UNIVERSITY OF PLYMOUTH

PUSL3190 Computing Individual Project

Project Proposal

AI-Powered Phishing Detection System

Supervisor: Mr. Chamara Dissanayake

Name: Ashen Abeysekara
Plymouth Index Number: 10899221
Degree Program: BSc (Hons) Computer Security

Table of Contents

Chapter 01 – Problem Statement

While looking for an idea for the project I read several articles, where many had phishing mentioned in some sort of way. And phishing has been here for a very long time. Also, with the rapidly growing technology phishing attacks have become much harder to detect. The Internet Crime Complaint Center (IC3) of FBI also reported that over $12.5 Billion was lost due to internet crime, and the leading problem of it was phishing by a huge margin. Initially most cyber-attacks start with phishing attacks as well, according to the CISA (Cybersecurity & Infrastructure Security Agency) “More than 90% of successful cyber-attacks start with a phishing email.” that alone tells us how big of a problem phishing is. Also, other sources like UK government, Sprinto, TechTarget, Terranova Security, Embroker, and KnowledgeHut also includes that phishing and AI & Generative AI phishing in their articles.

Because of these reasons traditional security measures are proving to be not good enough against modern phishing techniques for several critical reasons:

1. **Advanced Social Engineering**: Most modern phishing attacks have advanced psychological manipulation techniques that can bypass traditional security awareness training. This can be done through highly personalized messages using OSINT.
2. **Multiple Attack Ventors**: Although email is the most targeted attack vector for phishing, there are other attack vectors as well. Such as,
    a. Social media
    b. Text/SMS
    c. Professional networking platforms
3. **Dynamic Attacks**: Attackers rapidly change their infrastructure by,
    a. URL shorteners
    b. Temporary domains
    c. Compromised legitimate websites
4. **Limited Real-Time Protection**: Currently most solutions detect threats after the user has been exposed.

These problems create an urgent need for an intelligent, proactive and user-friendly security solution that can adopt to newer threats while providing real-time protection.

Chapter 02 – Project Description

2.1 Project Overview

The AI-Powered Phishing Detection System is a solution for a huge problem in the security field. This combines real-time browser protection and advanced AI to create a comprehensive defense against modern phishing attacks. This system takes multiple AI models working together to analyze different types of potential threats. This provides users with immediate protection and detailed analysis.

2.2 Project Objectives

By completing this project, I target to achieve these objectives,

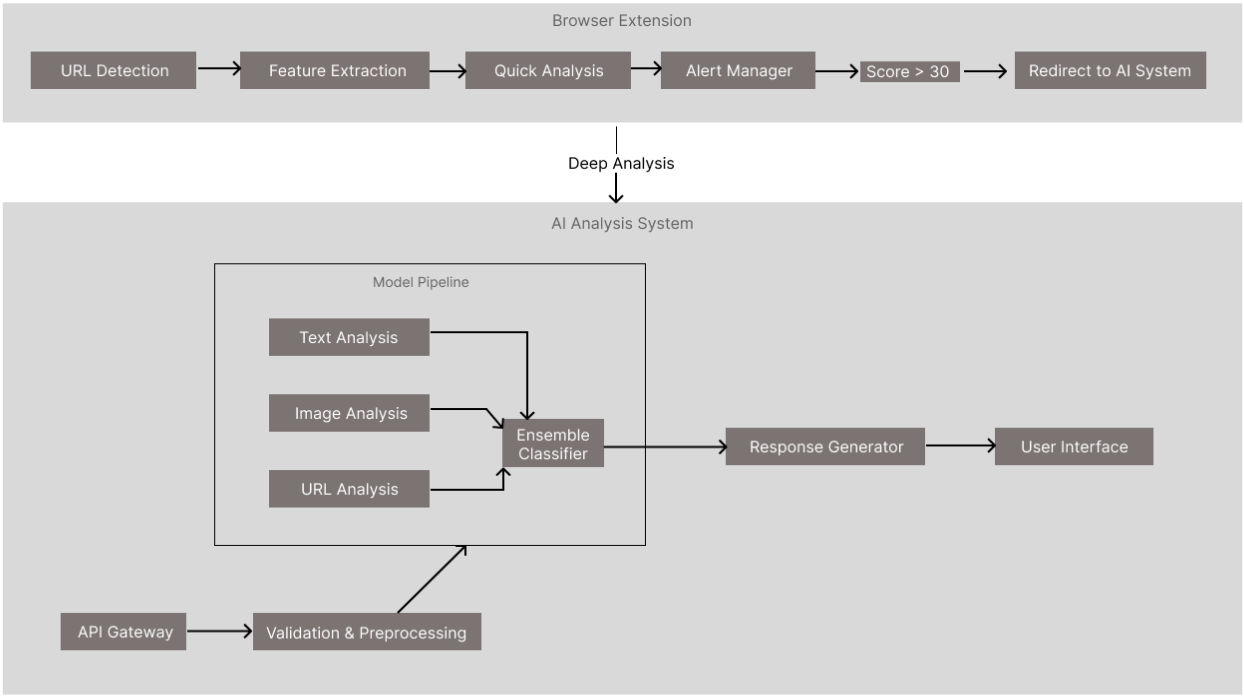
1. **Real-Time Threat Detection**
    - Develop a browser extension that monitors web activity in real-time.
    - Create an alert system for quick user notification.
    - Achieve detection speeds under 1 second for better user experience.

