Browser-Embedded, Legal-Aware Cybersecurity Co-Pilot for Myanmar with RAG, Multilingual Defense, and Privacy

Htet Myet Zaw
University of Information Technology
Yangon, Myanmar
htetmyetzaw@uit.edu.mm

Hpone Khant Naing
University of Information Technology
Yangon, Myanmar
hponekhantnaing@uit.edu.mm

Aung Kaung Myat
University of Information Technology
Yangon, Myanmar
aungkaungmyat_2022@uit.edu.mm

Htet Paing Linn
University of Information Technology
Yangon, Myanmar
htetpainglinn@uit.edu.mm

Thurein Lin
University of Information Technology
Yangon, Myanmar
thureinlin@uit.edu.mm

Dr. Thiri Thitsar Khaing
University of Information Technology
Yangon, Myanmar
thirithitsarkhaing@uit.edu.mm

We present Konbaung AI, a multilingual, privacy-preserving browser extension that integrates real-time cyber defense with contextual legal awareness for Myanmar's digital citizens. The system combines an Ollama-compatible large language model with Retrieval-Augmented Generation (RAG) over Myanmar Cyber Law and global privacy and security frameworks, while incorporating natural language–based phishing detection, URL heuristics, cryptographic file hashing with malware verification, and breach exposure analysis. A privacy-by-default design ensures local inference, minimal external dependencies, and the absence of remote user data storage. We formalize a threat model for browser-resident AI and propose a modular architecture that couples lightweight client processing with secure backend services. Experimental evaluation across detection accuracy, response latency, and user interaction indicates phishing and spam classification, close alignment with established malware verdicts, effective legal and privacy policy summarization in Burmese, and sustained engagement through adaptive gamification. Remaining challenges include dialectal speech processing and limited domain-specific datasets, offline-capable mechanisms, and broader cross-platform deployment.

Keywords—Cybersecurity, Browser Extensions, RAG, Myanmar Cyber Law, GDPR, PDPA, Phishing, NLP, Privacy by Design, VirusTotal, HaveIBeenPwned.

I. Introduction

Konbaung AI is a multilingual, AI-powered browser extension that redefines personal cybersecurity for Myanmar's digital citizens. It empowers users with an always-accessible, privacy-respecting AI chatbot trained on Myanmar's Cyber Laws while protecting them in real- time from phishing, scams, malware, and digital manipulation. This dual-purpose tool combines awareness and defense which are two critical layers often missing in cybersecurity adoption delivered through a gamified, hyper-local experience tailored for the Myanmar population.

Key problems being addressed include rising phishing attacks, lack of accessible cybersecurity education in Burmese, and low awareness of local cyber legislation. Over 63% of Myanmar's internet users primarily access the web through browsers and are vulnerable to cyber deception via social media and messaging apps. Konbaung AI fills this critical void by embedding protection directly into the user's browser and delivering education on-the-go.

A. Research Contributions

- RAG-enhanced LLM chatbot: Integration of Myanmar Cyber Law, GDPR, PDPA, and ISO-27001 guidance into a unified retrieval-augmented chatbot that returns bilingual, citation-anchored answers for in-browser legal guidance.
- Multi-modal threat detection pipeline: Combination of lightweight in-browser NLP classifiers for spam and phishing, URL heuristics, and cryptographic file hashing (SHA-256) with multi-engine malware intelligence lookup.
- Hybrid data-sovereignty design performs local inference while leveraging selective, consented threat-intelligence APIs and privacy-first bilingual policy analysis for constrained devices, minimizing PII transfer and maintaining user control.

B. System Overview

Konbaung AI is a unified browser extension integrating six modules: a RAG-based legal assistant covering Myanmar Cyber Law with GDPR, PDPA, and ISO 27001; natural language classifiers for scams, phishing, and malicious URLs; cryptographic file verification via SHA-256 with malware intelligence; breach analysis through public exposure databases; privacy policy summarization with Burmese translation; and entropy-based password strength evaluation. All inference is performed locally to preserve data sovereignty, with external queries limited to threat intelligence lookups. The extension further enhances user engagement by delivering real-time cybersecurity updates and educational resources.

