AI-Powered Email Phishing Detector

Elton Batista
Jordan Chelsey
Curtis Jones

Goal And Motivation

Phishing scams are defined as, "the fraudulent practice of sending emails or other messages purporting to be from reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers"

These emails typically target technologically illiterate people, such as children or the elderly. However, anybody can fall for these scams, posing a personal security risk to anybody who owns an email account.

The goal of this system is to analyze the user's inbox and use an AI model to discern which emails are "phishing" scams. Our goal is to help minimize risk from an individual's inbox by developing software to detect phishing emails.

Approach:
Email Analysis and Classification Using AI

By checking the subject, the sender, the body text, and any links attached to an email, a trained AI model will be able to classify emails based on risk analysis.

Using a dataset of known phishing emails, scams that are not normally flagged by a spam filter will be able to be discerned from legitimate emails using a combination of factors determined to be suspicious by the model.

Emails in an inbox are given a "risk score" based on the model's confidence that a given email is phishing.

Approach:
Gmail Integration Tools

By logging in through Google, an inbox from any domain can be checked for phishing scams.

The AI will be able to sift through and check for patterns in an inbox on top of the analysis of known phishing emails.

A variety of functions can be performed on high-risk emails, such as deletion and reporting.

Approach: Visualization

Emails with a higher risk are sorted to the top of the page. Each flagged email will be given an explanation as to why the model believes it is a scam. This gives users a better understanding of why an email should be removed from their inbox.

Several visual tools, such as charts and diagrams, will allow users to see the percentage of emails in their inbox that are flagged as "phishing emails", "suspicious", and "legitimate". By doing this, users gain a better understanding of how at risk they are to phishing scams on the internet.

Compared to Existing Tools…

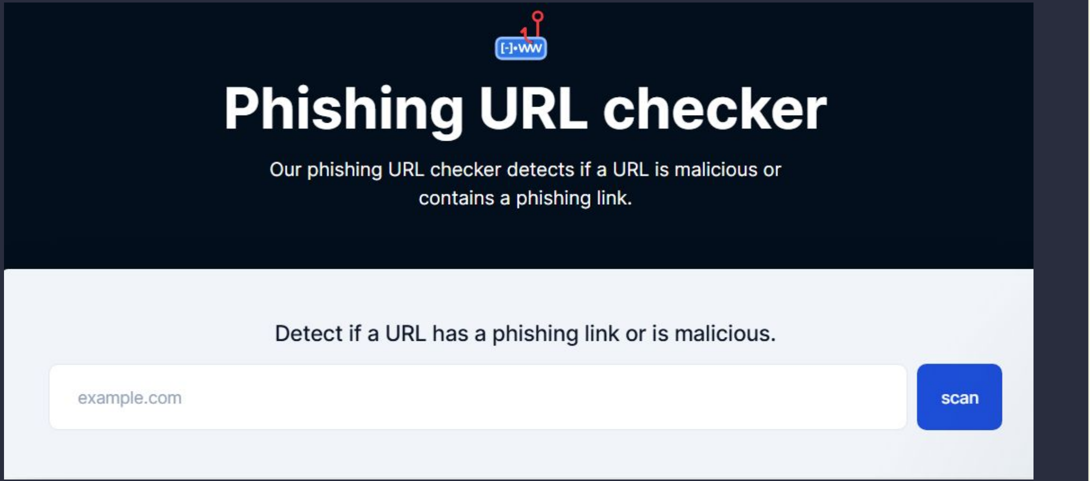
Most phishing detectors require pasting individual emails into text boxes on websites.

Existing tools lack any direct integration with the users inbox. This makes their systems tedious to use, especially when a user would like to check anywhere from hundreds to thousands of emails.

By introducing an AI baseline into this tool, we allow users to get dynamic feedback on each email as well.

In combination, these systems allow for an easy and intuitive way to filter your inbox compared to other systems.

For Example…



Phishing URL checker

Our phishing URL checker detects if a URL is malicious or contains a phishing link.