2. **Multi-Modal Analysis**
    - Build AI models capable of analyzing,
        - Text content using NLP.
        - Visual elements using computer vision
        - URLs using feature extraction and analysis
    - Implement ensemble learning for better threat assessment.

3. **User Education and Empowerment**
    - Develop an interactive chatbot interface for threat analysis.
    - Provide detailed explanations of detected threats.
    - Create educational resources for security awareness.

4. **Security Improvements**
    - Develop regular model updates for new threat patterns.
    - Develop flexible rule sets for upcoming attack vectors.
    - Build a scalable architecture for future improvements.

2.3 Technical Architecture



I designed this architecture using Figma if it is not clear you can view it with this link, https://www.figma.com/design/eW6QJJMp9RtpBvseSd3Rl2/AI-Powered-Phishing-Detection-System-Technical-Architecture?node-id=0-1&t=Mv97YhlR8bIN88Eq-1

2.4 Implementation Details

For this project I chose the currently most used technologies in the industry. So, this will be able to help me learn and power up my skills.

1. Frontend Technologies
    - React.js for web interface
    - Chrome Extensions API
    - WebSocket for real-time updates
    - Tailwind CSS for styling

2. Backend Technologies
    - Python for AI/ML components

    - FastAPI for REST endpoints
    - Redis for caching
    - PostgreSQL for data storage
3. AI/ML Components
    - PyTorch for deep learning
    - Transformers for NLP
    - Scikit-learn for traditional ML
    - OpenCV for image processing

2.5 Project Keywords

- Phishing Detection
- Open-Source Security Tool
- Real-Time Browser Protection
- Artificial Intelligence
- Natural Language Processing

Chapter 03 – Research Gap

The popularity of phishing attacks continues to be a significant concern in cybersecurity. Despite various advancements in detection technologies, multiple challenges persist that hinder the effectiveness of current phishing detection methods (Basit et al., 2020). The planned browser extension project aims to bridge the existing gaps with a focus on real-time analysis, multi-modal input capacity, and user interaction.

3.1 Current Challenges in Phishing Detection

Many existing phishing detection methodologies rely mainly on heuristic and rule-based systems that often struggle with new, innovative phishing tactics (Nadeem et al., 2023). For instance, traditional systems may prioritize elemental analysis of URLs, leading to high false positives when faced with novel phishing concepts that bypass established patterns. The proposed system's integration of a dynamic scoring mechanism capable of evaluating multiple parameters (including URL and image analysis) addresses this challenge head-on, providing more accurate and timely feedback (Basit et al., 2020) .

3.2 Emerging Trends in Phishing Techniques

As phishing attacks increasingly leverage generative AI to create misleading yet authentic-looking communications or websites, standard detection methods fail to keep pace with these advancements (Schmitt & Flechais, 2023). The adaptation of artificial intelligence into phishing tactics requires the development of more polished detection systems, like the proposed extension, which aims to continually learn from new data and user interactions, thus enhancing its detection capabilities over time (Eze & Shamir, 2024).

3.3 User Awareness and Engagement

The question of user education and engagement in recognizing phishing attempts remains largely unaddressed in current systems (Ansari et al., 2022). Many existing detection solutions alert users of potential threats but do not provide educational follow-ups or deeper analyses. This gap can lead to confusion among users, who may not fully understand the nature of the threat. The proposed project’s incorporation of an AI chatbot for immediate analysis and

education addresses this need for user interaction, empowering users to make informed decisions.

3.4 Personalization in Detection Mechanisms

Current phishing detection systems often operate in a one-size-fits-all manner, failing to consider individual user behavior or preferences (Basit et al., 2020). This lack of personalization can result in ineffective detection, as users may exhibit varying levels of vulnerability based on their behavior and tendencies. The proposed browser extension aims to analyze user behavior patterns alongside incoming threats, tailoring detection and alerts to the individual user, thereby enhancing the relevance and effectiveness of responses.

