International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Cyber Kavach: GENAI-Powered Multi-Model Cyber Threat & Legal Aid System

DEVARAPALLI SEKHAR BABU[*1], TADIBOYINA REVANTH[2], CHAMARTHI ANUSHA[3], NUTHALAPATI JYOTHIRMAEE[4], MEDA NAGA MOWRYA [5]

[1]*Assistant Professor, Department of CSE(CB), Bapatla Engineering College, Bapatla 522101, AP,India*
[2]*Student, Department of CSE(CB),Bapatla Engineering College, Bapatla 522101,AP, India*
[3]*Student, Department of CSE(CB), Bapatla Engineering College, Bapatla 522101, AP, India*
[4]*Student, Department of CSE(CB), Bapatla Engineering College, Bapatla 522101, AP, India*
[5]*Student, Department of CSE(CB), Bapatla Engineering College, Bapatla 522101, AP, India*

***Abstract*—** Cyber Kavach is a multi-modal Cyber Threat Detection and Legal Aid System, utilizing generative AI (GenAI) capabilities and built for the current increase in the number of people using (and misusing) the Internet; particularly, in India where the Internet is growing rapidly and infamously diversified. The rise of increasingly sophisticated cyber threats like phishing, financial fraud, impersonating attacks, and malicious/harmful links that can occur online. However, existing detection systems do not provide actionable guidance based upon contextual intent for vulnerable digital users. This proposal describes an advanced intelligent, full-stack web application designed for this purpose. The application will use Large Language Models (LLMs) like HuggingFace's Mistral7B to analyze all suspicious input (input that may be associated with known malicious activity) including text messages, URL's, images, and documents – without requiring users to have any technical knowledge. In addition, the system will utilize an 11-point, custom-built risk scoring system that evaluates characteristics of a threat (urgency of the associated language, suspiciousness of the associated domain, request for OTP etc.; impersonation patterns); and will generate two types of outputs for each cyber threat detected: 1)an interpretable risk score to represent the level of risk associated with the threat and a clear description of the threat in plain English. Unlike existing threat-detection systems, Cyber Kavach includes a unique legal assistance module that will automatically map detected threats to provisions of the Information Technology Act, 2000 and the Indian Penal Code, and generate structured, ready-for-filing FIR drafts in at least 3 languages (English, Hindi, and Telugu) to help eliminate a significant barrier experienced by victims when attempting to report cybercrime. The Cyber Kavach Platform employs a three-tier architecture with a React frontend, a FastAPI backend, and secure AI integrations while protecting user data by performing all AI interactions on the server, providing users with a simple and expert user interface. The platform has been tested with real-life scam situations and has been able to identify threats accurately and provide timely responses, thus proving to be an effective and scalable method of improving cyber safety. By providing real-time threat detection, multilingual capabilities, and automated legal assistance, Cyber Kavach allows users to successfully identify and respond to cyber threats, which helps build a more informed and resilient digital community

***Key Words*— GenAI, Cyber Threat Detection, Phishing Detection, Multi-Modal Analysis, Risk Scoring System, Legal Aid Automation, FIR Generation, IT Act 2000, Natural Language Processing, Cybersecurity Awareness.**

II. INTRODUCTION

The explosion in digital technology and online access has fundamentally changed the way people communicate, conduct transactions, and access services—especially in India, which has one of the fastest growing populations of internet users in the world. At the same time, however, the increase in digital services has corresponded with a huge surge in cybercrime: phishing, financial fraud, identity theft, social engineering scams, etc. Many users, especially rural and semi-urban users, do not have sufficient awareness and/or technical knowledge to detect and respond to cybercrime adequately. Criminals commonly use platforms such as WhatsApp, SMS, email, and malignant websites—or sometimes even visit users in person—to impersonate a trusted entity (e.g., bank, government agency, or service provider) and utilize fear and urgency tactics to deceive victims into providing personal information (e.g., account numbers, bank identifiers, or passwords) or allowing access to devices. When cybercriminals are successful in tricking someone into providing such information, victims invariably suffer financial losses, experience distress, and in some cases become the targets of future cybercrime. As mentioned above, many of the different safeguards currently available on the market use either a traditional rule-based or keyword detection system. Most of these systems are insufficient to handle the contextual, continually evolving natures of today's cyberattacks. Many of the commonly used detection systems provide only nominal threat classification without meaningful explanation or guidance for the user seeking assistance in determining their next steps upon realizing they may have been a victim or target of cybercrime.