Detect if a URL has a phishing link or is malicious.

This existing tool only checks for phishing links, assumingly just checking for keywords or pulling from a database of known links…

Algorithms and Tools / Technical Challenges

AI Models: LLM vs Machine Learning

LLMs (Large Language Models)

Pre-trained to understand existing language patterns

Can be highly susceptible to faults in output

Can better understand conceptual and abstract sentences

Machine Learning (Using a transformer Model)

Neural network

Good with natural language processing

Could be more accurate on a large data set than an LLM

Algorithms and Tools / Technical Challenges

Datasets

Training our AI model on a quality dataset is extremely important if we want to have accurate measurements on real emails.

Requirements for good data sets:

Contains the following components of an email:
Email of sender
Subject
Body

Includes examples of "phishy" links (I.e. shortened URLs)
Large data set

By using a larger data set, we reduce the margin of error in the trained model

Milestone 1

Compare and select quality datasets

Compare and select an AI model that fits the needs of the project through rigorous testing

Create a small demo using the selected AI model

Milestone 2

Use selected dataset to feed and train selected AI model

Gather testing data to forward test the detection model. Testing data can be:

- Real phishing scams from the team's inboxes
- Generated phishing scams using AI
- A reserved portion of the selected dataset

Use results of testing to tune detection model

Research Gmail Integration

Milestone 3

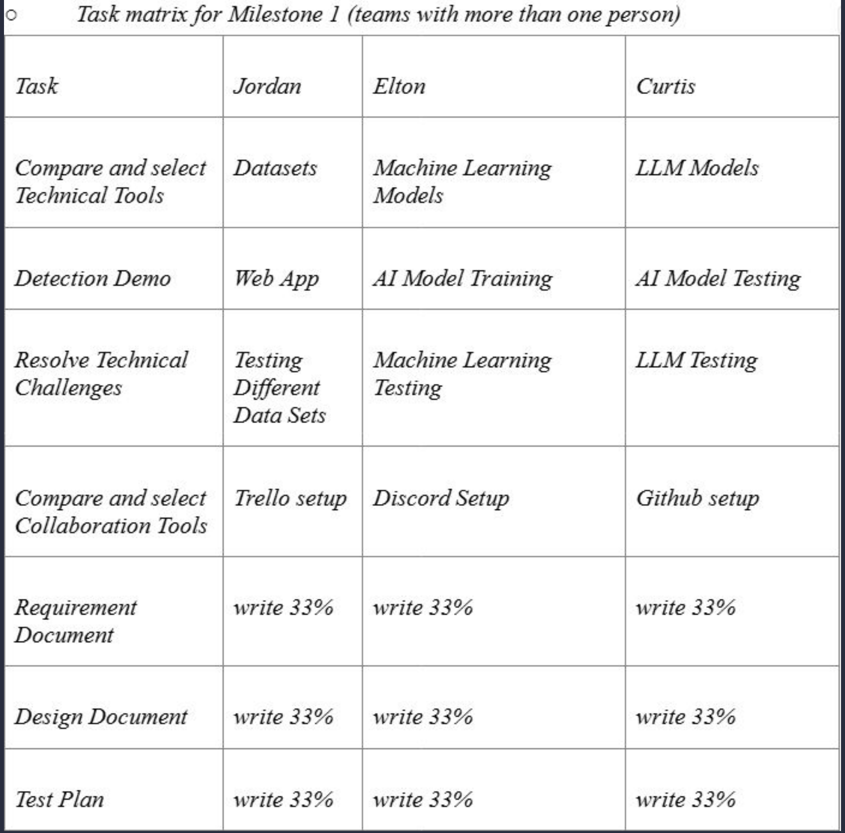
Complete Gmail Integration

- Feed actual user inbox messages directly into the detection model
- Allow the model to delete suspicious messages when given permission to do so