II. Background

A. Cybersecurity Challenges in Myanmar

Myanmar's internet adoption has grown rapidly but remains below some previously reported figures; as of January 2024, there were approximately 24.1 million internet users in Myanmar corresponding to ~44% penetration of the population (Kepios / DataReportal). At the same time, global phishing volumes reached record levels in 2023, with APWG reporting nearly five million unique phishing attacks in 2023 and elevated quarterly volumes since 2023. Regionally, investigative reporting and CERT advisories document a sustained increase in organized online scam operations and

cross-border scam centers. These trends motivate a locally adapted, privacy-preserving browser approach for Myanmar's users.

B. Regulatory and Data Privacy Landscape

Myanmar's cybersecurity environment operates under regulatory frameworks, including the Myanmar Cyber Law (2021), the GDPR from EU, PDPA from Singapore, and ISO 27001, creating complex multi-jurisdictional compliance requirements. Concurrently, users face significant privacy risks from cross-border data transfers, surveillance practices, and inconsistent corporate transparency. These challenges emphasize the need for privacy-preserving cybersecurity mechanisms and hybrid architectures that combine local inference with cloud intelligence that uphold data sovereignty.

III. Related Work

A. Multilingual Cybersecurity Systems

Most multilingual cybersecurity research has focused on high-resource languages. Phishing detection systems achieve 78–85% accuracy in English, Chinese, and Spanish [3], but accuracy falls to 45–62% for Southeast Asian languages. Cross-lingual transfer learning methods reach 73% in low-resource contexts [6], but existing approaches lack cultural adaptation and fail to account for Myanmar-specific linguistic variation.

B. RAG-Based Security and Legal Systems

Retrieval-Augmented Generation (RAG) has demonstrated strong potential in security, with reported performance of 87% in threat intelligence synthesis [4] and 82% in regulatory compliance [7]. However, current RAG systems remain limited by their single-jurisdiction scope, insufficient support for non-English terminology, lack of cultural grounding, and weak integration with operational security pipelines.

C. Browser-Embedded Security Solutions

Contemporary browser extensions rely on signature-based filtering, with detection rates between 82–89% for known threats [5]. Community-driven solutions provide reputation scoring but lack AI-driven analysis. Machine learning approaches report 91% detection [8], yet existing extensions remain limited in multilingual coverage, legal integration, privacy guarantees, and cultural adaptation.

D. Privacy-Preserving Threat Detection

Privacy-preserving cybersecurity methods have gained importance in response to heightened data protection concerns. Federated learning approaches achieve up to 85% accuracy in threat detection while retaining user data locally [10]. Nevertheless, challenges persist in balancing real-time performance, scalability, and integration. Hybrid architectures combining local processing with selective API integration have been proposed to mitigate these trade-offs.

E. Comparison to prior privacy-preserving systems

Two prior client-side studies, Off-the-Hook and PhishLang, demonstrated effective local phishing detection and LLM-based URL analysis but lacked legal reasoning, multilingual support, and privacy safeguards. Konbaung AI advances this line by integrating Myanmar Cyber Law retrieval, cryptographic hashing with prefix k-anonymity for breach checks, and a privacy-centered user interface. These extensions enhance compliance awareness while maintaining fully client-side processing and balanced performance.

IV. Methodology

A. Intelligent Spam Message Detection

The spam detection system employs a hybrid architecture that integrates AI-based classification with rule-based heuristics to mitigate the growing threat of spam and scam messages in Myanmar. The framework consists of three main components: a real-time text analysis interface, a classification engine based on a multilingual large language model (LLM), and a hybrid detection layer combining machine learning with pattern recognition.

1) Multi-Level Text and AI Analysis: The system performs layered analysis of user inputs, classifying content such as Legitimate, Spam, Scam, or Phishing. Advanced assessment incorporates probabilistic risk scoring, indicator detection, keyword evaluation, and context-aware recommendations. AI-driven insights deliver actionable strategies for prevention, mitigation, and response, ensuring comprehensive threat evaluation while preserving accuracy, transparency, and culturally relevant adaptation for Myanmar-specific content.

2) Hybrid Detection and Data Processing: A multilingual LLM powers hybrid detection through rule-based patterns, semantic assessment, historical analysis, and confidence scoring. Inputs undergo Myanmar-optimized validation, prompt generation, and secure LLM communication, with responses delivered via a localized interface. Analysis history is maintained under user control, supporting transparency and privacy while enabling dynamic calibration to risk tolerance and adaptive learning for evolving cyber threats.

