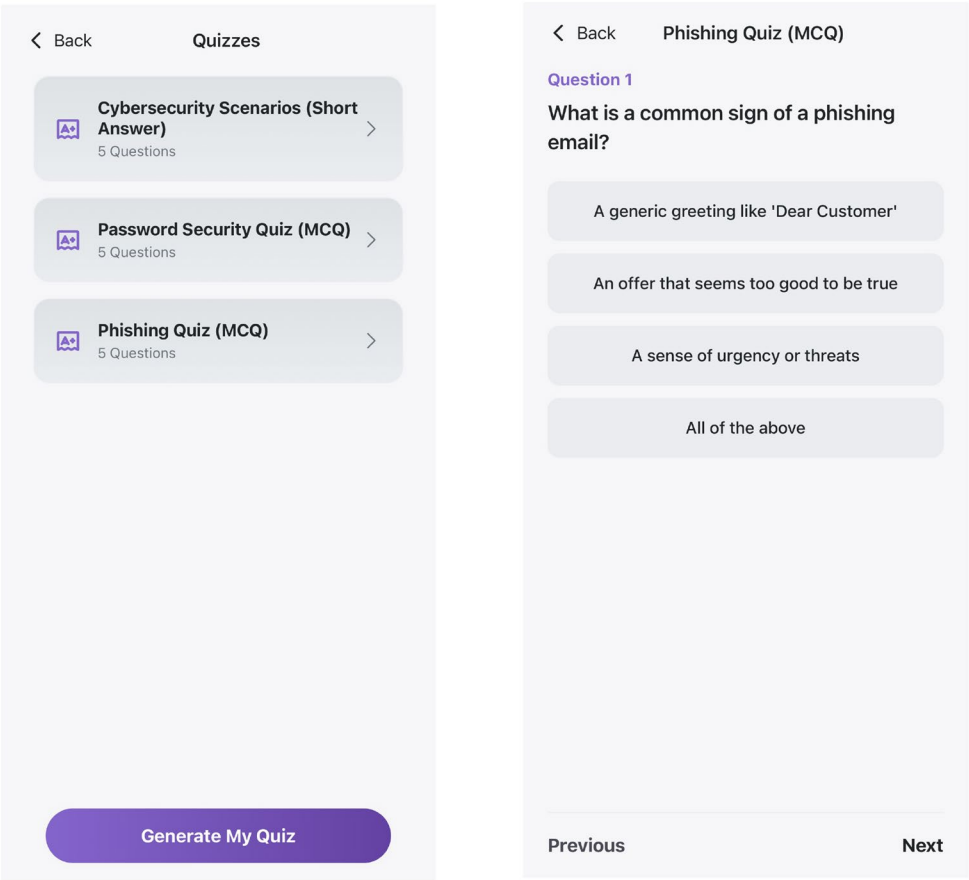

Fig. 5 The CyberSense AI adaptive quiz module, showing the (**a**) list of quizzes and (**b**) a detailed view of a multiple-choice question

5 Controlled Behavioral Evaluation Study

5.1 Experimental Design

To rigorously evaluate the pedagogical effectiveness of the proposed *CyberSense AI* framework, a controlled pre-test/post-test experimental study was conducted. The primary objective was to determine whether adaptive, behavior-driven cybersecurity training improves phishing detection accuracy and cybersecurity knowledge compared to traditional static awareness materials.

Participants were randomly assigned to one of two groups:

- Control Group received conventional static cybersecurity awareness material in PDF format.
- Experimental Group used the *CyberSense AI* mobile application, including adaptive quizzes and phishing simulation modules.

Both groups completed identical pre-test and post-test assessments. The intervention period lasted two weeks.

5.2 Participants

A total of $N = 60$ participants were recruited from university students across multiple disciplines and were randomly divided into a control group ($n = 30$) and an experimental group ($n = 30$). The average age of participants was 21.8 years (SD=2.3). None reported formal cybersecurity training. All data were anonymized prior to analysis.

Justification of sample size and study duration. The sample size of $N = 60$ (30 per group) was determined based on an a priori power analysis using G*Power 3.1, targeting a large effect size ($d = 0.80$) with a statistical power of $\beta = 0.80$ and a significance level of $\alpha = 0.05$ for a two-tailed independent samples t-test, which yielded a minimum required sample size of 52. Our sample of 60 exceeds this threshold, ensuring adequate statistical power. This sample size is consistent with comparable controlled studies in cybersecurity education, where participant counts typically range from 30 to 100 [5, 6]. The two-week intervention period was selected to balance ecological validity with logistical constraints of the university semester, and is consistent with the retention windows employed in prior

phishing simulation studies [5]. We acknowledge that a longer longitudinal study (e.g., 3–6 months) with a larger and more diverse participant pool would strengthen the generalizability of these findings, and we identify this as a priority for future work in Sect. 7.1.