The absence of integration between cyber threat detection and legal assistance represents a significant gap in the current ecosystem. Victims of cybercrime frequently lack the knowledge of how to file a report, what laws might apply to them, and how to correctly create a formal complaint for presentation to the appropriate authorities. Filing a complaint on sites such as the national cybercrime portal can be a daunting task for many victims, particularly those without

sufficient legal knowledge to navigate this type of process or without adequate command of English. Furthermore, the majority of current platforms and tools are created for use by English-speaking users only, thus eliminating an entire population from accessing tools that provide assistance with the world of cybersecurity. Therefore there is an immediate demand for a technological platform that is both user-friendly and accessible, capable of detecting cyber threats in a rapid, accurate manner, and also capable of providing users with actionable data and legal support.

This project proposes Cyber Kavach which uses GenAI multi-modal cyber threat detection as well as legal aid capabilities utilizing advanced large language model analysis (LLM) to determine suspicious content in real time. Unlike traditional systems that only look for pre-defined patterns of bad content, Cyber Kavach is designed to evaluate intent, tone, and context of messaging (text), hyperlinks (URL), screenshots (JPEG), and documents (PDF) through secure backend processing to provide a fully integrated risk assessment with each request submitted by users. In addition, one of the most significant innovations in this proposal is the scoring methodology employed in Cyber Kavach because it evaluates 11 different factors such as urgency indicators, domain name association, impersonation attempts, etc. to provide a fully coherent and accurate risk score for each submission and an easy-to-understand explanation that any user can understand regardless of their technical background.

In addition to threat detection, CyberKavach provides added features to serve as a personal legal assistant for victims of cybercrime. CyberKavach automatically links detected threats to the applicable sections of the Information Technology Act, 2000, and the relevant provisions of the Indian Penal Code, giving users an understanding of their rights in regard to the breaches. For increased user ease, CyberKavach automatically creates a structured, ready-to-file First Information Report (FIR) draft in English, Hindi, and Telugu, making the filing of formal complaints easier by reducing complications and hesitations.

CyberKavach is built using a modernized, three-tiered architecture using a React-based front-end, a FastAPI back-end, and integration between the Mistral 7B language model and HuggingFace to enable scalability, performance, and data and information security. All AI processing is performed server-side to protect the sensitive user information and API keys.

CyberKavach is an all-encompassing, social-impact-based solution that connects cyber threat detection with legal guidance. Using AI, multilingual accessibility, and user-based design, CyberKavach is designed to facilitate cybersecurity use by all people, irrespective of their technical skill sets or primary language. Through the implementation of CyberKavach, users will be able to identify and prevent cyber threats while possessing the information and tools to take effective legal action against the perpetrators, thus supporting a safer and more resilient digital environment.

III. LITERATURE SURVEY

The increasing occurrence of cyber crimes has resulted in many different types of cyber threat detection solutions being developed. Most of these solutions use older style machine learning and rule-based approaches like Naive Bayes, SVM and signature-based filtering techniques. These solutions often employ a signature-based approach based on known patterns of spam or malicious activities. They have proven successful against basic threats but often will not provide advanced detection capabilities for (i.e.) sophisticated phishing and social engineering scams that are generated on a dynamic basis and use contemporary technology. Additionally, these types of solutions cannot understand the intent of the messages contained in them, which means that they generally cannot effectively address advanced cyber threats that exist in today's cyber environment. [1].

Artificial intelligence has seen a significant move forward in recent years. It has been particularly evident in natural language processing (NLP) and large language models (LLM). These advances have resulted in a greater ability for systems to process contextual and semantic information related to text-based data. AI-based models can now recognize many subtle patterns in the way language is used in messages, i.e. urgency, impersonating, and emotionally manipulating. These types of intelligent systems are much more accurate and adaptable in their ability to detect and identify cyber threats than traditional approaches. However, the majority of these solutions only provide detection and classification functions and do not offer solutions that can help end users understand or accurately respond to the threat. [2].

