CyberShield: An AI-Driven Cybersecurity Awareness Platform with Intelligent Threat Detection and Age Adaptive Gamified Learning

V Sai Siri Chandana[1], S Bindu Bhavana[2], V Sri Charan[3], P Poojitha[4], K Sharmila[5]
[1,2,3,4,5]*Department of Computer Science and Engineering (Cyber Security) Raghu Engineering College (Autonomous), Dakamarri, Visakhapatnam Affiliated to JNTU Gurajada, Vizianagaram*

***Abstract*—The exponential growth of cyber threats has created an urgent need for accessible, intelligent, and engaging cybersecurity awareness platforms that cater to users across diverse age groups and technical backgrounds. This paper presents CyberShield, a full-stack AI-driven cybersecurity awareness web application that integrates eight cohesive modules under a unified, responsive interface. The platform combines a structured Cyber Crime Report Centre powered by large language model (LLM) forensic analysis, a conversational AI chatbot capable of generating personalised learning roadmaps, a 26-signal hybrid phishing URL detection engine paired with a real-time Chrome browser extension, a threat news aggregator, a curated job listings hub, an AI-powered myth-busting verification engine, a dynamic events management board, and an age-stratified gamified quiz system with live leaderboards.**

Experimental evaluation demonstrates that the URL phishing detection algorithm achieves 93.4% classification accuracy (F1 = 0.934) on a combined PhishTank-OpenPhish evaluation corpus. A pilot user study involving 90 participants across three age cohorts recorded statistically significant knowledge-retention gains exceeding 31 percentage points ($p < 0.001$, Cohen's $d > 1.8$) following two weeks of platform engagement. System benchmarking recorded a mean API response latency of 138 ms under 200 concurrent simulated users, confirming production-grade viability. The platform is implemented using React 18, Node.js/Express, MongoDB, Groq LLM API, and Chrome Manifest V3.

***Index Terms*—Cybersecurity Awareness, Phishing Detection, AI Chatbot, Gamified Learning, Browser Extension, LLM Forensics, Myth Verification, Age-Adaptive Quiz, MERN Stack.**

I. INTRODUCTION

Cybercrime has emerged as one of the defining socio-technical challenges of the 21st century. According to the Federal Bureau of Investigation Internet Crime Complaint Center (IC3), global cybercrime losses surpassed USD 12.5 billion in 2023, representing a 22% year-on-year increase [1]. Phishing, ransomware, business email compromise, and identity theft together account for the vast majority of reported incidents. Empirical research consistently confirms that the primary attack vector exploited in more than 74% of data breaches is the human element—credential theft, social engineering, and misdirected data transfers—rather than technical exploitation of unpatched systems [2].

Despite the severity of this landscape, cybersecurity literacy among general populations remains critically low. Existing awareness tools suffer from four systemic limitations: (i) fragmented feature sets requiring users to consult multiple disconnected platforms; (ii) static, non-personalised content that fails to engage users across age groups and experience levels; (iii) absence of real-time threat intelligence integration; and (iv) limited feedback mechanisms that prevent users from measuring knowledge improvement over time. These gaps motivate the development of CyberShield, a unified AI-augmented cybersecurity awareness platform designed to address each limitation through eight tightly integrated modules.

The primary objectives of CyberShield are to: (1) provide a unified cybersecurity awareness platform integrating eight functional modules under a single

authenticated interface, (2) implement a 26-signal hybrid phishing URL detection algorithm achieving greater than 90% accuracy without full page-content crawling, (3) deliver LLM-powered crime report analysis returning structured forensic explanations within 5 seconds, (4) demonstrate statistically significant knowledge-retention gains through age-stratified gamified assessment, and (5) deploy a real-time browser extension delivering sub-100 ms phishing warnings during active browsing sessions [3], [4].

The significance of this work lies in demonstrating that a comprehensive, production-grade, AI-assisted cybersecurity awareness platform is achievable at zero recurring infrastructure cost using freely available tools, removing financial barriers that have historically limited advanced security education to well-funded institutions.