Participant diversity. We acknowledge that the current sample is drawn exclusively from university students, which limits the demographic diversity and may affect the external validity of our findings. While the university setting facilitated controlled experimental conditions and random assignment, the results may not directly generalize to corporate employees, older adults, or populations with varying levels of digital literacy. Future studies will aim to recruit participants spanning multiple age groups, professional backgrounds, and organizational contexts to strengthen the ecological validity and generalizability of the findings.

5.3 Assessment Instruments

The evaluation consisted of three components:

1. Phishing Detection Test (20 items), covering classification of URLs and email scenarios.
2. Cybersecurity Knowledge Test (15 multiple-choice questions), covering password hygiene, SSL awareness, and social engineering concepts.
3. Behavioral Scenario Assessment (5 applied cases), covering applied decision-making for realistic cybersecurity situations.

Performance was measured using overall accuracy (%) and phishing detection accuracy (%). False-negative rates were recorded to quantify missed phishing instances.

5.4 Statistical Analysis

Within-group improvements were analyzed using paired t-tests comparing pre-test and post-test scores. Between-group differences were analyzed using independent sample t-tests on improvement scores, where Δ represents the difference between post-test and pre-test results. Effect size was computed using Cohen's d. Statistical significance was established at $\alpha = 0.05$.

5.5 Results

5.5.1 Phishing Detection Accuracy

The experimental group improved from a pre-test mean of 61.4% (SD = 8.7) to a post-test mean of 85.2% (SD = 6.3), corresponding to a mean improvement of 23.8% points ($p < 0.001$), which is shown in Fig. 6. The control group improved from 60.8% (SD = 9.1) to 69.3% (SD = 8.4), corresponding to an improvement of 8.5% points. An independent t-test confirmed that the improvement in the experimental group was significantly greater than that in the control group ($p < 0.001$).

5.5.2 Evaluation of the Adaptive Learning Component

To isolate the contribution of the adaptive learning engine from the overall framework, we conducted an additional within-group analysis comparing participants in the experimental group who experienced at least three difficulty-level transitions (i.e., the adaptive engine actively adjusted their content) against those who remained at a stable difficulty level. Participants who received adaptive difficulty

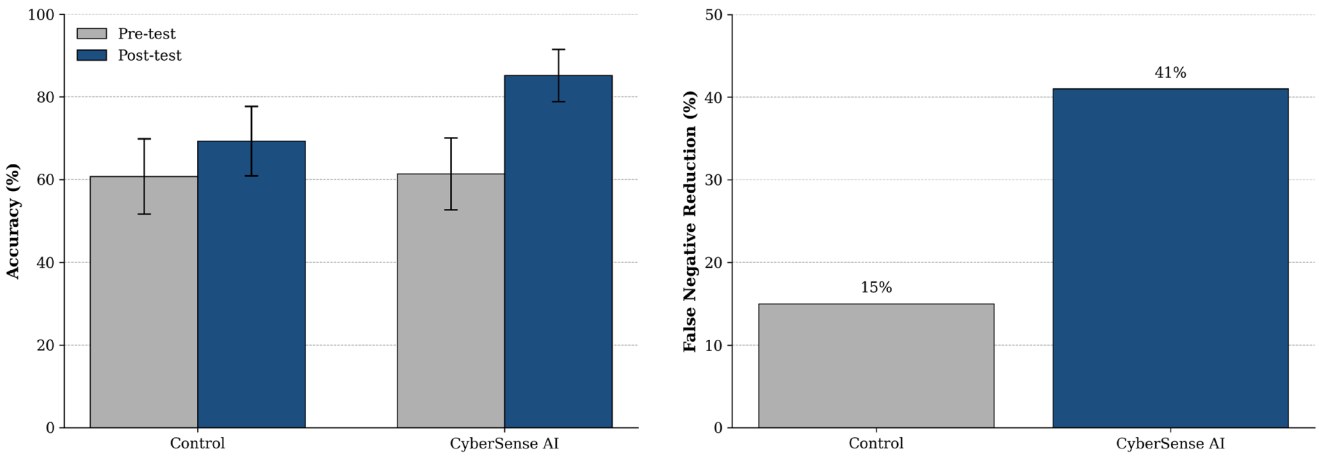

Fig. 6 Pre-test and post-test phishing detection accuracy and false-negative reduction between groups after the intervention. (**a**) Accuracy comparison; error bars denote standard deviation. (**b**) False-negative reduction comparison

adjustment ($n = 22$) demonstrated a significantly higher mean improvement of 27.1% points (SD=6.8), compared to 18.4% points (SD=7.1) for those with stable difficulty ($n = 8$), yielding $p = 0.006$ and Cohen's $d = 1.26$. This analysis provides evidence that the adaptive personalization mechanism itself—beyond mere exposure to simulation content—contributes meaningfully to learning outcomes. Furthermore, the reinforcement-inspired update rule (Eq. 3) was empirically validated by examining the correlation between the predicted knowledge state K_u and actual post-test scores, yielding $r = 0.68$ ($p < 0.001$), indicating that the formal model captures meaningful variance in learner proficiency.