B. Multi-Layered Phishing URL Detection

The phishing detection framework applies a layered approach integrating rule-based heuristics, machine learning, and threat intelligence services to identify malicious URLs. The system architecture includes feature extraction, multi-API verification, machine learning classification, and a consensus-based scoring mechanism.

1) Feature Extraction and Analysis: A 15-dimensional pipeline evaluates URL attributes including TLDs, subdomain depth, path length, lexical characteristics, and character entropy. Suspicious domains and excessive subdomains are flagged, while over forty phishing-related keywords and brand typo squatting patterns are detected. These features provide the foundation for accurate risk assessment and machine learning classification.

2) Multi-API Threat Intelligence: Parallel queries are executed against Google Safe Browsing, PhishTank, URLVoid, and VirusTotal. Fault-tolerant execution with local caching enhances reliability, reduces latency, and provides real-time verification. Intelligence from multiple sources complements feature-based evaluation, ensuring adaptive detection of known and emerging phishing threats with relevant consideration for Myanmar URL patterns.

3) Classification and Risk Scoring: A curated forty-seven-feature dataset supports machine learning classification with adaptive weighting to enhance accuracy. Final risk scores from zero to one hundred integrate blacklist status, lexical analysis, SSL/TLS checks, content evaluation, and reputation signals, classifying URLs as safe, suspicious, or phishing with continuous cached verification for resilience.

C. Privacy Policy Intelligent Analysis

The privacy policy analyzer employs a multi-stage processing pipeline to interpret complex legal documents with high accuracy. The architecture includes document ingestion supporting multiple formats (PDF, DOCX, TXT), an OCR-enabled text extraction engine, a semantic content analysis pipeline, and an AI-based summarization module designed for legal compliance.

1) Natural Language Processing Framework: The preprocessing stage applies to OCR, parsing, hierarchical structuring, and multilingual normalization. Semantic analysis extracts key entities, clusters related sections, assesses tone and complexity, and maps content to relevant legal frameworks for compliance-aware interpretation.

2) Summary Generation Strategies: Summarization integrates extractive, abstractive, and hybrid methods using legal-aware models. Outputs include hierarchical multi-level summaries that highlight data usage, user rights, obligations, and compliance risks.

3) AI-Enhanced Document Intelligence: The summarization pipeline leverages large language models with specialized prompt engineering for privacy policy comprehension. Translation capabilities provide English-to-Myanmar conversion with semantic fidelity and cultural adaptation, ensuring accessibility for local users.

D. Cryptographic File Verification System

The file verification system employs a layered security architecture integrating cryptographic hashing, multi-engine malware detection, and AI-driven threat intelligence, implemented as a browser extension with modular layers for interface, processing, and external intelligence services.

1) Hash-Based File Verification Methodology: SHA256 hashing serves as the core mechanism for file identification and verification, preserving user privacy by preventing content exposure while maintaining linear time complexity and efficient memory utilization.

2) Multi-Engine Malware Detection Integration: File hashes are compared against aggregated antivirus databases, with multi-engine scanning used for unknown files. Classification thresholds categorize files as malicious, suspicious, or safe based on detection counts, enabling probabilistic risk modeling.

3) AI-Powered Threat Intelligence Analysis: Complementing traditional scanning, the system applies transformer-based LLMs for contextual threat analysis. The framework generates structured risk assessments, actionable recommendations, and educational summaries, with Myanmar-language outputs to improve accessibility.

E. Privacy-Preserving Detection of Data Breaches

The breach detection system uses a privacy-preserving methodology to identify compromised credentials while minimizing exposure to sensitive user data. K-anonymity is enforced through cryptographic hashing with prefix-based querying, ensuring that only partial hash information is transmitted to external services, preserving confidentiality without compromising detection accuracy.

1) K-Anonymity Implementation Framework: Email and password breach detection apply SHA-1 hashing with prefix-based querying to preserve k-anonymity. Only the first five characters of each hash are sent to external breach databases, keeping raw credentials private while enabling accurate compromise detection. Password verification uses the same method, supporting scalable, privacy-preserving breach analysis across multiple threat intelligence sources.

2) Multi-Source Breach Intelligence Integration: The system aggregates data from multiple breach intelligence sources, combining temporal analysis, severity scoring, and confidence-based aggregation to provide comprehensive breach detection. Context-aware guidance is delivered in Burmese to improve accessibility for local users.