3.5 Real-Time Detection Capabilities

Timeliness of detection is a vital aspect of effective phishing prevention that many existing systems overlook (IEEE, n.d.). Traditional detection tools often analyze data post-factum, which can be too late for users who may have already been compromised. The introduction of a real-time scoring system that quantifies suspicious URLs adds significant value, allowing users immediate risk assessments as they navigate online spaces.

3.6 Integration of Multi-Modal Inputs

Existing research often focuses on singular forms of data, not remembering the importance of multi-modal analysis in phishing detection (Basit et al., 2020). By allowing inputs such as images, text, and URLs, the proposed AI provides a more comprehensive and nuanced approach to identifying potential threats, which is necessary given the evolving complexity of phishing schemes.

3.7 Challenges of Machine Learning in Phishing Detection

Despite the potential of machine learning to enhance phishing detection, many models currently in use suffer from limitations in training data diversity and adaptability (Eze & Shamir, 2024). The proposed project emphasizes employing machine learning techniques that can process diverse input forms and continuously evolve based on new data, thus addressing existing models' rigidity.

The proposed research aims to overcome these critical gaps through an innovative platform that integrates user experiences, real-time evaluations, multi-modal analysis, and educational assistance, ultimately enriching the landscape of phishing detection methodologies and significantly contributing to ongoing cybersecurity efforts (Schmitt & Flechais, 2023).

Chapter 04 – Requirements Analysis

4.1 Functional Requirements

There are three main functional requirements needed for this. Browser extension, AI analysis system, and user interface. I will point out each of its functional requirements.

1. Browser Extension
    a. Real-time URL analysis – When a user visits a website it will analyze the URL, because most phishing websites have misspelt words.
    b. Automatic threat scoring – This will produce a score from 0 to 100 depending on the URL analysis.
    c. Alert generation – This will have different notifications depending on the score of the threat scoring.
    d. Redirection to AI system – If the threat scoring is more than 30, then it will show a redirection to the user to check if it is a phishing attempt with the AI system.
    e. Settings customization

2. AI Analysis System
    a. Multi-modal input processing – To have different types of inputs, such as images, URLs and text.
    b. Detailed threat analysis
    c. Result explanation
    d. Historical analysis storage

3. User Interface
    a. Intuitive chat interface
    b. Multiple input methods – Images, text and URLs
    c. Clear threat indicators – Add colors according to the threat level
    d. Educational resources

4.2 Non-Functional Requirements

There are also three main non-functional requirements needed for this. Performance, Security, and Usability. I will point out each of its importance.

1. Performance
    a. Browser extension response time should be less than 1 second.
    b. AI analysis completion time should be within 3 to 10 seconds.
    c. 99% system uptime for availability.

2. Security
    a. Encrypted data transmission.
    b. Secure user data storage.
    c. Privacy-preserving analysis.

3. Usability
    a. Intuitive interface.
    b. Clear error messages, so it will be easier for the user.
    c. Responsive design.
    d. Accessibility compliance.

4.3 User Requirements

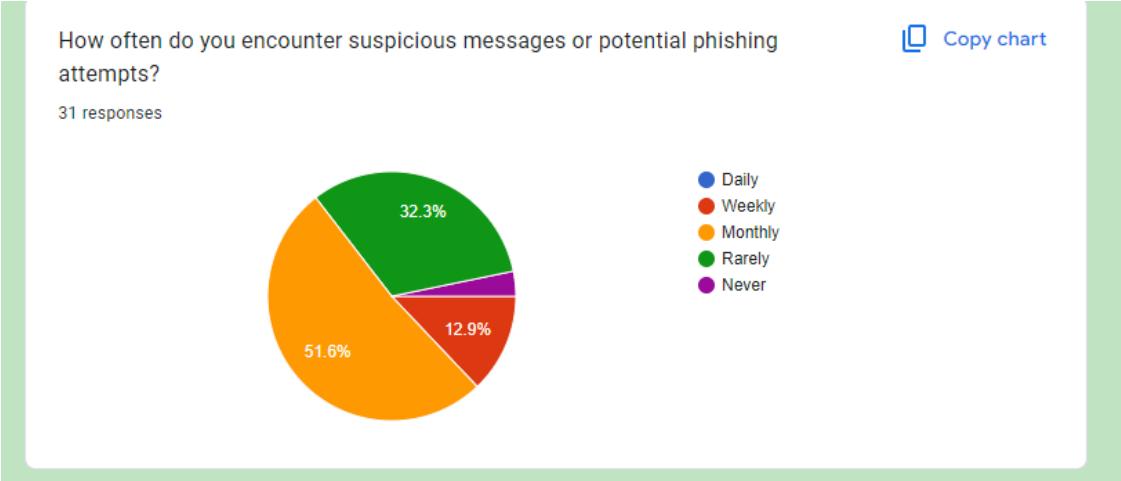
I also created a survey to collect user requirements for this project, the results of that survey are shown below from the screenshots. I will add the statistics of the top three responses from some questions. To view all the questions and statistics it can be viewed from the below link to the full screenshot or from the excel sheet,