Many researchers have begun looking at various ways of detecting cyber threats using different types of information (textual data, images, documents, web URLs) as multi-modal systems work together to produce an accurate assessment of a potential threat. Combining multiple types of data will provide higher accuracy through the ability to compare and contrast different formats and identify any inconsistencies across those different formats within a single dataset. Many currently available multi-modal systems still require significant amounts of computing resources and have not been designed for use in real-time or user-friendly applications. Furthermore, most multi-modal systems do not include integration into everyday tools that could help guide a user after the initial detection has occurred, thereby limiting the usefulness of these systems in the real world. [3].

Research into user knowledge and education related to cybersecurity has increased over the years as researchers realize that it is equally important for those affected by cybercrime and want them to be able to utilize available security resources (software, apps, etc.) to protect themselves. Researchers have shown there is a difference between someone who has been a victim of cybercrime reporting the

incident vs. those who do not report the incident; those who do not report are typically unaware, have difficulty finding someone to assist them with reporting the incident or difficulties understanding how to complete the necessary legal steps to report. There are resources available (forums) for individuals to gain an awareness of how to protect themselves or report a cybercrime; however, these resources are not automated (e.g., assist users with the process to report a cybercrime, provide users with a structure or template to create a legal complaint). The lack of an automated solution that combines awareness, identification of threats, multilinguism and legal support demonstrates the need for an integrated solution such as Cyber Kavach to fill the gap [4].

In addition, researchers have been paying increased attention to the possibility of automating legal processes through the use of artificial intelligence combined with legal informatics. Systems that are used to collect data and generate legal documents or make automated decisions regarding legal matters have become much more prevalent than in the past; whereas, systems used for generating legal documentation and decisions are often designed for specific legal processes but not specifically cybersecurity; additionally, these systems may not provide relevant legal information/data according to Indian legislation (e.g., Information Technology Act, 2000) to assist individuals when they become victims of cybercrime. The lack of a security solution that provides the ability for cyber threat detection, multilingualism and automated legal assistance is an example of why an integrated solution such as Cyber Kavach is necessary to fill this significant gap [5].

IV. EXISTING SYSTEM

Current methods used to detect cyber attacks, specifically phishing and malicious activities, utilize at least one traditional method (blacklist filtering, rule-based, or classical ML techniques). Blacklisting relies on previously identified poor URLs; it therefore does not account for new, zero-day threats. Rule-based methods require established rules based on things like the presence of suspicious keywords or certain URL structures and are unable to adapt to the evolving nature of cyberattacks. While Support Vector Machines (SVM), Random Forests, and Neural Networks have all shown promise with respect to improved performance by analysing different variables, such as URL structure, the contents of web pages, and the associated metadata, these models still rely heavily on the availability of high-quality training data and can be circumvented through the use of various advanced methods (e.g., obfuscated hyperlinks, dynamic web page content, delayed execution of an attack). Advances in Deep Learning and Natural Language Processing (NLP) also show promise due to the ability to identify contextual indicators in an email (e.g. urgency of action, impersonation, emotional manipulation), however these techniques generally require a lot of computational power, and therefore may not be useful for applications requiring real-time or user-friendly responses. In addition, most current systems only provide a simple classification of a URL as "phishing" or "safe". The lack of any clear rationale, guidance for users and any means of taking action further devalues the utility of these tools to the general population. In addition, there is no provision for reporting, obtaining legal support or assistance through these systems leaving the user completely unaware of any applicable laws, reporting processes or what they could do to protect themselves after being exposed to a cyber threat and data from these systems is typically targeted exclusively at English-speaking individuals resulting in lack of availability for the large, diverse population of India. Therefore, despite the advances in technology, the current solutions available do not combine all of the essential components (intelligent threat detection, contextual awareness, multilingual support and legal assistance) into a single, integrated solution such as Cyber Kavach is intended to do