F. Legal Cybersecurity Knowledge System

The legal cybersecurity module implements a microservice architecture for real-time processing of legal queries. It leverages transformer-based LLMs for cross-jurisdictional analysis and legal reasoning.

1) Legal Knowledge Integration Framework: The system integrates a multi-jurisdictional legal corpus including Myanmar Cyber Law, GDPR, PDPA, and ISO 27001. Content is preprocessed, encoded, and indexed for efficient retrieval. Contextual prompt engineering enables accurate interpretation of queries, generating harmonized compliance recommendations that satisfy local requirements while aligning with international standards. Multilingual responses and structured reasoning support real-time legal guidance.

2) Domain Classification and Query Processing: Queries are categorized into cybersecurity, data protection, digital rights, and general law. Contextual prompt engineering facilitates accurate classification and response generation, with explanations provided in Burmese for clarity.

3) API Integration and Performance Metrics: The model is accessed through a structured API with controlled parameters to ensure safe and coherent output. Reliability measures include error handling, request timeouts, and fallback mechanisms.

G. Myanmar Script fragmentation (Unicode ↔ Zawgyi)

Input text is first scored using a Zawgyi detector to identify non-Unicode segments, which are then converted to Unicode through Rabbit and Google Myanmar Tools, followed by normalization. A syllable-based tokenizer optimized for Burmese orthography processes the cleaned text. Controlled ablation showed a 2.7–3.9 F1 improvement in spam and phishing detection, confirming the effectiveness of the normalization pipeline.

V. System Architecture and Design Overview

The privacy-first browser extension executes all interface and decision logic locally, preserving data sovereignty. Users interact through a side panel, context menu, or new-tab dashboard, with transient logs and chat history retained on-device; external services require explicit consent.

A RAG-powered local AI assistant retrieves passages from Myanmar Cyber Law and international standards (GDPR, PDPA, ISO 27001) to generate grounded bilingual responses, supported by lightweight classifiers for intent verification. The knowledge base incorporates legal documents, bilingual corpora, phishing URLs, and curated scam data for Myanmar-specific threat detection. On-demand analyses leverage NLP, URL heuristics, malware intelligence queries, and cryptographic hashing. Privacy policies are summarized locally and translated into Burmese bullet-point

digests. A breach checker queries public exposure databases to provide actionable remediation guidance.

The new-tab dashboard delivers curated security news, alerts, MITRE ATT&CK guidance, interactive quizzes, and contextual scans. Core operations are both in-browser and new-tab dashboard, with minimal and transparent network calls, delivering an integrated, privacy-preserving cybersecurity experience.

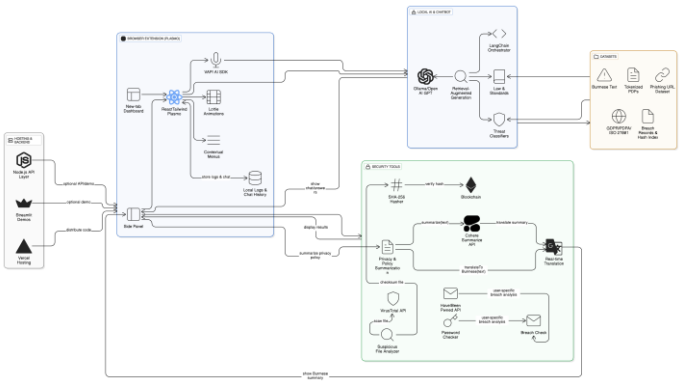

Fig. 1. *High level system architecture of the system*

VI. System Implementation Details

A. Frontend Integration and UX

The privacy-first browser extension executes all interface and decision logic locally. Users interact via a side panel, context menu, or new-tab dashboard, with ephemeral data stored locally under user control. Performance optimizations include memoization, lazy loading, and lightweight styling to ensure accessibility, keyboard navigation, high-contrast themes, and optional voice interaction.

B. LLM, RAG, and Orchestration

For evaluation, we used an Ollama-compatible local LLM runner with a quantized Llama 2 family model. For the experiments reported here, we selected Llama 2 (13B) quantized to 8-bit (ggml/llama.cpp format) to balance response quality and in-browser/edge latency. The selection rationale: 13B offers substantially higher factuality for legal reasoning than 7B in our internal pilot while remaining feasible on a dedicated local host via Ollama and ggml quantization. In production deployments, the system can switch to a smaller 7B variant for low-resource devices at the cost of lower legal reasoning fidelity.