Image link: survey responses.png

Excel sheet link: User Requirements Survey for AI-Powered Phishing Detection System (Responses).xlsx

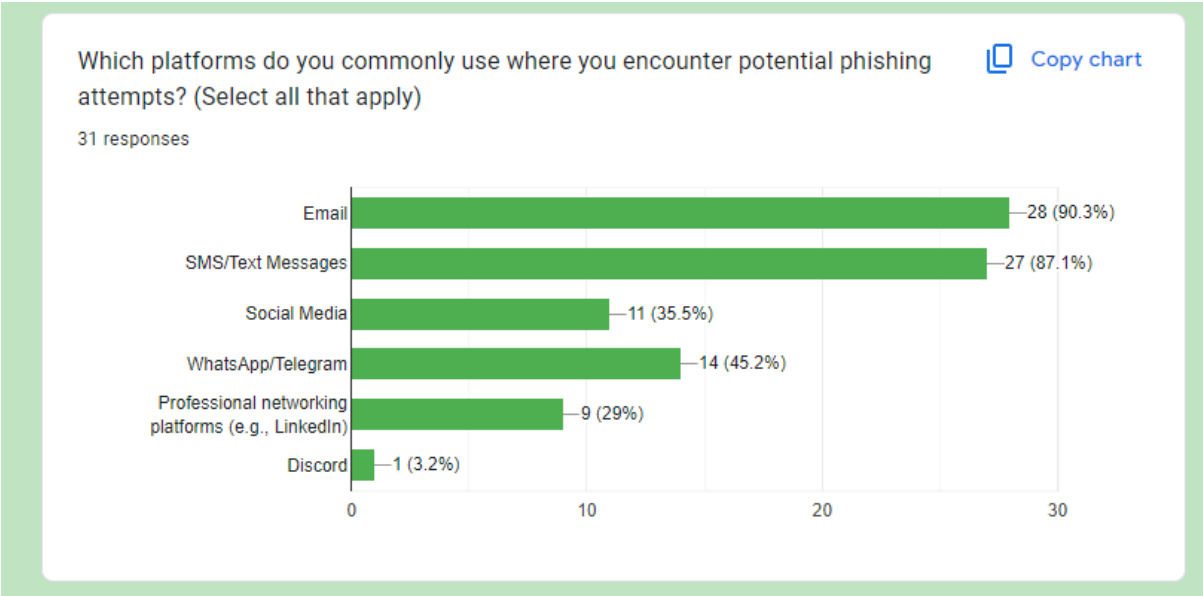
Also, the survey form can be viewed from this link,

https://forms.gle/z521WxCM7joxi9ch6



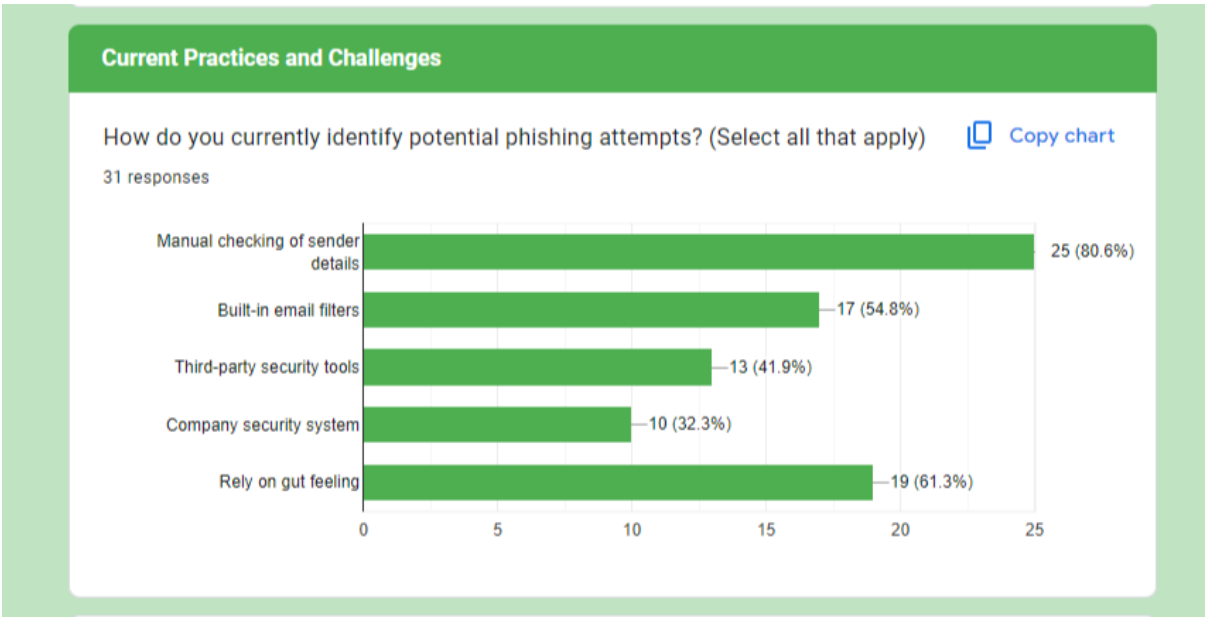
How often do you encounter suspicious messages or potential phishing attempts?