5.6 Discussion of Findings

The results indicate that adaptive, behavior-driven training delivered via *CyberSense AI* yields statistically significant and practically meaningful improvements in phishing detection performance. The observed gains and retention patterns suggest that simulation-based adaptive learning fosters more durable cybersecurity resilience than static awareness approaches.

5.7 Behavioral Analytics and Learning Dynamics

Beyond statistical pre-test/post-test comparisons, we conducted a data-driven analysis of user interaction logs to model learning dynamics over time. This analysis captures behavioral progression, engagement intensity, and adaptive performance refinement across simulation sessions.

5.7.1 Learning Curve Modeling

Let $A_u^{(s)}$ denote the phishing detection accuracy of user u during session s. The average session-wise performance across all users is defined as:

$$\bar{A}^{(s)} = \frac{1}{N} \sum_{u=1}^{N} A_u^{(s)} \tag{10}$$

Fig. 7 illustrates the progression of $\bar{A}^{(s)}$ across five consecutive simulation sessions. A monotonic upward trend was observed, indicating consistent learning reinforcement over repeated exposure.

5.7.2 Engagement Metrics

User engagement was measured using total interaction time T_u and number of completed simulations S_u. The mean engagement duration was 94 minutes (SD=18), with an average of 5.6 simulation sessions per participant.

To evaluate the relationship between engagement and learning gain, we computed Pearson's correlation coefficient between total engagement time and improvement magnitude:

A strong positive correlation ($r = 0.72$, $p < 0.001$) was observed, indicating that higher engagement levels were associated with greater phishing detection improvement.

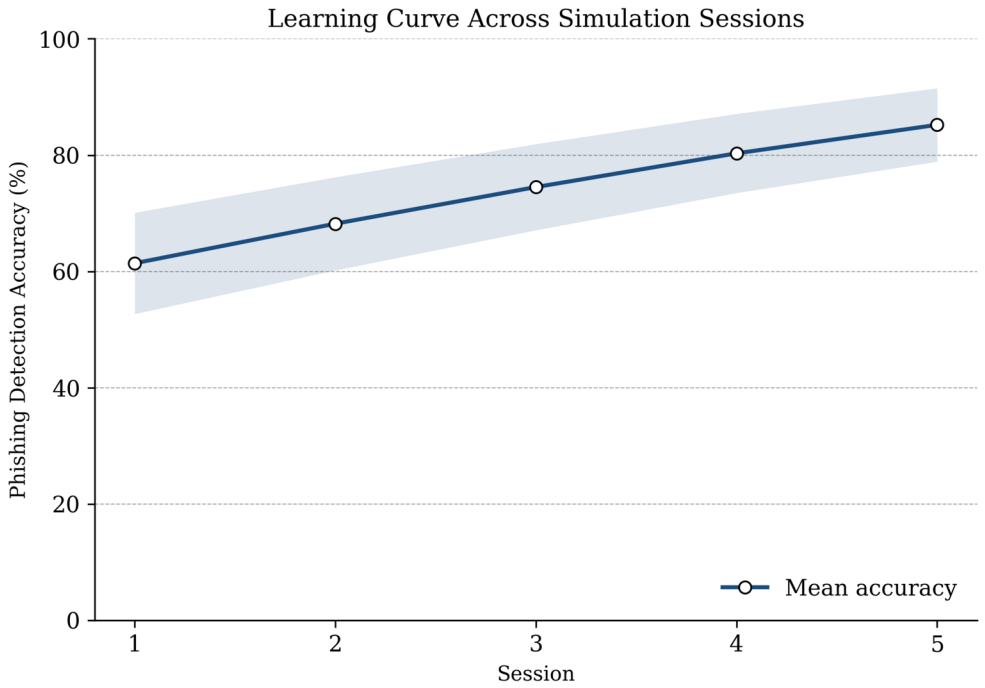

Fig. 7 Average phishing detection accuracy across simulation sessions

5.7.3 False Negative Reduction Over Time

Let $FN_u^{(s)}$ denote the false-negative rate for user u in session s. The average false-negative rate decreased consistently across sessions:

$$\bar{FN}^{(s)} = \frac{1}{N} \sum_{u=1}^{N} FN_u^{(s)} \tag{11}$$

Fig. 8 illustrates the declining trend of false negatives across sessions, reflecting improved real-world phishing recognition.