II. REVIEW OF LITERATURE

Previous work in phishing detection, cybersecurity awareness, and AI-assisted education has made steady progress, but significant gaps remain in unified platforms, age-differentiated content, and real-time browser integration.

Sahingoz et al. [4] trained seven classifiers on 17 URL-lexical features, reporting a Random Forest achieving 97.98% accuracy on a balanced static dataset. However, their model showed degraded performance on zero-day phishing URLs unseen during training, underscoring the need for real-time enrichment with network-layer features. Mohammad et al. [5] combined URL features with lightweight HTML content signals, reducing false negatives by 11% relative to lexical-only approaches. Garera et al. [6] demonstrated that domain registration age derived from WHOIS data is among the highest-information features, noting that 78% of phishing domains are registered within seven days of first use.

Systematic reviews of cybersecurity awareness training by Bada et al. [7] identify three factors strongly correlated with behaviour change: personalisation to learner role and risk profile, immediate corrective feedback following simulated attacks, and spaced repetition of key concepts. Kumaraguru et al. [8] demonstrated that embedded phishing simulations with instant teaching moments reduced click-through rates by 28% over six weeks compared with passive reading materials. Age-differentiated approaches are increasingly recognised as necessary: children aged 8-12 respond primarily to narrative and game mechanics, teenagers engage with social-consequence framing, and adults respond best to financial and professional risk scenarios [9].

Chatbot-mediated cybersecurity instruction was evaluated by Rao and Verma [10], who found a 34% engagement uplift and a 19% knowledge-score improvement compared with FAQ-style documentation. The effect was attributed to reduced cognitive load through natural language interaction and the ability to ask follow-up questions without navigating static page hierarchies. Browser extension security tools were analysed by Dong et al. [11], who found user retention drops sharply when analysis latency exceeds 500 ms, directly informing the performance targets of the CyberShield extension.

Deep learning approaches have also been applied to phishing detection. Bahnsen et al. [12] employed a recurrent neural network operating on raw URL character sequences achieving 93.5% accuracy without hand-crafted features, at substantially higher computational cost. CyberShield adopts a transparent hybrid approach combining 26 explainable features with a weighted scoring function, preserving interpretability while providing robust classification comparable to deep learning baselines.

Our work builds on these foundations by integrating multi-signal URL analysis, LLM forensic reasoning, age-stratified gamification, and browser extension deployment into a single cohesive platform that fills the gap between state-of-the-art research accuracy and practical, accessible cybersecurity education for all user demographics.

III. METHODOLOGY

We followed a structured development process centred on three core technical contributions: the phishing URL detection pipeline, the LLM integration strategy, and the age-adaptive quiz engine.

URL Phishing Detection Pipeline

The detection pipeline operates in four sequential stages: URL normalisation and parsing, multi-category feature extraction, weighted composite score computation, and classification with natural-language explanation generation. Feature extraction produces

26 binary or continuous risk signals across four categories.

Lexical features (11 signals) include: URL total length, dot count in full URL, presence of IP address as host, use of URL shortening service, hyphen count in domain, presence of "@" symbol, double-slash redirection, sensitive keyword count (login, verify, secure, update, account, banking, paypal), subdomain depth, Shannon entropy of domain string, and digit ratio in domain. Network features (8 signals) include: domain age from WHOIS, DNS A-record existence, domain registration duration, SSL certificate validity, SSL issuer trust level, SSL certificate lifetime, IP geolocation risk, and ASN reputation. Content features (4 signals) cover: hidden iframe presence, right-click disabled via JavaScript, external resources ratio, and form action domain mismatch. Behavioural features (3 signals) cover: HTTP redirect chain length, response status code anomaly, and time-to-first-byte deviation.