- Monthly: 51.6%
- Rarely: 32.3%
- Weekly: 12.9%



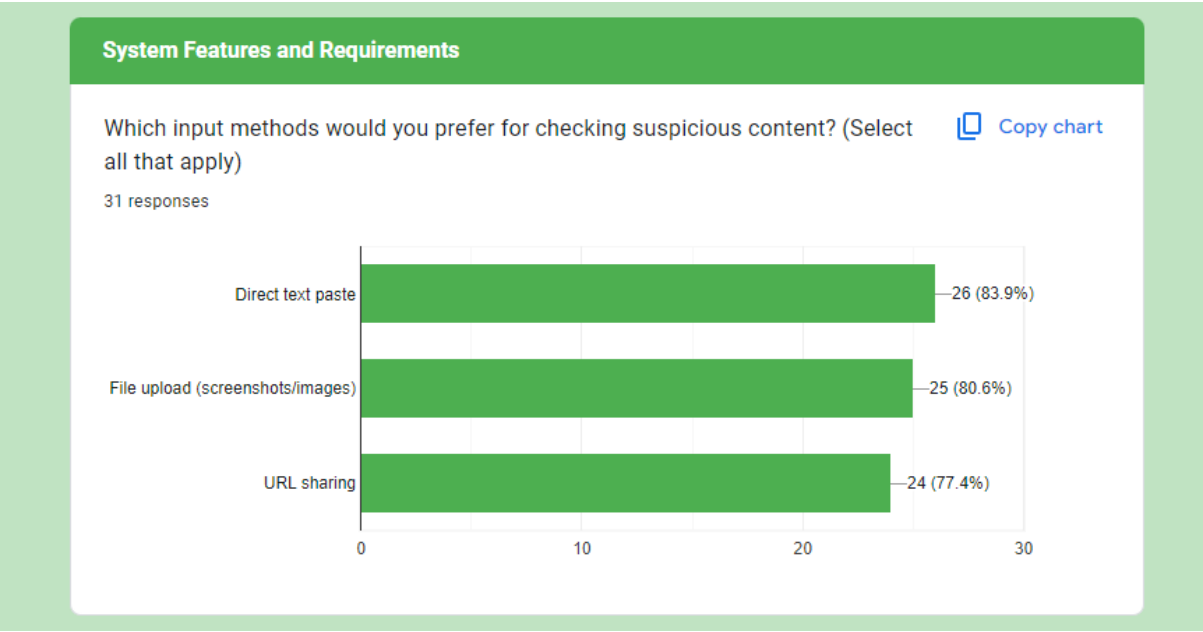
Which platforms do you commonly use where you encounter potential phishing attempts?

- Email: 28 (90.3%)
- SMS/Text Messages: 27 (87.1%)
- WhatsApp/Telegram: 14 (45.2%)