Begin Web Interface Construction

- Implement Google OAuth
- List the user's messages on the dashboard

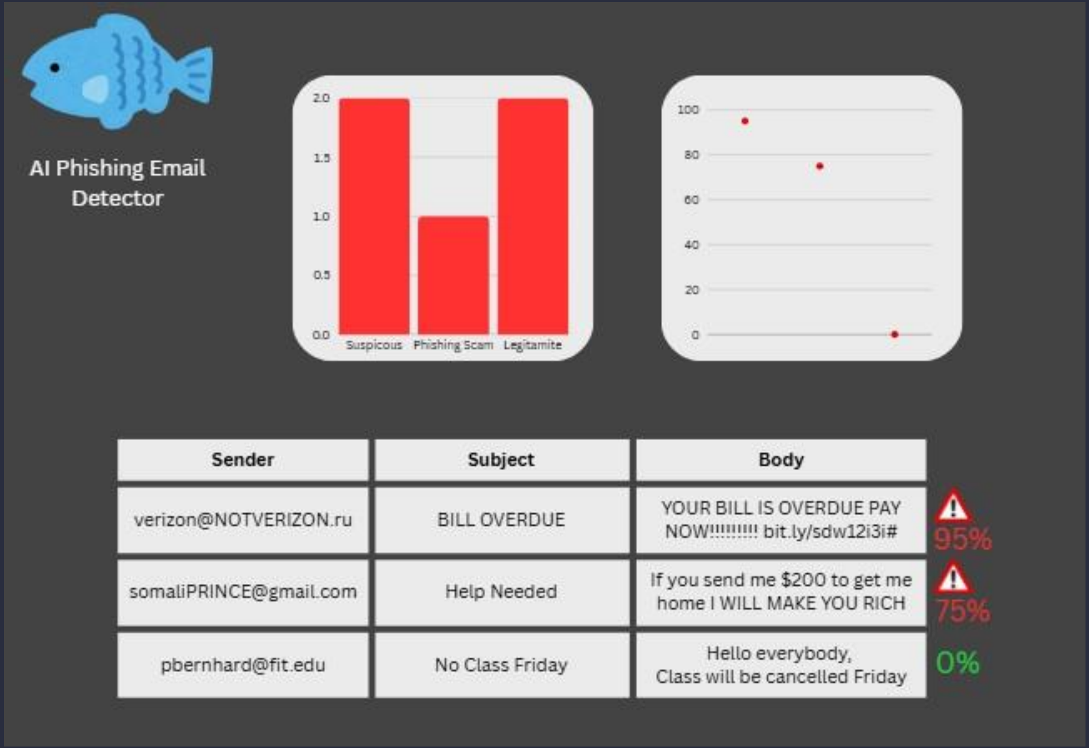
Task Matrix



- *Task matrix for Milestone 1 (teams with more than one person)*

Task	*Jordan*	*Elton*	*Curtis*
Compare and select Technical Tools	*Datasets*	*Machine Learning Models*	*LLM Models*
Detection Demo	*Web App*	*AI Model Training*	*AI Model Testing*
Resolve Technical Challenges	*Testing Different Data Sets*	*Machine Learning Testing*	*LLM Testing*
Compare and select Collaboration Tools	*Trello setup*	*Discord Setup*	*Github setup*
Requirement Document	*write 33%*	*write 33%*	*write 33%*
Design Document	*write 33%*	*write 33%*	*write 33%*
Test Plan	*write 33%*	*write 33%*	*write 33%*

Example UI



AI Phishing Email Detector

Sender	Subject	Body	
verizon@NOTVERIZON.ru	BILL OVERDUE	YOUR BILL IS OVERDUE PAY NOW!!!!!!!!!! bit.ly/sdw12i3i#	⚠ 95%
somaliPRINCE@gmail.com	Help Needed	If you send me $200 to get me home I WILL MAKE YOU RICH	⚠ 75%
pbernhard@fit.edu	No Class Friday	Hello everybody, Class will be cancelled Friday	0%