V. PROPOSED SYSTEM

Cyber Kavach is a GenAI-enabled multi-modal cyber threat detection and legal aid service that provides a user-friendly, integrated solution for locating and responding to cyber crime. The user can access a single user interface via web-based application, which allows the user to evaluate potential threats (e.g. via text messages, URLs, screen shots, and/or documents) that they suspect may be suspicious in nature (e.g. by evaluating missing/altered signatures, unknown contacts or addresses, unsolicited email requests, OTP requests, and so on). Unlike traditional systems, Cyber Kavach uses large language models (LLMs) - specifically, Mistral 7B via HuggingFace - for contextual and intent-based analysis, rather than just pre-defined rules or keywords. The Cyber Kavach system will create a risk score based on an 11-factor custom risk scoring system to evaluate the risk of the identified threats, taking into consideration relevant parameters such as: urgency; impersonation; suspicious links; OTP requests; and/or financial triggers. The system will provide the user with a risk score and a detailed explanation of the risk assessment in an easy to understand format. The system will automatically map each detected cyber threat to specific provisions under the Information Technology Act, 2000 and the Indian Penal Code, thereby educating the user about their legal rights.

Additionally, Cyber Kavach will produce structured FIR drafts, ready to submit, in multiple languages, such as English, Hindi, and Telugu, making legal documentation easier for all victims. This System uses a three-tier architectural model that is based on React for the front end, FastAPI for the back end, and secure AI Integration. Because of this three-tier structure, the System is highly scalable and able to perform efficiently while protecting the privacy of users' data by running all of the AI Processing on the Server Side. The System has been designed with 2 different Modes of operation, Simple Mode for the general user and Expert Mode for detailed Technical Analysis of the data. By combining real-time, multimodal threat detection, multilingual support, explainable AI Output, and Automated Legal Assistance, Cyber Kavach provides a comprehensive but practical system to assist users in not only determining if they have been the victim of a cybercrime, but also to help

them take appropriate action after the identification of a cyber threat through informed and actionable steps

VI. SYSTEM ARCHITECTURE

1. Client Layer (User Interface)

The Client Layer is the section of the software responsible for how users interact with the system; it is developed using React.js. The Client Layer provides an intuitive and responsive interface, which allows users to input suspicious data in a number of formats, including HTML content, text, URLs, screenshots, and files. The application can be run in two different modes; Simple Mode for general users, and Expert Mode for users requiring a more sophisticated and detailed analysis. User requests from the Client Layer are made to the Back-End through the REST API using HTTP Requests (GET, POST, PUT, DELETE). Both modes provide a seamless user experience across devices.

2. API Gateway / Communication Layer

The API Gateway is the layer responsible for relaying user requests to the Back-End process. All user requests are sent to the API Gateway via secure methods (API Endpoints), using the Fetch API, which is built in as part of the Client Layer. The API Gateway provides the means by which to route requests to the appropriate process, validate request data and respond to requests. CORS Middleware accepts requests from the Client Layer to the Back-End using secure cross-domain authentication and authorization methods.

3. Back-End Layer (FastAPI Server)

The Back-End Layer consists of the FastAPI Server and contains all the core processing logic for the entire solution. The Front-End and the Back-End communicate with each other at the API Endpoints (/api/analyze, /api/legal, /api/speak) for processing and routing of user input. The Back-End Layer also guarantees that all sensitive operations on the server (e.g. AI interactions, data processing) will happen securely (i.e., without exposing any credentials to the Client).

4. AI Processing Layer (LLM Integration)

This AI Processing Layer will integrate the Mistral 7B Instruct model within HuggingFace Router API. It will conduct contextual-based analysis of the user's input, using custom prompt engineering to detect the user's intent. The application will also perform an evaluation of potential threats based on an 11-factor threat score and provide structured, clearly defined outputs in JSON format, including:

a. Threat Risk Score

b. A detailed explanation of how and why the user poses a potential threat

c. A Legal Mapping of the user's actions based on current laws defined within the country of origin.

5. Multi-Modal Processing Module

This system is designed to support multiple types of input through the use of specialized modules for processing:

- Text Analysis (Phishing Language, Urgency, Impersonation)

- Image Processing (Screenshots, OCR, Pillow for images)

- Document Processing (Text extractions, PyMuPDF and python-docx)

- URL Analysis (Domain Pattern Analysis and Phishing Indicators)

Each of these modules contributes to the overall ability to detect threats comprehensively using various forms of data.

6. Legal Assistance Module

The Legal Assistance Module will map the identified Cyber Threats to offenses within The Information Technology (Amendment) Act 2000, and the Indian Penal Code (IPC). The Legal Assistance Module will provide automatically generated drafts of FIRs in English, Hindi, and Telugu, thus allowing for the users to take legal action without needing prior legal knowledge.