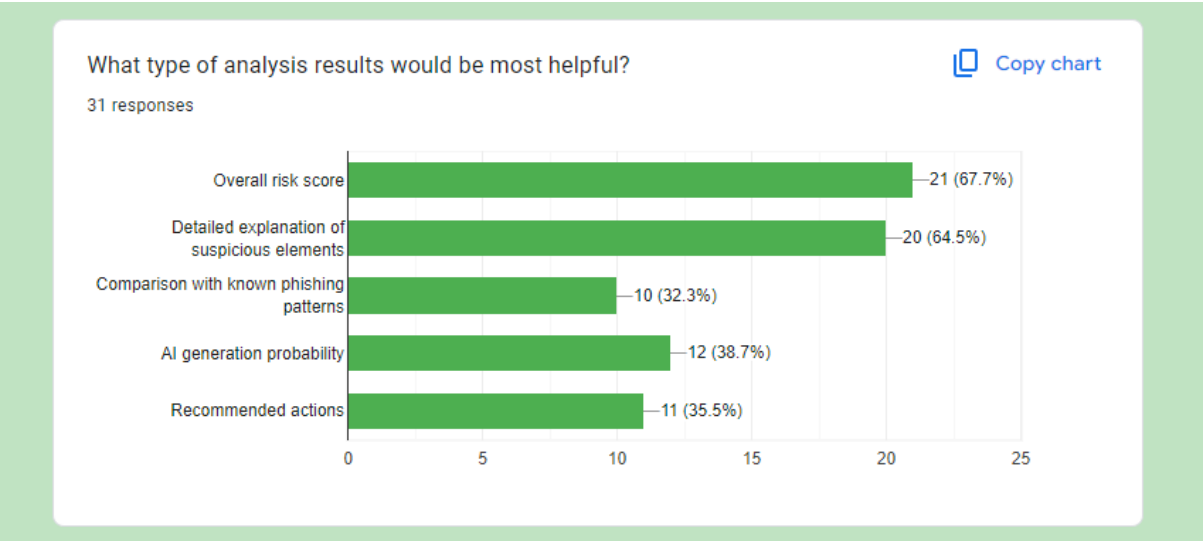
How do you currently identify potential phishing attempts?

- Manual checking of sender details: 25 (80.6%)
- Rely on gut feeling: 19 (61.3%)
- Built-in email filters: 17 (54.8%)



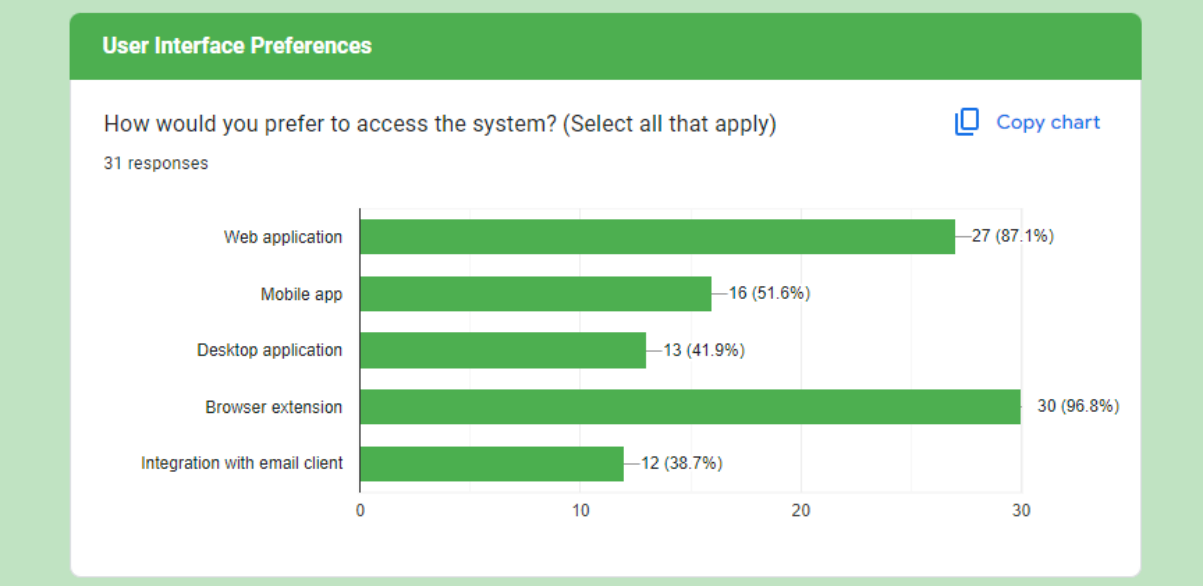
Which input methods would you prefer for checking suspicious content?