Embeddings & retrieval**:** Document embeddings for the RAG store were computed with a multilingual sentence-embedding model (paraphrase-multilingual-mpnet-base-v2) to support cross-language retrieval across Burmese and English corpora. The embedding vectors were indexed using a lightweight FAISS index, and retrieval is top-k (k=8) with an overlap reranker (cross-encoder) applied to the top candidates.

Runtime & quantization**:** Local inference uses Ollama with ggml quantized models (8-bit) and caching of embeddings & retrieved passages to reduce repeated work. For demonstration measurements, the LLM served via a local backend (Ollama runner) with explicit request timeouts and throttling. Safety layers include prompt-scoping, injection filters, and deterministic citation generation for all legal outputs.

C. Datasets and Knowledge Engineering

For reproducibility and clarity, we summarize dataset sources, sizes, annotation procedure, and splits used for training/evaluation. Where data cannot be publicly released (privacy/legal constraints), representative samples and data schemas are provided in the project repository.

Phishing URL dataset (URLs): 20,000 URLs (10,000 labelled malicious, 10,000 benign) aggregated from PhishTank and Google Safe Browsing feeds and augmented by mmCERT local reports collected between 2019–2024. Malicious labels use community/verifier labels (PhishTank) and cross-checked with VirusTotal reports. Train/ Val/ Test split: 70/15/15 stratified by time and TLD. Source APIs: PhishTank (API) and Google Safe Browsing.

Spam / scam message corpus: 12,000 short messages collected from public complaint boards and volunteer contributions with manual annotation (three annotators; majority vote). The dataset includes mixed-script Burmese (Unicode and Zawgyi examples), English-Burmese code-switching, and social-media artifacts. Split: 70/15/15.

Privacy policy corpus: 120 full privacy policies (mobile apps and web services) double-annotated for key facts; evaluation uses a gold-standard of human summaries (two annotators and adjudicator).

Malware / file hash dataset: ~10,000 SHA-256 hashes collated from VirusTotal historical reports and curated sample files (known malware families and benign files). Hash lookups were performed via the VirusTotal API.

Breach verification dataset: For k-anonymity experiments, we used anonymized prefix hashes and synthetic samples derived from HaveIBeenPwned API patterns; no raw credentials were stored. HaveIBeenPwned API and methodology informed the k-anonymity design.

Annotation & quality control: Class labels were double or triple annotated; inter-annotator agreement (Cohen's κ) exceeded 0.78 for primary tasks. For classifiers, we performed stratified sampling across dialectal and mixed-script inputs to ensure representative evaluation.

D. Security Tools, Consent, and Hosting Architecture

Analyses run locally with optional trusted services, following privacy by design, minimal permissions, tamper detection, user control, and transparent consent. Sensitive data is never stored persistently, and external components operate statelessly with scoped API keys and opt-in telemetry. A hybrid deployment hosts interfaces on secure clouds while inference and retrieval execute in-browser to uphold data sovereignty. Implementation employs modern web frameworks, transformer models, lightweight NLP for phishing and spam detection, and cryptographic primitives, with selective APIs for translation, breach checks, and malware intelligence.

E. Sustainability and maintenance plan

To remain operationally current, the system employs three coordinated update mechanisms: (1) a weekly crawler that ingests new legal and regulatory texts, (2) modular adapters that synchronize threat-intelligence feeds (PhishTank, Google Safe Browsing, VirusTotal) with provenance tracking, and (3) a CI/CD pipeline that reindexes embeddings and validates knowledge assets through regression testing. Legal updates triggering compliance changes undergo editorial review and committee approval. The architecture supports hot-swappable quantized models

for low-resource clients and staged rollouts for major LLM updates to mitigate regression risks.

VII. Experimental Results

A. Experimental Setup

Experiments were conducted on Chrome 126 (MV3) under Windows 11 with an Intel i7-12700H and 16 GB RAM. Spam and phishing classifiers combine TF-IDF features with a DistilBERT fine-tune using AdamW, learning rate 2e-5, batch size 16, three epochs, and max length 256. URL-based detection employed a Random Forest with 100 trees, max depth 12, and five-fold cross-validation. Privacy-policy summarization fine-tuned a summarization head for three epochs at 1e-5 learning rate using hybrid extractive/abstractive scoring. All models used 70/15/15 train/val/test splits, reporting macro-averaged Precision, Recall, F1, Accuracy, 95% bootstrap confidence intervals, and empirical p50/p95 latency over ≥300 trials.