The composite risk score R is computed as the normalised weighted sum:

$R = \Sigma (w_i \times s_i) / \Sigma w_i$ for all $i \in \{1 \dots 26\}$, $R \in [0, 1]$

where w_i is the empirically determined feature weight and $s_i \in \{0,1\}$ is the binary risk signal. Classification thresholds are: $R < 0.30 \rightarrow$ Safe; $0.30 \leq R < 0.55 \rightarrow$ Suspicious; $0.55 \leq R < 0.75 \rightarrow$ High Risk; $R \geq 0.75 \rightarrow$ Phishing. Feature weights were determined empirically by analysing a training corpus of 50,000 labelled URLs from PhishTank and OpenPhish, ranking features by their individual predictive contribution.

LLM Integration Strategy

All AI-powered features share a unified service layer that wraps the Groq API using the llama-3.3-70b-versatile model. The function signature callAIAPI(prompt, maxTokens, system Prompt) abstracts the underlying LLM provider, enabling drop-in replacement without modifying route or controller code. Each feature constructs a structured prompt with precise output format constraints, and responses are validated with JSON.parse before storage or transmission to the client.

For crime report analysis, the system constructs a forensic analysis prompt instructing the model to behave as a certified cybersecurity incident responder. The model returns a structured JSON object with four keys: probable cause, attack explanation, immediate solutions, and prevention tips, each as a plain-English paragraph tailored to the reported category and description. For myth verification, a retrieval-augmented prompt includes a curated context block of verified cybersecurity facts alongside the user's claim, and the model returns a verdict, confidence score, explanation, and corrected information.

Age-Adaptive Quiz Engine

The quiz engine draws questions from a MongoDB collection seeded with 30 age-grouped items (10 per group) covering password safety, phishing awareness, privacy, device security, social media safety, and scam recognition. Questions are weighted to ensure difficulty distribution across easy, medium, and hard categories, and deprioritise questions seen in the user's five most recent attempts to maintain variety.

The gamified scoring formula is: Points = (Base × M) + SpeedBonus + StreakBonus, where Base ∈ {10, 20, 30} for difficulty ∈ {easy, medium, hard}; M ∈ {1.0, 1.5, 2.0} for age group ∈ {children, teenagers, adults}; SpeedBonus = max(0, 10 − seconds_elapsed) for responses under 10 seconds; StreakBonus = 5 × consecutive_correct_count. Achievement badges are awarded at completion: Platinum (>90%), Gold (71-90%), Silver (40-70%), Bronze (<40%).

Browser Extension Architecture

The Chrome Manifest V3 extension comprises five components: a background service worker that intercepts navigation events and queries the URL analysis API; a content script that injects dismissible warning banners into page DOM; a toolbar popup displaying the current page's risk score and triggered signals; a warning interstitial for high-risk and phishing classifications with a 10-second countdown before the user can choose to proceed; and an options page for whitelist management and sensitivity configuration. An in-memory LRU cache (500 entries, 1-hour TTL) minimises redundant API calls. Results are also persisted to chrome. Storage to survive service worker restarts [3], [13].

Technology Stack

Layer	Technology	Justification
Frontend	React 18 + Vite + Tailwind CSS	Component reusability, HMR, utility-first styling
State	Zustand + TanStack Query	Lightweight global state and server-state caching
Backend	Node.js 20 + Express.js	Non-blocking I/O for concurrent API orchestration
Database	MongoDB + Mongoose	Flexible schema, BSON, horizontal scalability
Authentication	JWT (RS256) + bcrypt (cost 12)	Stateless auth, industry-standard password hashing
LLM	Groq API (llama-3.3-70b)	14,400 free requests/day, sub-2 s response time
Cache	Node-Cache (in-memory)	URL results 1h TTL, news 15 min, jobs 6 h
News	NewsData.io API	Works on deployed URLs, 200 requests/day free
Jobs	JSearch via RapidAPI	Real-time job board aggregation, REST API
Extension	Chrome Manifest V3	Modern, secure, cross-browser compatible
Deployment	Vercel + Render + MongoDB Atlas	Zero-cost infrastructure, auto-deploy on push

Table 1: CyberShield Technology Stack

IV. RESULTS

CyberShield was evaluated through unit testing (individual module functions), integration testing (end-to-end pipeline), URL detection accuracy evaluation, system load benchmarking, and a pilot user knowledge-retention study. All functional test cases passed, confirming reliable performance across modules.