- Direct text paste: 26 (83.9%)
- File upload (screenshots/images): 25 (80.6%)
- URL sharing: 24 (77.4%)



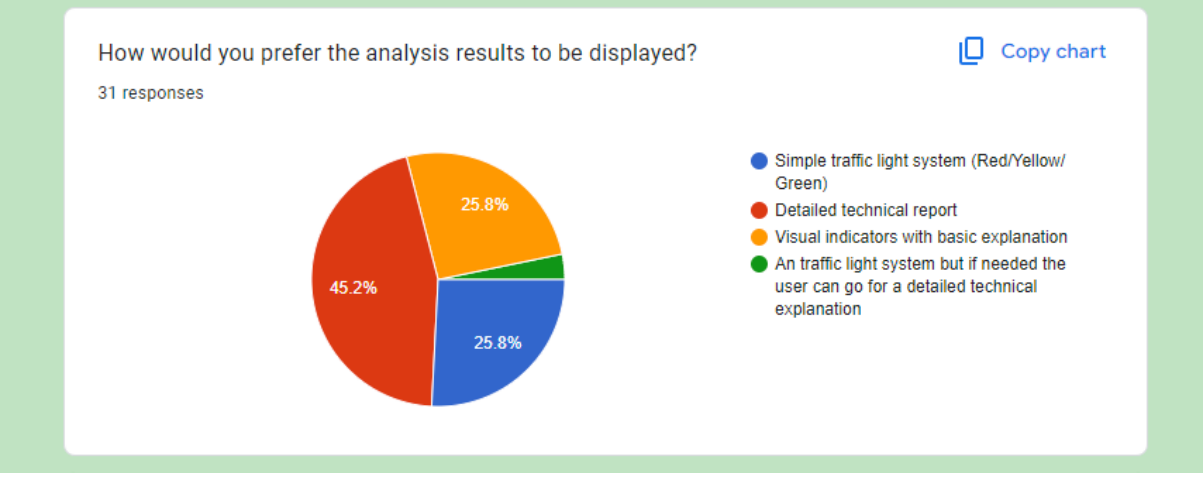
What type of analysis results would be most helpful?

- Overall risk score: 21 (67.7%)
- Detailed explanation of suspicious elements: 20 (64.5%)
- AI generation probability: 12 (38.7%)



How would you prefer to access the system?

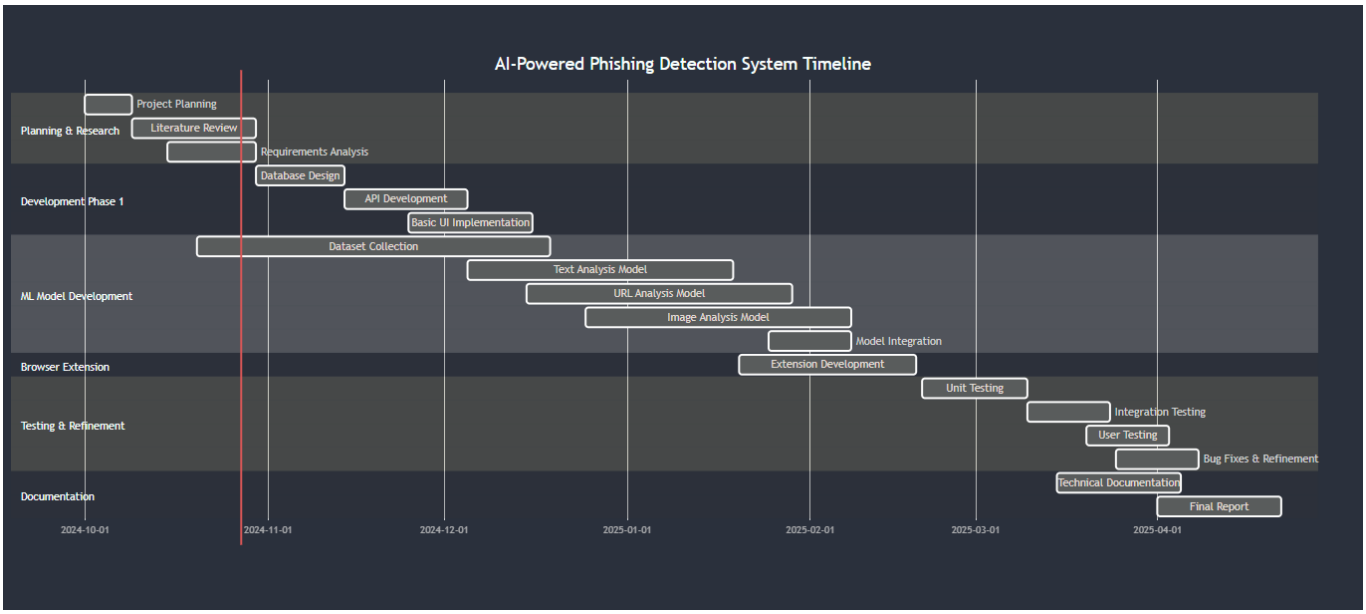
- Browser extension: 30 (96.8%)
- Web application: 27 (87.1%)
- Mobile app: 16 (51.6%)



How would you prefer the analysis results to be displayed?