B. Quantitative Results

Unicode normalization increased F1 scores by 2.7–3.9 points for spam and phishing detection. Code-switching decreased Precision by ~1.1 points; context-aware RAG prompts recovered 0.8–1.3 points. Out-of-distribution phishing kits showed URL lexical entropy and subdomain depth features contributed $\Delta F1 \approx +2.1$. Calibration via Platt scaling achieved ECE < 0.06; operating thresholds were chosen to maintain ≤2% false positives in safety-critical contexts.

C. Accuracy Evaluation

Table I summarize performance across modules. Spam and phishing detection was evaluated on mixed Myanmar/English corpora with curated local scams and out-of-distribution kits; thresholds were optimized to maximize F1 under false-positive rates ≤2.5%. The Privacy Policy Summarizer was evaluated on 120 policies with gold-standard, double-annotated summaries; F1 reflects the key-fact extraction overlap. Cyberlaw RAG outputs were assessed for answer equivalence with bilingual citations.

TABLE I. ACCURACY TESTING

Methods	Prec (%)	Rec (%)	F1 Score (%)	Acc (%)
SpamDetector	94.2	92.8	93.5	93.1
PrivacyPolicySummarizer	91.7	89.4	92.1	90.7
CyberLaw Training	88.9	87.2	89.5	88.3
FileAnalyzer	86.4	85.1	87.8	86.4
Phishing Detection System	89.5	88.0	90.1	89.1
Data Breach	93.4	92.7	94.1	93.4

D. Latency and Reliability

Experiments were conducted on Chrome 126 (MV3) under Windows 11 (Intel i7-12700H, 16 GB RAM) with 200±20 ms RTT emulated WAN conditions. Each module handled up to 6 concurrent requests; performance was measured over N≥300 trials, including parsing, formatting, and UI update. Medians (p50) and tail (p95) latencies are reported in Table II.

To reduce perceived wait and enhance trust, we implemented progressive results showing passage titles and partial summaries immediately, streaming with token-by-token rendering for long answers, prefetching common legal topics and prior queries, micro-interactions with actionable tips, and a fallback providing short, citation-anchored summaries if the full LLM response times out.

TABLE II. LATENCY AND RELIABILITY

Module	p50 (s)	p95 (s)	Success (%)
Legal Query (RAG Chat)	2.3	3.8	94.7
Phishing URL Detection	1.5	2.7	96.3
Data Breach (k-Anon)	1.8	2.2	98.9
SpamDetector	0.23	0.41	99.4
Privacy Summarizer	2.9	4.6	95.1
File Hash (SHA-256)	0.90	1.20	99.6

E. Resource Footprint

Table III shows CPU and memory usage per module. Hashing is streamed with chunked reads (1–4 MB working buffers), resulting in bounded CPU usage and sublinear memory scaling. The RAG chat UI exhibits peak memory during long bilingual responses and citation rendering.

TABLE III. BROWSER RESOURCE FOOTPRINT

Module	CPU (%)	Mem (MB)	Peak Mem (MB)
RAG Chat UI	10 ±1	180	260
Phishing Checker	7 ±1	110	150
Spam Detector	4 ±1	95	120
Privacy Summarizer	9 ±1	170	240
File Analyzer (hash)	8 ±1	130	170

F. Throughput and Efficiency

Table IV summarizes throughput (transactions per second, TPS), external API calls per operation, and cache hit rates. Measurements were performed with six concurrent requests on a single tab. Warm cache significantly improves p50 latency and reduces API calls.

TABLE IV. MODULE THROUGHPUT AND CACHE EFFICIENCY

Module	TPS (ops/s)	API Calls/op	Cache Hit (%)
Spam Detector	12.5	0	100
Phishing Checker	2.4	2-4	61
RAG Chat	0.5	0-1	54
Privacy Summarizer	0.35	1	18
Data Breach	1.1	1	72

G. Quality and Safety Indicators

Table V reports false-positive rates, grounded citation coverage, summarizer factual errors, and privacy preservation metrics. Classifier metrics are macro-averaged, with 95% confidence intervals from 10,000 bootstrap replicates. Summarizer factuality was verified against human-curated key facts by a second reviewer.