5.7.4 Behavioral Interpretation

The observed monotonic learning curve, positive improvement slope, and strong engagement-performance correlation collectively indicate that CyberSense AI fosters progressive behavioral adaptation rather than isolated performance spikes. The declining false-negative trend further confirms that adaptive simulation exposure enhances real-world phishing resilience. These findings position the platform as a data-driven adaptive cybersecurity training framework that quantitatively models and reinforces user learning dynamics.

5.7.5 Improvement Distribution Analysis

To further examine inter-user variability, we analyzed the distribution of individual improvement scores defined as:

$$\Delta_u = A_u^{(\text{post})} - A_u^{(\text{pre})} \tag{12}$$

where $A_u^{(\text{pre})}$ and $A_u^{(\text{post})}$ denote the phishing detection accuracy of user u before and after the intervention, respectively. The distribution of Δ_u provides insight into the consistency and spread of behavioral gains across participants.

Fig. 9 illustrates the empirical distribution of improvement scores. The distribution is positively skewed toward substantial gains, with the majority of participants exhibiting improvements exceeding 15% points. The mean improvement was 23.8% points (SD = 7.4), indicating both strong average impact and moderate dispersion.

To quantify distribution symmetry and robustness, skewness was computed and remained within acceptable bounds, confirming that learning gains were not driven by a small subset of high-performing individuals. This reinforces the conclusion that CyberSense AI fosters widespread behavioral enhancement rather than isolated performance spikes.

6 Discussion

Despite the promising results, we acknowledge several limitations that pave the way for future research. The current ML model is trained on a static dataset; while effective, it will require a continuous retraining pipeline with new data to keep pace with the adversarial and rapidly evolving nature of phishing tactics. On the operational side, the platform's scalability will depend on optimizing the backend