- Detailed technical report: 45.2%
- Simple traffic light system (Red/Yellow/Green): 25.8%
- Visual indicators with basic explanation: 25.8%

Chapter 05 – Time Frame / Timeline



This Gantt chart was created using mermaid (https://mermaid.js.org). Also made an Gantt chart in excel to track the completion of the project. You can view it using the link below,
AI-Powered Phishing Detection System Gantt chart.xlsx

Details of the time frame if the Gantt chart is not clear,

1. **Planning and Research**
    a. Project planning – 8 days (from October 1st to 9th of 2024)
    b. Literature review – 21 days (from October 9th to 30th of 2024)
    c. Requirements analysis – 15 days (from October 15th to 30th of 2024)
2. **Development Phase 1**
    a. Database design – 15 days (from October 30th to November 14th of 2024)
    b. API Development – 21 days (from November 14th to December 5th of 2024)
    c. Basic UI implementation – 21 days (from November 25th to December 16th of 2024)
3. **ML Model Development**
    a. Dataset collection - 60 days (from October 20th to December 20th of 2024)
    b. Text analysis model – 45 days (from December 5th of 2024 to January 20th of 2025)

c. URL analysis model – 45 days (from December 15th of 2024 to January 30th of 2025)
d. Image analysis model – 45 days (from December 25th of 2024 to February 8th of 2025)
e. Model integration – 14 days (from January 25th to February 8th of 2025)

4. Browser Extension

a. Extension Development – 30 days (from January 20th to February 20th of 2025)

5. Testing and Refinement

a. Unit testing – 18 days (from February 20th to March 10th of 2025)
b. Integration testing – 14 days (from March 10th to March 24th of 2025)
c. User testing - 14 days (from March 20th to April 3rd of 2025)
d. Bug fixes and refinement – 14 days (from March 25th to April 8th of 2025)

6. Documentation

a. Technical documentation – 21 days (from March 15th to April 5th of 2025)
b. Final report – 21 days (from April 1st to April 21st of 2025)

Referencing / Bibliography

- **Federal Bureau of Investigation (FBI), 2023.** Internet Crime Report 2023. [pdf] Internet Crime Complaint Center (IC3). Available at: https://www.ic3.gov/AnnualReport/Reports/2023_IC3Report.pdf (Accessed 28th October 2024).
- Cybersecurity and Infrastructure Security Agency CISA. (n.d.). *Shields Up: Guidance for Families | CISA*. [online] Available at: https://www.cisa.gov/shields-guidance-families.
- Basit, A., Zafar, M., Liu, X., Javed, A.R., Jalil, Z. and Kifayat, K. (2020). A comprehensive survey of AI-enabled phishing attacks detection techniques. Telecommunication Systems, [online] 76(1). doi:https://doi.org/10.1007/s11235-020-00733-2.
- Schmitt, M. and Flechais, I. (2023). Digital Deception: Generative Artificial Intelligence in Social Engineering and Phishing. SSRN Electronic Journal. [online] doi:https://doi.org/10.2139/ssrn.4602790.
- Eze, C.S. and Shamir, L. (2024). Analysis and Prevention of AI-Based Phishing Email Attacks. Electronics, [online] 13(10), p.1839. doi:https://doi.org/10.3390/electronics13101839.
- Nadeem, M., Syeda Wajiha Zahra, Muhammad Noman Abbasi and Ahmed, W. (2023). Phishing Attack, Its Detections and Prevention Techniques. [online] ResearchGate. Available at: https://www.researchgate.net/publication/374848676_Phishing_Attack_Its_Detections_and_Prevention_Techniques.
- ieeexplore.ieee.org. (n.d.). PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System. [online] Available at: https://ieeexplore.ieee.org/abstract/document/9082616.
- Ansari, M.F., Sharma, P.K. and Dash, B. (2022). Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training. International Journal of Smart Sensor and Adhoc Network., 3(3), pp.61–72. doi:https://doi.org/10.47893/ijssan.2022.1221.

- Bhavsar, V., Kadlak, A. and Sharma, S. (2018). Study on Phishing Attacks. International Journal of Computer Applications, [online] 182(33), pp.27–29. Available at: https://www.researchgate.net/publication/329716781_Study_on_Phishing_Attacks.
- Alsharnouby, M., Alaca, F. and Chiasson, S. (2015). Why phishing still works: User strategies for combating phishing attacks. International Journal of Human-Computer Studies, 82, pp.69–82. doi:https://doi.org/10.1016/j.ijhcs.2015.05.005.