TABLE V. QUALITY AND SAFETY INDICATORS

Indicator	Value	Notes
Spam FP rate	2.1%	Sensitivity tuned for safety
Phishing FP rate	1.8%	Safety threshold 2–4
RAG grounded-citation coverage	93%	0–1 citation missing per query
Summarizer factual error rate	4.6%	Key-fact overlap
Privacy/API data exfiltration	0	No PII transmitted

- *Spam FP rate / Phishing FP rate*: false positive rate computed as FP/(FP+TN) on the test set using the operating threshold used in the study (thresholds chosen to keep FP ≤ 2.5% in safety-critical modes).
- *RAG grounded-citation coverage*: percent of queries where the final LLM answer contained at least one retrieved passage of citation that matched a human-verified ground truth passage (measured by overlap and adjudication).
- *Privacy/API data exfiltration*: binary indicators measured via static analysis and logs to confirm whether any PII fields (email, phone, full credentials) were transmitted to external services during the test harness.

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

Konbaung AI is a privacy-first, multilingual cybersecurity co-pilot tailored for Myanmar. The system integrates legal intelligence, threat detection, and user education into a browser-embedded platform that preserves data sovereignty. Evaluations demonstrate high accuracy in phishing, spam, and breach detection, robust latency and resource efficiency, and effective privacy policy summarization in Burmese. By combining RAG-powered legal reasoning, AI-driven threat analysis, and localized datasets, Konbaung AI empowers users to identify risks, understand their implications, and maintain safer digital practices.

B. Future Extensions and Research Directions

Future extensions will focus on expanding accessibility, resilience, and coverage, including:

- Deploying an offline-capable LLM chatbot for low-connectivity environments.
- Supporting multimodal and voice-based interactions for Burmese and ethnic dialects.
- Enhancing cross-platform integration and partnerships for real-time scam alerts, expert validation, and fraud detection.
- Incorporating continuous learning and multimodal threat analysis to maintain alignment with evolving cyber threats.

REFERENCES

[1] Chrome Developers, "Extensions Documentation (Manifest V3)," Available: https://developer.chrome.com/docs/extensions/,

[2] Plasmo Framework, "Build Browser Extensions with Plasmo," Available: https://docs.plasmo.com/

[3] LangChain, "LangChain Framework for Python and JavaScript," Available: https://python.langchain.com/ and https://js.langchain.com/

[4] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP," arXiv:2005.11401, 2020.

[5] Cohere, "Summarize API," Available: https://docs.cohere.com/docs

[6] VirusTotal API, https://docs.virustotal.com/reference/overview

[7] HaveIBeenPwned, "API v3," Available: https://haveibeenpwned.com

[8] FastAPI, "FastAPI Documentation," https://fastapi.tiangolo.com/

[9] OWASP, "Phishing Prevention Cheat Sheet," Available: https://cheatsheetseries.owasp.org/

[10] ENISA, https://www.enisa.europa.eu/publications/cyber-hygiene

[11] MITRE, "ATT&CK," Available: https://attack.mitre.org/

[12] mmCERT, "Advisories and Alerts," https://www.mmcert.org.mm/

[13] Google Cloud Translation: https://cloud.google.com/translate/docs

[14] Ethers.js, "Documentation," Available: https://docs.ethers.org/

[15] DataReportal, "Digital 2024: Myanmar," *Kepios Global Digital Insights*, Jan. 2024. Available: https://datareportal.com/reports/digital-2024-myanmar

[16] Anti-Phishing Working Group (APWG), "Phishing Activity Trends Report 2023", Available: https://apwg.org/trendsreports/

[17] PhishTank, "PhishTank Developer API Information", Available: https://phishtank.org/api_info.php. Accessed: Sep. 11, 2025.

[18] Google Myanmar-Tools, "Zawgyi Detector and Converter," *GitHub Repository, Available*: https://github.com/google/myanmar-tools.

[19] Off-the-Hook, "Client-Side Phishing Detection Browser Add-on" *Aalto University Technical Report*, 2022. Available: https://aaltodoc.aalto.fi/bitstreams/8635e4c1-fe27-46c1-afa6-7d21db15f20e/download

[20] PhishLang, "Client-Side LLM-Assisted Phishing Detection," *arXiv preprint* arXiv:2408.05667, 2024.

[21] Information Commissioner's Office (ICO), "International Transfers – A Guide", Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/international-transfers/international-transfers-a-guide/