AI in Social Engineering and Phishing Campaigns: Spam Detector

Introduction

- Phishing and social engineering are major cyber threats.
- Traditional spam filters are rule-based and easily bypassed.
- AI techniques (ML and DL) offer adaptive and intelligent detection.
- Aim: Build an AI-based Spam Detector using modern techniques.

Problem Statement

- Cybercriminals use deceptive messages to trick users.
- Traditional spam filters fail against modern phishing techniques.
- Static systems can't adapt to new attack patterns.
- Need for intelligent systems with better context understanding.

Solution Overview

- Use of AI/ML/DL models to detect phishing in emails.
- Train models on real datasets with labeled spam/phishing data.
- Deploy system with real-time detection and high accuracy.
- Models: Naive Bayes, SVM, Random Forest, LSTM.

Tool Architecture

- Modular components: UI, Preprocessing, Feature Extraction, Classifier.
- Trained models integrated for real-time classification.
- Web interface for user input and output display.
- Confidence score and suggested actions provided.

Code/Tool Breakdown

- Language: Python; Tools: Scikit-learn, TensorFlow, NLTK.
- Preprocessing: Cleaning, Tokenization, Vectorization.
- Feature extraction from email header, body, URLs.
- Serialized models for efficient deployment (.pkl/.h5).

Sample Output

- Input: Suspicious email content.
- Model: LSTM.
- Output: ⚠️ Phishing Email Detected.
- Confidence: 97.3%.

Real-World Use Cases

- Organizations detecting phishing at scale.
- Banks and e-commerce securing customer interactions.
- Educational institutions preventing cyber fraud.
- Integration into email clients and webmail systems.

Model Evaluation Results

- Naive Bayes: 91.2% Accuracy.
- SVM: 93.5% Accuracy.
- Random Forest: 95.0% Accuracy.
- LSTM: 97.3% Accuracy (best performance).

Market Relevance

- Growing phishing threats (36% of breaches).
- Rising demand for AI-powered email security.
- Adoption across sectors: Finance, Health, Education.
- Commercial potential as SaaS or email plugin.

Ethical Considerations

- Ensure data privacy and fairness in training.
- Avoid biases and explain AI decisions.
- Prevent misuse of AI for crafting attacks.
- Maintain user control over filtering decisions.

Future Enhancements

- Real-time browser/email client plugins.
- Adaptive learning with user feedback integration.
- Support for multiple languages and message types.
- Voice phishing (vishing) and SMS (smishing) detection.

THANKYOU