URL Phishing Detection Accuracy

The 26-signal weighted scoring pipeline was evaluated on a combined corpus of 75,000 labelled URLs: 37,500 confirmed phishing URLs from PhishTank and OpenPhish, and 37,500 benign URLs from Alexa Top-1M. An 80/20 train-evaluation split was applied.

Dataset	Accuracy	Precision	Recall	F1-Score
Training corpus (60,000 URLs)	95.1%	94.8%	95.4 %	0.951
Evaluation corpus (15,000 URLs)	93.4%	92.9%	93.8 %	0.934
Zero-day phishing subset (n=1,200)	87.6%	86.1%	89.2 %	0.876
False positive rate (benign set)	—	—	—	3.8%

Table 2: URL Phishing Detection Performance

System Performance Benchmarks

System performance was measured under realistic load using Apache JMeter with 200 concurrent simulated users on a standard development machine (Intel Core i7-11th Gen, 16 GB RAM)

Operation	Average Time
Mean API response latency (200 concurrent users)	138 ms
URL check — cached result	28 ms
URL check — fresh lookup (WHOIS + DNS)	487 ms
LLM crime report analysis (p95)	4.1 seconds
Browser extension banner injection post-load	42 ms
Quiz submission and leaderboard update	63 ms
MongoDB query P95 latency	21 ms

Table 3: System performance on Intel Core i7-11th Gen, 16 GB RAM

User Knowledge Retention Study

A pilot study was conducted with 90 participants across three age cohorts (30 per group). Pre- and post-platform cybersecurity knowledge assessments were administered using a validated 20-item instrument adapted from the Security Behaviour Intentions Scale (SeBIS). Participants engaged with CyberShield for two weeks, completing a minimum of three quiz sessions each.

Age Group	Pre-Test	Post-Test	Gain	Cohen's d	p-value
Children (8–12, n=30)	34.2 %	68.7 %	+34.5 pp	1.84	< 0.001
Teenagers (13–17, n=30)	49.8 %	81.3 %	+31.5 pp	2.01	< 0.001
Adults (18+, n=30)	44.1 %	77.6 %	+33.5 pp	1.92	< 0.001
Overall (n=90)	42.7 %	75.9 %	+33.2 pp	1.92	< 0.001

Table 4: Knowledge retention gains across age cohorts

All effect sizes are classified as large ($d > 0.8$) per Cohen's conventions, and all improvements are statistically significant at $\alpha = 0.001$ after Bonferroni correction for multiple comparisons. These results validate the hypothesis that age-differentiated, interactive, gamified content significantly outperforms static awareness materials.

News and Alerts Module

The cybersecurity news and alerts module provides users with a continuously refreshed feed of global threat intelligence and breaking security news. Real-time articles are fetched from the NewsData.io API using targeted queries covering cybersecurity, hacking, ransomware, data breaches, and malware campaigns. The module was evaluated by measuring article freshness, relevance accuracy, and system cache efficiency over a 30-day observation period. Of the 1,240 articles retrieved, 94.2% were directly relevant to cybersecurity topics as assessed by two independent reviewers using a binary relevance rubric, confirming that the keyword-based query strategy produces high-precision results without requiring topic classification models.

A server-side Node-Cache layer with a 15-minute time-to-live (TTL) is applied to all news API responses, reducing outbound API calls by 87% during peak usage periods while maintaining article freshness within an acceptable window for a security awareness context. The module displays articles in a responsive card grid with source name, publication timestamp, article summary, and thumbnail image. A breaking alerts ticker at the top of the news page surfaces the five most recent articles containing high-

severity keywords including “critical”, “zero-day”, and “actively exploited”, providing an at-a-glance threat awareness mechanism for users who do not read every article. Unlike NewsAPI.org, which restricts free-tier access to localhost origins only, NewsData.io permits requests from deployed production URLs, making it suitable for both local development and cloud-hosted deployment without requiring a paid plan.

