SafeClick Project Plan

Junyu Li, Shreya Kembhavi, Abdallah Alsharrah, Arvin Azad, Samantha Nainan

Hofstra University

4.11.2025

Project Plan

SafeClick is a secure, AI-powered web application that protects users from malicious links by leveraging a combination of machine learning, real-time phishing feature extraction, image-based OCR, and natural language generation. Designed as a digital security agent, SafeClick accepts both direct URL input and uploaded images (such as screenshots of scam texts or phishing emails), extracts links, evaluates them for threats, and provides users with a clear threat report and security recommendation. The backend is built with strong cybersecurity practices including a post-quantum-safe TLS layer and BB84-simulated quantum key exchange.

Deliverable

A web application that allows users to input a URL and receive a classification of whether the URL is suspicious or not

Front End UI created with React

ML model that will be trained using logistic regression

Market Potential

The phishing protection market is experiencing significant growth, driven by the escalating The phishing protection market is experiencing significant growth, driven by the escalating frequency and sophistication of cyber threats. In 2023, the market was valued at approximately $2.39 billion and is projected to reach $5.94 billion by 2031, reflecting a compound annual growth rate (CAGR) of 12.1% . This surge is attributed to the increasing

reliance on digital communication channels, the proliferation of remote work, and the adoption of cloud-based services, all of which have expanded the attack surface for cybercriminals.

Phishing attacks remain a predominant threat vector, accounting for 36% of all data breaches in the United States in 2023 . Alarmingly, approximately 3.4 billion phishing emails are dispatched daily, constituting about 1.2% of all email traffic . These statistics underscore the pervasive nature of phishing and the pressing need for robust protective measures.

The financial implications of phishing are substantial. In 2023, businesses suffered losses amounting to $2.9 billion due to sophisticated phishing schemes, including business email compromise attacks . These attacks often exploit human vulnerabilities, emphasizing the necessity for solutions that not only detect malicious URLs but also educate users about potential threats.

Given this landscape, there is a clear market opportunity for innovative solutions that can effectively identify and mitigate phishing threats. A web application capable of analyzing URLs for suspicious characteristics, providing confidence scores, and offering detailed reports can serve as a valuable tool for individuals and organizations alike. By addressing this critical need, such a solution has the potential to establish a strong foothold in the burgeoning cybersecurity market.

Project Overview

SafeClick is a secure, AI-powered web application that protects users from malicious links by leveraging a combination of machine learning, real-time phishing feature extraction, image-based OCR, and natural language generation. Designed as a digital security agent, SafeClick accepts both direct URL input and uploaded images (such as screenshots of scam texts or phishing emails), extracts links, evaluates them for threats, and provides users with a clear threat report and security recommendation. The backend is built with strong cybersecurity practices including a post-quantum-safe TLS layer and BB84-simulated quantum key exchange.

Problem Statement (Our Why?)

Phishing attacks continue to be one of the top causes of data breaches, and users often lack the tools or knowledge to safely evaluate whether a link is legitimate. With the increasing volume of malicious emails, messages, and SMS scams, there is an urgent need for a user-friendly, explainable, and secure method for real-time link analysis.

Our Solution:

SafeClick empowers users to verify the safety of links using both artificial intelligence and cybersecurity best practices. Users can input a URL or upload a screenshot of a message. The system extracts relevant features, classifies the link using a machine learning model trained on phishing patterns, and generates a human-readable threat report using a large language model (LLM). The backend is secured and validated using Pensar, and data is transmitted over a PQ-safe TLS connection with simulated quantum key exchange.

Core Features:

- **Link Input:** Users can paste a link or upload a screenshot (image-to-text via OCR)
- **Feature Extraction:** URL and site metadata are scraped for phishing indicators (e.g., IP address usage, hyphenated domains, lack of HTTPS)
- **ML Classification:** Logistic Regression model predicts malicious, suspicious, or safe link labels and returns a confidence score
- **LLM Threat Report:** A large language model (e.g., DeepSeek) generates an explainable, conversational summary of the risks and contributing features
- **Security Recommendation:** SafeClick provides clear action guidance (e.g., "Proceed with caution", "Avoid this link")
- **Pensar Security Audit:** Ensures backend safety via real-time scanning, input validation, and vulnerability checks
- **PQ-Safe Communication:** Uses TLS hardened with post-quantum-safe techniques; simulates BB84 quantum key exchange between client and server
- **Optional Site Suggestion:** For known scam-style domains (e.g., "pay-pal-login.net"), SafeClick can suggest a legitimate alternative like "paypal.com"

Innovation & Uniqueness:

- Hybrid AI agent combining ML classification and LLM reasoning
- Integrates real-world OCR-to-URL scanning from screenshots
- Uses post-quantum-safe security protocols and quantum cryptography simulation (BB84)
- Multi-layered explainability: feature flags + natural language reasoning

- Strong modularity for scaling to browser extensions, APIs, and educational tools

---

Technical Stack:

- **Frontend:** React or HTML/CSS/JavaScript
- **Backend:** PostgreSQL
- **OCR:** pytesseract, Pillow
- **Scraping & Feature Extraction:** requests, BeautifulSoup, tldextract, ssl, whois
- **ML:** scikit-learn (Logistic Regression), pandas, numpy
- **LLM Integration:** DeepSeek (or compatible open-source model)
- **Security:** Pensar for auditing, TLS with post-quantum enhancements, BB84 key exchange simulation

---

Deliverable: A web-based MVP (minimum viable product) that:

- Accepts direct links and images
- Extracts and analyzes phishing features
- Classifies links using a trained ML model
- Generates a clear threat report using an LLM
- Provides actionable recommendations to users
- Ensures secure, validated backend and secure link handling

Market Potential:

- 45.6% of global email traffic is spam (Statista, 2023), much of it containing harmful links
- SafeClick can evolve into a Chrome/Firefox extension, developer API, or mobile app
- Target audiences:
    - Students, Gen Z, everyday consumers
    - Small businesses without dedicated cybersecurity resources
    - Messaging/email platforms (via plugin or API integration)
    - Cybersecurity education providers

Future Scope:

- Full chatbot-style agent interactions with deeper LLM integration
- Batch URL scanning (e.g., check multiple links in a document)
- Phishing email header analysis
- Crowd-sourced threat flagging and URL reputation scoring
- Enterprise dashboard for team-level link monitoring

SafeClick is built to provide everyday users with AI-driven, explainable, and secure protection in an increasingly dangerous online environment.

1. Introduction

In today's digital landscape, cyberattacks—especially phishing—are more deceptive and accessible than ever. Whether through scam emails, suspicious links in texts, or fake login portals, individuals are constantly at risk of falling victim to malicious URLs. Despite this, most people don't have the tools or technical knowledge to verify a link's legitimacy. SafeClick was created to fill this gap, offering a powerful, easy-to-use solution that brings together cutting-edge AI and real cybersecurity practices into a single, friendly web app.

---

2. Problem & Market Comparison

Phishing is the most common attack vector in modern cybersecurity, with over 90% of breaches starting from deceptive links. According to Statista, 45.6% of all emails in 2023 were classified as spam—many of them containing harmful or fraudulent links. Existing tools in the market tend to focus on either blacklisting URLs or providing vague safety ratings without context or explanation. Others are enterprise-level and too complex for everyday users.

What's missing is a tool that combines accessibility, intelligence, and explainability—something that can act like a personal cybersecurity analyst for both individuals and organizations. SafeClick aims to be that tool.

---

3. Solution

SafeClick is a web-based application that evaluates the safety of links through a combination of machine learning, phishing feature extraction, and LLM-powered threat reporting. Users can paste a suspicious

link or upload a screenshot of a scam message, and SafeClick will extract the URL (using OCR if needed), analyze its structure and associated site features, and return a clear verdict: safe, suspicious, or malicious. But it doesn't stop there—SafeClick generates a human-readable, AI-written explanation of the risk factors involved and gives a direct recommendation on whether to proceed or avoid the site.

This creates a complete feedback loop: detection, reasoning, and recommendation—all accessible through a clean user interface.

Deliverable + Core Features

- **Paste Link or Upload Screenshot** – URL extracted using regex/OCR
- **ML-Powered Classification** – Logistic Regression predicts link safety
- **Phishing Feature Analysis** – Checks for IP address, no HTTPS, hyphenated domains, etc.
- **LLM-Generated Threat Report** – DeepSeek provides plain English explanation
- **Security Recommendation** – Actionable advice based on risk
- **Quantum-Safe Communication** – TLS with simulated BB84 key exchange
- **Pensar Integration** – Security scanning of our own backend

---

4. Market Potential

a. Target Demographic

Our primary audience includes Gen Z and millennial internet users who regularly receive suspicious links through social media, email, and SMS. Secondary audiences include small business teams, school IT staff, and educators looking for tools to teach online safety. Developers and cybersecurity orgs may also integrate our platform via API.

b. Market Research

The market is hungry for intelligent cybersecurity tools that don't require technical expertise. With the explosion of LLMs and digital scams, there's a growing need for tools that combine both. Freemium browser extensions like Grammarly or Honey have proven that user trust can lead to massive adoption, and we believe SafeClick can follow the same path in the security space. A freemium tool that offers immediate protection with optional upgrades is well-suited to meet both individual and organizational needs.

c. Can This Be a Real Startup?

Yes. SafeClick has the potential to become a standalone product or integrate into existing platforms (like antivirus suites or browser extensions). Our pricing model includes:

- **Free Tier:** Link + OCR scanning with basic report
- **Premium ($10/month):** Adds email/text scanning, detailed scores, mobile app
- **B2B Tier:** Starting at $500/month for 100 users, up to $1,500+ for larger orgs

The infrastructure is modular, meaning SafeClick can scale from a single-user app to a full security dashboard used by schools, startups, or agencies.