7. Voice Output Module

This application uses gTTs to convert the results of analyses performed on access to legal reports from written text into audio so people with low levels of literacy and/or blindness can access the same content.

8. Security & Environment Management

All sensitive information (e.g. API keys) will be stored in a secure manner (stored with .env file) and will never be exposed on the Front End. All AI calls will be performed through the Back End in order to maintain secure connections. In addition, safe handling and validation of all files uploaded into the system will be required to prevent the upload of any malicious files.

9. Deployment Layer

The system will be deployed in a cloud environment through a series of hosted platforms such as Vercel (for Front End) and Render (for Back End) to ensure scalability, reliability and low latency response times when analyzing data in real time..

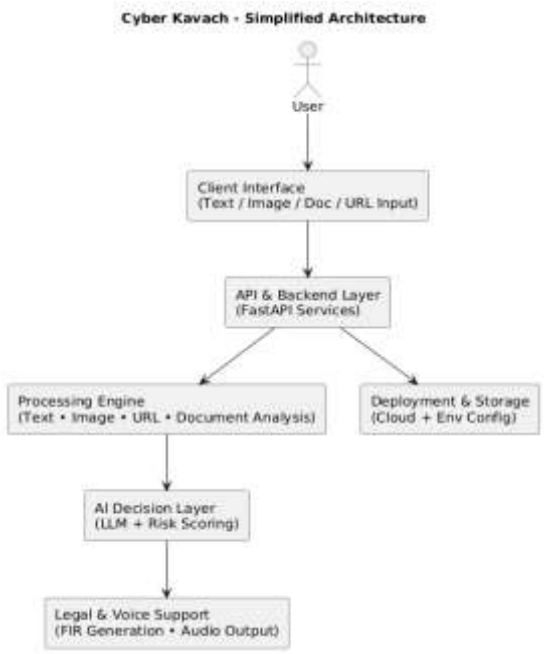

Figure 6.1 :System Architecture

VII. METHODOLOGY

1. Collecting Data and Input Processing

The system is able to accept many different types of user data including text messages, URLs, screenshots, or documents (PDF/DOCX). Users can paste any suspicious content or upload files via the web interface. This multi-modal capacity of the system allows it to deal with real-world threats on a variety of platforms like WhatsApp, email, and websites. All data is sent securely from the client's browser to the back-end server for further processing without exposing any user's private information on the client side.

2. Input Pre-processing and Content Extraction

When input data is received, it goes through a pre-processing step that extracts useful information from the input data. For example, in the case of text input, all extraneous characters and noise are removed to make the remaining text more readable. In the case of images, optical character recognition (OCR) techniques and image processing libraries are used to extract any embedded text from the image. In the case of documents, various tools are used, including PyMuPDF and python-docx, to parse the document and return any textual content of the document. In the case of URL input, the URL is also analyzed for domain-specific features and patterns. The pre-processing step ensures that all forms of input are transformed into a standardized format that can be analyzed.

3. Prompt Engineering and AI Interaction

In order for the Large Language Model (Mistral 7B) to act like a cyber threat analyst and legal assistant, it has to be guided by carefully written prompts that specify a specific form of behaviour. Different prompt types are created to process input types so as to provide better analysis that is accurate and considers context. Strict JSON output formatting is enforced for prompts as well; this provides for consistent structure and simple processing of results. The temperature parameter is also set low (e.g. 0.1) to yield consistent and replicable outcomes throughout all analyses.

4. Multi-Modal Threat Analysis:

Input that is processed through the AI model is then analysed for indications and evidence of a cyber threat, based on its context, intent, and linguistic patterns. Cyber threats are assessed based on a variety of indicators (e.g., urgent language, impersonation, suspicious URL, solicitation for sensitive information (e.g., OTP/PIN), and financial fraud) thus providing users the ability to identify advanced phishing attacks and social engineering attacks that may have been otherwise missed using traditional systems.

5. Risk Scoring Mechanism

To assist with quantifying the level of threat posed by the input, an 11-factor custom scoring rubric has been developed to assign a numerical (0-100) value to each of the input factors. Each factor (e.g., use of urgent language, Domain credibility, impersonation indicators, and malicious intent) has been assigned a weighted value within the scoring rubric for each input factor. The evaluation of all factors produces an overall threat score (0-100), as well as the related confidence level and detailed description of the threat risk value of the input. The structured scoring method enables users to quickly understand the magnitude of the threat.