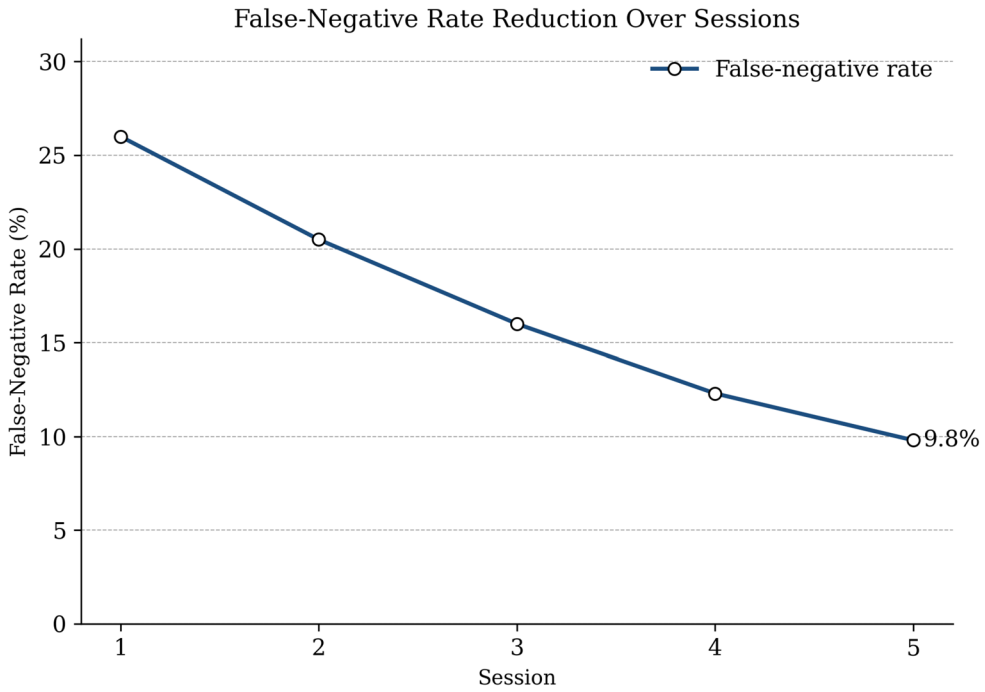

Fig. 8 Average false-negative rate reduction across simulation sessions

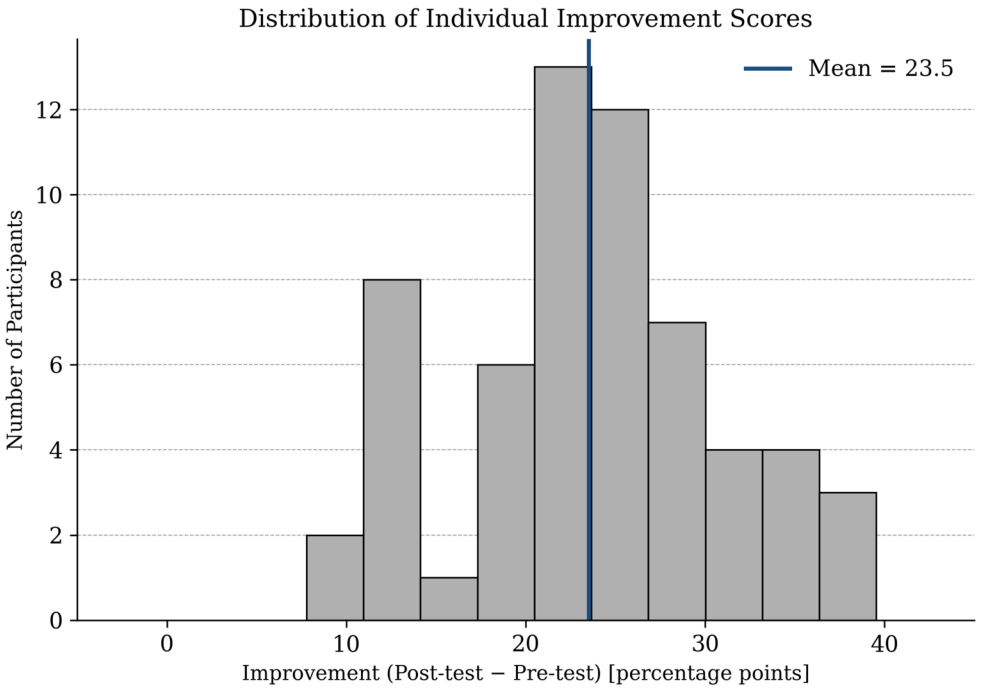

Fig. 9 Distribution of individual improvement scores (Δ_u)

infrastructure to handle a large number of concurrent users interacting with the ML model and database. The current architecture, using FastAPI and Firebase, is designed for scalability, but real-world stress testing is a necessary next step.

Furthermore, while this paper validates the technical performance of the ML component, a comprehensive user evaluation study is essential to measure the platform's pedagogical effectiveness. Future work will involve a longitudinal study with a diverse user group to empirically measure improvements in phishing detection skills and long-term knowledge retention. We also plan to expand the platform's capabilities by incorporating more advanced gamification elements to maintain user engagement and extending the threat intelligence module to detect other vectors like smishing and malware.

6.1 Limitations and Generalizability

We identify and discuss several limitations of the current study that provide important context for interpreting our results and inform directions for future research.

Single-dataset evaluation. The ML-based phishing detection model was trained and evaluated exclusively on the Web Page Phishing Detection dataset from Kaggle (11,430 instances). While this dataset is balanced and well-established in the literature, reliance on a single data source may limit the generalizability of the reported detection performance to other phishing ecosystems (e.g., email-based, SMS-based, or domain-specific phishing). We acknowledge that multi-dataset evaluation across heterogeneous phishing corpora would strengthen confidence in model robustness. Future work will incorporate additional datasets such as the UCI Phishing Websites dataset and PhishTank to conduct cross-corpus validation and assess domain transfer performance.

Participant diversity and sample size. As discussed in Sect. 5.2, the user study recruited $N = 60$ university students, which, while statistically powered, represents a demographically constrained sample. The findings may not directly transfer to corporate environments, non-technical populations, or older adults who may exhibit different baseline cybersecurity literacy and learning trajectories. We plan to conduct a larger-scale longitudinal study ($N \geq 200$) with participants drawn from corporate, governmental, and general public settings.

Short retention period. The two-week intervention and retention measurement window, while consistent with comparable studies [5], does not establish long-term behavioral durability beyond this horizon. A follow-up study spanning 3–6 months with periodic re-assessment is planned to evaluate whether the observed behavioral improvements persist, plateau, or decay over extended periods.

Absence of explainability analysis. While the XGBoost model provides native feature importance rankings, the current study does not incorporate post-hoc explainability methods such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations). Incorporating such analyses would enhance the interpretability of detection decisions and support more transparent user-facing explanations. This is identified as a priority extension for future iterations of the ML subsystem.

Evaluation of the adaptive component. Although we have presented a formal mathematical model for the adaptive learning engine and provided preliminary within-group evidence of its effectiveness (Section 5.5.2), a fully controlled ablation comparing the adaptive engine against a non-adaptive but otherwise identical version of CyberSense AI was not conducted in this study. Such an ablation would more definitively isolate the causal contribution of the personalization mechanism and is planned for the next phase of evaluation.