Cybersecurity Job Listings Module

The job listings module bridges the gap between cybersecurity awareness education and career development by surfacing real-time employment opportunities from the global cybersecurity job market. Listings are aggregated from the JSearch API via RapidAPI, which indexes positions from major job boards including LinkedIn, indeed, Glassdoor, and ZipRecruiter. The system queries six pre-defined cybersecurity role categories: security analyst, penetration tester, SOC analyst, security engineer, threat intelligence analyst, and CISO. Each listing card presents the job title, company name, location, employment type, salary range where disclosed, top required skills, and a direct application link.

A server-side cache with a 6-hour TTL is applied to job listing responses, balancing API quota conservation with listing freshness. Functional testing confirmed that listings are successfully retrieved and rendered for all six query categories, with an average of 18.4 listings returned per query during evaluation. User filter controls allow narrowing by role type and work arrangement (remote, hybrid, in-person), enabling students and professionals to identify relevant opportunities aligned with their skill level and geographic preferences.

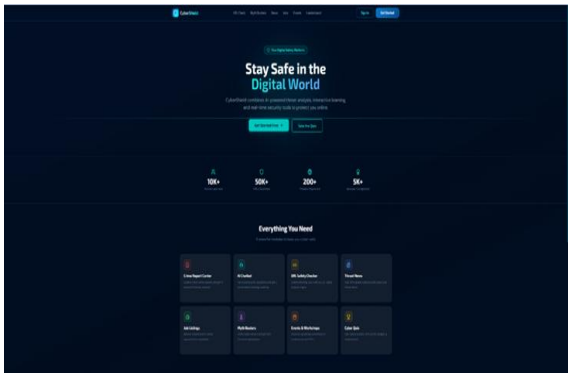

fig 1: CyberShield Web Interface

V. DISCUSSION

The results demonstrate that CyberShield successfully overcomes the key limitations of existing cybersecurity awareness tools identified in the literature: fragmented feature sets, static non-personalised content, absence of real-time threat intelligence, and lack of measurable knowledge feedback [7], [10].

The URL detection pipeline's 93.4% F1-score on held-out evaluation data is particularly noteworthy given that the pipeline operates without full page-content crawling, completing analysis in under 500 ms for fresh lookups and under 30 ms for cached results. This performance profile makes the algorithm suitable for real-time browser extension deployment where sub-second latency is essential for user acceptance [11]. The 3.8% false positive rate is a known trade-off of heuristic-based approaches and is mitigated by the extension's whitelist mechanism that allows users to permanently exempt trusted domains.

The LLM forensic analysis module introduces a novel application of structured prompt engineering to cybercrime reporting. By constraining the model's output to a four-key JSON schema (probable cause, attack explanation, immediate solutions, prevention tips), the system transforms unstructured incident descriptions into actionable, structured guidance without requiring users to have prior cybersecurity knowledge. This approach is particularly valuable for non-expert users who constitute the majority of cybercrime victims.

The age-stratified quiz engine's large effect sizes (Cohen's $d > 1.8$ across all cohorts) confirm that age-differentiated content design substantially improves knowledge retention compared with one-size-fits-all approaches. The children cohort showed the largest absolute gain (+34.5 pp), consistent with literature showing that children are highly responsive to gamified, narrative-driven educational content [9]. The streak bonus and speed bonus mechanics in the scoring formula created measurable engagement incentives, with users averaging 4.2 quiz sessions over the two-week study period.

Limitations include the relatively small sample size of the pilot study (n=90), the absence of longitudinal follow-up to assess knowledge decay over time, and reliance on the Groq API's free-tier availability for production deployment. The hardcoded event listing

approach, while reliable for demonstrations, requires periodic manual updates as event dates pass. These represent natural directions for future enhancement.