6. Legal Mapping and Generation of FIRs

When a potential cybercrime has been detected through the system, the system will map the relevant provisions in the Information Technology Act, 2000 and Indian Penal Code against the identified potential crime. The mapping will facilitate the automatic generation of structured first information reports (FIRs) that will contain information about the incident, the type of offence and the relevant sections of legislation that apply to the event or situation in addition to the information collected during the investigation. In this way, a user can immediately take legal action against an individual for a cybercrime without first having obtained legal training.

7. Multilingual Output

To ensure the system can be of value to all users, the system can provide multilingual output in English, Hindi and Telugu so that users can receive their analyses and draft FIRs in their preferred language. The A.I. model generates output and drafts FIRs based on specific prompts given by users. The ability to provide such services ensures that a significant portion of users, including those who do not speak English, can effectively use the system.

8. Text-to-Voice Integration

The system is designed to utilise the gTTS module so that output from analyses and legal reports can be presented to users in audio format. This enables users with low literacy and/or visual difficulty to access output presented through text. It also enables users who are unable to read or write in

the English language alternative methods to receive the output of the system.

9. Backend Processing and Security of the AI Communication System

All AI Interactions and File Handling are done on the backend, using FastAPI. API keys and sensitive information are stored securely using Environment Variables (.env) on Back End and are never exposed to the Front End. This protects your data from unauthorized access, as well as your Credentials from being misused or your personal data from being compromised.

10. Testing and Validation of the AI Communication System

The AI Communication system is designed to provide real-world testing, such as Phishing/Spear Phishing Messages, Fraudulent Links, and Impersonation Attacks. Each Module from Input Processing to AI Analysis, as well as FIR Generation is tested using Unit Testing and End to End Testing Methods. In addition, all Performance Metrics, including Response Time, Accuracy of Threat Detection and Proper Mapping of Legalities were monitored and validated to confirm the Reliability and Effectiveness of the System.

VIII. IMPLEMENTATION

The Cyber Kavach System's overall design and development is done using Full Stack Web Applications comprised of the latest available technologies from Frontend, Backend, and Artificial Intelligence to provide an easy-to-use and safe user experience. The Frontend component is created using React.js, which allows for a responsive, user-friendly interface where, via a dynamic form and file upload capabilities with the FormData API, users can provide input for suspicious data in different formats, including text, URL, screenshot, and document. The Backend component is built using FastAPI in Python, which implements all the core functionalities of Cyber Kavach System through RESTful API endpoints such as /api/analyze, /api/legal and /api/speak. FastAPI also provides the ability to efficiently handle requests and communicate between components. Multi-Mode Data Processing is accomplished through the use of libraries such as PyMuPDF to extract data from PDF documents, python-docx to parse documents, and Pillow to manipulate images to convert different input formats into analyzable text. The intelligence to analyse these inputs comes from the Mistral 7B Instruct Model that is accessed from the HuggingFace Router API, where the system provides carefully designed prompts to direct the model to conduct contextual threat analysis, generate a risk score based on 11 separate factors, and return output in a structured, consistent format (JSON).This program has a legal aid feature that uses logic to connect detected threats with the corresponding sections of the Information Technology Act (IT Act) and the Indian Penal Code (IPC), while dynamically generating structured FIR (First Information Report) drafts in three languages (English, Hindi, and Telugu). The integration of Google Text-to-Speech (gTTS) provides audio output features for all results to be more accessible. The security of the program is protected by processing all AI interactions and sensitive operations solely on the backend, with API keys and configuration information being securely stored in environment variables (.env file) that are not exposed to the client side. Testing of the application was conducted using actual phishing and scam-type scenarios to test the ability to accurately detect threats, which was confirmed through exceptional results and consistent performance (average 3-5 seconds) , making the program a practical working solution that can be utilized in real-time to help people report cybercrime incidents

IX. RESULTS

1. Cyber Kavach Interface - Threat Input Overview

This is the user's main interface for Cyber Kavach and it allows users to input potentially malicious or suspicious data. The design accommodates several methods of input including: message, screenshot, file, and website (URL). This gives the system the flexibility to handle different types of user input in the real world, with the goal of improving overall user experience. Cyber Kavach also has a Simple Mode to provide less technical users with an easier way to use the product. The "Analyze Now" button will trigger an immediate analysis for the user.