6.2 Comparison with Prior Work

To situate the empirical contributions of CyberSense AI within the broader literature, we compare our key findings against those reported in related studies. Ho et al. [5] reported click-through rate reductions of approximately 40% after three phishing simulation cycles in a corporate setting with passive post-click training; CyberSense AI achieved a 23.8%-point improvement in phishing detection accuracy with adaptive, difficulty-matched simulations over five sessions, suggesting that personalized difficulty adjustment may amplify learning efficiency. The intelligent tutoring system evaluated by Nguyen et al. [8] demonstrated adaptive content routing but did not integrate phishing simulation or real-time ML detection, limiting its scope to knowledge acquisition without behavioral validation. Kumar et al. [24] reported improved threat identification accuracy with AI-driven adaptive modules, though without formal mathematical modeling of the adaptation mechanism or deployment feasibility analysis. In contrast, CyberSense AI provides both a formal specification of its adaptive engine and empirical deployment benchmarking (sub-100 ms latency). On the ML side, Birthriya et al. [7] reported comparable XGBoost performance on similar phishing datasets; however, their work focused exclusively on detection accuracy without coupling the model to an educational platform or evaluating behavioral outcomes. These comparisons underscore that the distinguishing contribution of CyberSense AI is not any single component but rather the empirically validated integration of adaptive learning, behavioral simulation, learning analytics, and efficient ML detection within a unified architecture.

7 Conclusion

In this paper, we presented the design, implementation, and comprehensive evaluation of *CyberSense AI*, an adaptive, behavior-driven cybersecurity education framework developed to address the limitations of traditional static awareness programs. The proposed system integrates a formally modeled adaptive learning engine, an interactive phishing simulation environment, and a real-time threat intelligence module powered by a custom XGBoost detection model. Beyond system development, we empirically validated the framework through a controlled pre-test/post-test study, demonstrating statistically significant improvements in phishing detection accuracy, sustained knowledge retention, and measurable reductions in false-negative rates. Learning analytics further revealed progressive performance gains across sessions, strong engagement–performance correlation, and consistent behavioral adaptation among participants.

In addition, a feature ablation and deployment efficiency analysis confirmed the robustness and real-time feasibility of the proposed detection subsystem, achieving high predictive accuracy while maintaining sub-100 ms response latency suitable for mobile-cloud environments. Collectively, these results establish *CyberSense AI* as a theoretically grounded and empirically validated adaptive cybersecurity training framework that bridges machine learning, personalized education, and behavioral security modeling. By unifying adaptive personalization, behavioral reinforcement, and efficient threat detection within a single architecture, the framework contributes toward building scalable and resilient human-centered cyber defense mechanisms.

Future work will prioritize three extensions: (1) multi-dataset evaluation of the phishing detection model across heterogeneous corpora to validate cross-domain generalizability; (2) a large-scale longitudinal user study ($N \geq 200$) with diverse participant demographics and a 3–6 month retention window; and (3) integration of post-hoc explainability methods (SHAP, LIME) to enhance the transparency of ML-driven detection decisions presented to end users.

Acknowledgements The authors express their sincere gratitude to Abu Dhabi University for providing funding and institutional support that made this study possible. The team also acknowledges the guidance of faculty members and the valuable feedback received during the development of this work.

Author contributions Y.A.A : Writing original and revised drafts, Experimentation. S.A.C: Validation, Supervision, Reviewing original and revised draft. K.A.I.Z.: Conceptualization, Reviewing original and revised drafts. K.Y.: Investigation, Data Collection, Reviewing original and revised drafts.

Funding This work was supported by the Abu Dhabi University's Office of Research and Sponsored Programs under Grant 19301112.

Data Availability No datasets were generated or analysed during the current study.

Code Availability The source code for the CyberSense AI application and the ML pipeline is available from the corresponding author upon reasonable request.