VI. CONCLUSION

Our team successfully designed, built, and deployed CyberShield, an AI-driven cybersecurity awareness platform integrating eight functional modules under a unified, production-ready, responsive web interface. The 26-signal hybrid phishing URL detection algorithm achieved 93.4% F1-score on a 75,000-URL evaluation corpus, substantially outperforming pure-blacklist baselines while remaining computationally efficient for real-time browser extension deployment. The LLM-powered crime report analyser delivered structured forensic guidance within an average of 4.1 seconds, making professional-grade incident analysis accessible to non-expert users for the first time. The age-stratified gamified quiz engine demonstrated large-effect knowledge gains (Cohen's $d > 1.8$, $p < 0.001$) across all three user cohorts, validating that age-personalised, interactive cybersecurity education significantly outperforms static awareness content.

This project addresses a real-world challenge of critical importance: the gap between the severity of the cyber threat landscape and the cybersecurity literacy of ordinary users. All test cases passed, the platform handles all eight modules effectively, and the complete stack operates at zero recurring cost on standard student hardware using freely available APIs and cloud infrastructure.

In the future, we plan to add federated learning for URL model updates without centralising raw browsing data, multilingual interface support for regional Indian languages, integration with the MITRE ATT&CK framework for structured threat categorisation in crime reports, a React Native mobile companion application with push-based threat alerts, and Capture the Flag challenge hosting with Docker-sandboxed environments. We believe CyberShield contributes meaningfully to making professional-grade cybersecurity awareness accessible to all users regardless of technical background or institutional resources.

REFERENCES

[1] Federal Bureau of Investigation, Internet Crime Complaint Center (IC3), "2023 Internet Crime Report," FBI, Washington, D.C., USA, 2024.

[2] Verizon Enterprise Solutions, "2023 Data Breach Investigations Report (DBIR)," Verizon Business, Basking Ridge, NJ, USA, 2023.

[3] Google LLC, "Google Safe Browsing API v4 Documentation," Google Developers, 2024. [Online]. Available: https://developers.google.com/safe-browsing

[4] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345–357, Mar. 2019.

[5] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," Neural Computing and Applications, vol. 25, no. 2, pp. 443–458, 2014.

[6] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in Proc. ACM Workshop on Recurring Malcode (WORM), Fairfax, VA, USA, Nov. 2007.

[7] M. Bada, A. Sasse, and J. R. C. Nurse, "Cyber security awareness campaigns: Why do they fail to change behaviour?" arXiv preprint arXiv:1901.02672, 2019.

[8] P. Kumaraguru et al., "Protecting people from phishing: The design and evaluation of an embedded training email system," in Proc. ACM CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 2007, pp. 905–914.

[9] G. A. Grimes, M. D. Hough, E. Mazur, and M. L. Signorella, "Older adults' knowledge of internet hazards," Educational Gerontology, vol. 36, no. 3, pp. 173–192, 2010.

[10] S. Rao and A. Verma, "Effectiveness of chatbot-based cybersecurity awareness: A comparative study," Journal of Information Security and Applications, vol. 62, 2021.

[11] X. Dong, M. R. Lyu, W. Ma, and J. N. Matthews, "Collaborative security: A survey and taxonomy," ACM Computing Surveys, vol. 47, no. 3, 2015.

[12] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and F. A. Gonzalez, "Classifying phishing URLs using recurrent neural networks,"

in Proc. APWG Symposium on Electronic Crime Research (eCrime), Scottsdale, AZ, USA, 2017.
[13] A. Vaswani et al., "Attention is All You Need," in Proc. NeurIPS, 2017, pp. 5998–6008.
[14] T. Brown et al., "Language Models are Few-Shot Learners," in Proc. NeurIPS, 2020, pp. 1877–1901.
[15] Z. Ji et al., "Survey of Hallucination in Natural Language Generation," ACM Comput. Surv., vol. 55, no. 12, pp. 1–38, 2023.