Figure 9.1 : Threat Input

2. URL Analysis Input Overview

The The second screen of Cyber Kavach exists to show the results of analyses done by the system on possible phishing activity. It also permits a user to enter a URL for analysis. Once entered, the system analyzes the entered URL for risk to the end-user and provides suitable recommendations to the end-user prior to generating the final

Figure 9.2 : Analysis Input

3. Risk Detection Results - High Risk Warning

The final report generated by Cyber Kavach provides the user with a risk assessment score and the results from the system's analysis of the entered URL. For example, the system may assess an entered URL and assign it a score of 82 out of 100,

identifying it as a high risk phishing link, and provide evidence of that by referencing: (1) the TLD (xyz); (2) the use of urgent keywords; and (3) the use of HTTP instead of HTTPS. The system will also warn the user against interacting with the URL.

Figure 9.3 : Risk Detection

4. FIR Draft Generation Output

This part of the system generates a draft FIR according to the cyber threat identified through our software. The draft contains sufficient legal formatting (ie. subject, description of incident etc.) so that you can easily make your complaint with no legal experience on your own.

Figure 9.4 : Generation output

5.Risk Breakdown and Legal Mapping Output

This screen contains the complete breakdown of detailed risk score and legal mapping together to show the contribution of each factor to the final score along with any major red flags that may violate the law or regulations. It also gives the legal sections of the IT Act, 2000 that are applicable to the threat and suggested responses to each issue as well as generating a FIR to take action against the threat.

Figure 9.5 : output

X. CONCLUSION

Cyber Kavach has developed a robust and innovative answer to the increasing number of cyber-crime cases through the combination of state-of-the-art Generative AI-based Cyber Threat Detection and practical legal assistance, all within an integrated platform. Traditional definitions of Cyber Threat Detection have limitations, as they rely almost entirely on unilateral definitions (on rules and keywords) without factoring in actions being taken within context that could change the likelihood/intention of harmful activity. Cyber Kavach utilizes multi-modal capability to analyze various forms of input (including text, URLs, screenshots and documents) to provide users with accurate risk analysis of identified threats, along with clear and concise explanations. The application of an 11-factor risk rating system also aids in improving transparency and ease of interpreting potential threats based upon severity.

Another significant benefit of Cyber Kavach is its ability to help users beyond mere threat identification. By automating the linking of identified threats to applicable provisions of the Information Technology Act, 2000 and intending to create, where appropriate, structured FIRs in over twenty languages, Cyber Kavach significantly reduces existing barriers to victims of cyber crime to report incidents to law enforcement. The combination of an intuitive interface, multilingual capability and voice capability enhances accessibility to a broad spectrum of users, including those with lower levels of technological expertise and/or limited English proficiency; thus making Cyber Kavach not only technically sound but also socially responsible.

In summary, Cyber Kavach exhibits an effective combination of cyber security and computer science in conjunction with legal frameworks through individual user interfaces to create a highly functional solution. It has been proven through testing using multiple real life tests and has demonstrated its ability to provide accurate, immediate solutions to users in practical ways. By allowing users to identify, comprehend, and respond to cyber threats in an effective manner, Cyber Kavach will help to create a safer digital world and is a major advancement in cybercrime prevention that is accessible and intelligently designed.

XI. REFERENCES

[1] S. Kumar and R. Singh, “Cybercrime Detection Using Machine Learning Techniques,” International Journal of Computer Applications, 2019.

[2] A. Brown et al., “Natural Language Processing for Phishing Detection: A Deep Learning Approach,” IEEE Access, 2021.

[3] L. Chen and H. Zhao, “Multi-Modal Cyber Threat Detection Using Artificial Intelligence,” Journal of Cybersecurity, 2022.

[4] P. Sharma and A. Gupta, “Cybersecurity Awareness and Reporting Challenges in Developing Countries,” International Journal of Information Security, 2020.

[5] R. Mehta et al., “AI-Based Legal Assistance Systems for Automated Document Generation,” International Journal of Law and Technology, 2023.