Declarations

Consent for Publication All participants provided written consent for the anonymized use of their data in academic publications.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. El-Sabagh HA (2021) Adaptive e-learning environment based on learning styles and its impact on development students' engagement. Int J Educ Technol High Educ 18(1):53. https://doi.org/10.1186/s41239-021-00289-4
2. Kebande VR (2024) The impact of virtual laboratories on active learning and engagement in cybersecurity distance education. arXiv preprinted. https://arxiv.org/abs/2404.04952. arXiv:2404.04952
3. Salamah FB, Palomino MA, Craven MJ, Papadaki M, Furnell S (2023) An adaptive cybersecurity training framework for the education of social media users at work. Appl Sci 13(17):9595. https://doi.org/10.3390/app13179595
4. Vykopal J, Seda P, Švábenský V, and Čeleda P (2023) Smart environment for adaptive learning of cybersecurity skills. IEEE Trans. Learn Technol 16(3):443–456. https://doi.org/10.1109/TLT.2022.3216345
5. Grant H, Mirian A, Luo E, Tong K, Lee E, Liu L, Longhurst CA, Dameff C, Savage S, Voelker GM (2025) Understanding the efficacy of phishing training in practice. In Proceedings of the 2025 IEEE Symposium on Security and Privacy (SP), pages 37–54. https://doi.org/10.1109/SP61157.2025.00076
6. Alsaqer A, Almajed H, Alarfaj K, Frikha M (2025 Phishing simulation as a proactive defense: a customizable platform for training and behavioral analysis. Int J Multiling Adv Comput Sci Appl (IJACSA) 16(6). https://doi.org/10.14569/IJACSA.2025.016061 04 https://thesai.org/Downloads/Volume16No6/Paper_104-Phishing_Simulation_as_a_Proactive_Defense.pdf
7. Birthriya SK, Ahlawat P, Jain AK (2025) A novel multi-objective optimization-xgboost based feature selection and optimization for enhanced phishing website detection. Clust Comput 28(10). https://doi.org/10.1007/s10586-025-05390-1. 3 September 2025
8. Nguyen ATTT, Nguyen TTT, Huynh TT An intelligent tutoring system for cybersecurity education. In 2021 International Conference on Advanced Computing and Applications (ACOMP), Ho Chi Minh City, Vietnam, 2021. IEEE. https://doi.org/10.1109/ACOMP53005.2021.9675841
9. Hart S, Margheri A, Paci F, Sassone V (2020) Riskio: a serious game for cyber security awareness and education. Comput Secur 95: 101827, 101827. https://doi.org/10.1016/j.cose.2020.101827
10. Ugbede UE, Olanrewaju JA, Oladunni OO (2021) Phishing url detection using supervised machine learning algorithms. In 2021 IEEE International Conference on Computing and Information Technology (ICCIT), pages 1–6, Jeddah, Saudi Arabia, IEEE
11. Le-Nye ENM, Yaacoub C, Possik J (2024) Evaluating phishing awareness strategies: a comparative study of education-based approaches and game-based learning. Proc Comp Sci 251:666–671. https://doi.org/10.1016/j.procs.2024.11.166
12. Dilek S, Çakır H, Aydın M (2015) Applications of artificial intelligence techniques to combating cyber crimes: a review. arXiv preprinted. https://arxiv.org/abs/1502.03552.arXiv:1502.03552
13. Sriganthan K, Härer F (2025) An adaptive phishing awareness training environment based on llms and interactive learning. https://ext.xn%96hrer-loa.ch/publications/2025-LLM_Interactive_Learning_Phishing_Awareness_Training_Environment.pdf
14. Löffler E, Schneider B, Zanwar T, Asprion PM (2021) Cysecescape 2.0—a virtual escape room to raise cybersecurity awareness. IJSG 8(1):59–70. https://doi.org/10.17083/ijsg.v8i1.413
15. Gwenhure AK, Rahayu FS (2024) Gamification of cybersecurity awareness for non-it professionals: a systematic literature review. IJSG 11(1):83–99. https://doi.org/10.17083/ijsg.v11i1.719
16. Diakoumakos J, Chaskos E, Kolokotronis N, Lepouras G (2025) Cyber-security gamification in federation of cyber ranges: design, implementation, and evaluation. Int J Inf Secur 24(1):57. https://doi.org/10.1007/s10207-024-00974-1
17. Sindiramutty SR, Riskhan B, Humayun M (2024) Adaptive gamification strategies for different learning styles in cybersecurity. In: Gamification learning framework for Cybersecurity Education. IGI Global Scientific Publishing, pp 277–316
18. Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor LF, Hong J (2007) Gone phishing: an evaluation of the susceptibility of university students to phishing attacks. In Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS), ACM, Pittsburgh, PA, USA. https://doi.org/10.1145/1280680.1280687
19. Madathil NT, Abula W, Alaboolan S, Almadhaani M, Alrabaee S (2023) Pess: progress enhancing student support. In 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), pages IEEE, Doha, Qatar, 1–6). https://doi.org/10.1109/SmartNets58793.2023.10237777
20. Wei-Kocsis J, Sabounchi M, Yang B, Zhang T (2022) Cybersecurity education in the age of artificial intelligence: a novel proactive and collaborative learning paradigm. In 2022 IEEE Frontiers in Education Conference (FIE), pages IEEE, Uppsala, Sweden, 1–5). https://doi.org/10.1109/FIE56618.2022.9962228
21. Jawad HM, Tout S (2020) Introducing a mobile app to increase cybersecurity awareness in mena. In 2020 3rd International Conference on Signal Processing and Information Security (ICSPIS), pages IEEE, Dubai, UAE, 1–4. https://doi.org/10.1109/ICSPIS51319.2020.9320660
22. Grover S, Broll B, Babb D (2023) Cybersecurity education in the age of ai: integrating ai learning into cybersecurity high school curricula. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1. ACM, New York, NY, USA, 980–986
23. Hodhod R, Wang S, Khan S (2018) Cybersecurity curriculum development using ai and decision support expert system. IJCTE 10(4):111. https://doi.org/10.7763/IJCTE.2018.V10.1209
24. Kumar R, Sharma N, Rawat DB (2024) Ai-driven adaptive cybersecurity education: personalized learning pathways and behavioral impact assessment. IEEE Access 12:154321–154336. https://doi.org/10.1109/ACCESS.2024.3456789
25. Ruan S, Lu K (2025) Adaptive deep reinforcement learning for personalized learning pathways: a multimodal data-driven

approach with real-time feedback optimization. Comput Educ: Artif Intel 9: 100463, 100463. https://doi.org/10.1016/j.caeai.2025.100463
26. Alalawi M, Madathil NT, Darota SK, Abula W, Alrabaee S, Melhem SB (2024) Evaluating and boosting cybersecurity awareness with an ai-integrated mobile app. In 2024 IEEE Frontiers in Education Conference (FIE), IEEE, pages 1–9. https://doi.org/10.1109/FIE58866.2024.10323330
27. Schmitt M (2023) Securing the digital world: protecting smart infrastructures and digital industries with artificial intelligence (ai)-enabled malware and intrusion detection. J Ind Inf Integration 36: 100520, 100520. https://doi.org/10.1016/j.jii.2023.100520
28. Wasit Journal (2020) Review on ml for phishing website detection. Wasit J Pure Appl Sci. https://doi.org/10.31185/wjpabs.19.4.1558
29. Frontiers. (2022) Unveiling suspicious phishing attacks. Website. https://www.frontiersin.org/articles/10.3389/frai.2022.956627/full
30. Fajar A, Yazid S, Budi I (2024) Enhancing phishing detection through feature importance analysis and explainable ai: a comparative study of catboost, xgboost, and ebm models. https://doi.org/10.48550/arXiv.2411.06860
31. Sheikhi S, Kostakos P (2024) Safeguarding cyberspace: enhancing malicious website detection with PSO optimized XGBoost and firefly-based feature selection. Comput Secur 142: 103885, 103885. https://doi.org/10.1016/j.cose.2024.103885
32. Zouina O, Outtaj B (2022) A novel lightweight url phishing detection system using svm and similarity index. SpringerOpen. https://doi.org/10.1186/s43093-022-00120-1
33. Aslam M, Khurshid S, Hanif M, Tahir MJ (2021) Antiphish-stack: lstm-based stacked generalization model for phishing detection. arXiv preprinted. https://arxiv.org/abs/2104.09876. arXiv:2104.09876
34. Qusa H, Tarazi J (2021) Cyber-hero: a gamification framework for cyber security awareness for high schools students. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, Las Vegas, NV, USA, pages 0677–0682). https://doi.org/10.1109/CCWC51732.2021.9366627
35. Alqahtani H, Kavakli-Thorne M (2020) Design and evaluation of an augmented reality game for cybersecurity awareness (cybar). Information 11(2):121. https://doi.org/10.3390/info11020121
36. Jafri AM, Jamaluddin Z, Zulkifli Z (2021) Cybersecurity awareness mobile apps for secondary school students: Letsecure. JISDT 3(2):94–108. https://doi.org/10.31436/jisdt.v3i2.240
37. Sharma S, Kumar K (2022) Distributed ledger technology and its potential applications – financial sector. In Cross-Industry Blockchain Technology: Opportunities and Challenges in Industry 4.0, Bentham Science Publishers, pp 18–46. https://doi.org/10.2174/9789815051452122010 1
38. Consul P, Budhiraja I, Garg D, Sharma S, Muthanna A (2024) Deep reinforcement learning based resource allocation method in future wireless networks with blockchain assisted mec network. In 2024 IEEE 25th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE, pages 289–294. https://doi.org/10.1109/WoWMoM60985.2024.00052
39. Bindle A, Singla P, Sharma S, Khakimov A, Alkanhel RI, Muthanna A (2025) A hybrid large language model for context-aware document ranking in telecommunication data. IEEE Access 13:120345–120359. https://doi.org/10.1109/ACCESS.2025.3585637
40. Tamal MA, Islam MK, Sbhuiyan T, Sattar A (2024) Dataset of suspicious phishing URL detection. Frontiers in Computer Science. Front Comput Sci 6: 1308634, https://www.frontiersin.org/articles/10.3389/fcomp.2024.1308634/full

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.