AI-Driven Phishing Attacks Detection in Indian Enterprises

Phishing remains one of the most prevalent cyber threats affecting enterprises globally, with a significant rise in sophisticated, AI-generated phishing campaigns. The emergence of Large Language Models (LLMs) has enabled attackers to craft highly contextual, grammatically flawless, and personalized phishing emails that bypass traditional signature-based and rule-based email security systems. Indian enterprises face an additional challenge due to multilingual communication environments, particularly the widespread use of Tamil-English code-mixed content in corporate email exchanges. This increases vulnerability to Business Email Compromise (BEC) attacks and targeted spear-phishing campaigns that conventional filters fail to detect.

This research proposes the design and development of an AI-driven phishing detection framework specifically tailored for Indian enterprises. The objective of the system is to move beyond keyword-based filtering and instead detect malicious intent, behavioral anomalies, and linguistic patterns within multilingual email communications.

The proposed framework consists of three core components:

1. **Behavioral Email Pattern Analysis Module** – This module analyzes historical communication patterns between users to identify anomalies such as unusual sender behavior, abnormal request urgency, deviations in writing style, or unexpected financial instructions.

2. **Multilingual NLP Engine** – Leveraging Natural Language Processing techniques, this module processes English, Tamil, and Tamil-English code-mixed emails. Techniques such as tokenization, sentiment analysis, transformer-based language models (e.g., fine-tuned BERT variants), and contextual embeddings will be used to detect semantic inconsistencies and phishing indicators in regional language communication.

3. **Intent Detection Model** – A deep learning classifier will be developed to identify malicious intent (e.g., credential harvesting, invoice fraud, urgency-based manipulation) rather than relying solely on static keywords. Transformer-based architectures and attention mechanisms will be used to capture contextual meaning.

Technologies Used

- **Programming Languages:** Python
- **Machine Learning & NLP Libraries:** TensorFlow / PyTorch, Scikit-learn, Hugging Face Transformers
- **Language Models:** Fine-tuned multilingual BERT or IndicBERT models
- **Data Processing:** Pandas, NumPy
- **Email Parsing:** IMAP/SMTP protocol integration
- **Deployment:** Web-based dashboard using Flask or FastAPI
- **Database:** PostgreSQL / MongoDB for email metadata storage

Development Methodology

The system will be developed in phases:

1. **Dataset Collection & Labeling:** Collect multilingual phishing and legitimate email datasets relevant to Indian enterprises.
2. **Preprocessing:** Clean, tokenize, and normalize multilingual and code-mixed text.
3. **Model Training:** Train supervised classification models to distinguish phishing from legitimate emails.
4. **Behavioral Profiling:** Establish baseline communication behavior for users and detect deviations.
5. **Integration & Testing:** Integrate modules into a unified detection framework and evaluate performance using metrics such as accuracy, precision, recall, F1-score, and false positive rate.

6. **Deployment & Validation:** Deploy as a prototype email security gateway or plugin and validate in a simulated enterprise environment.

   The expected outcome of this research is a scalable and adaptive phishing detection framework capable of identifying AI-generated, multilingual phishing emails with higher detection accuracy compared to traditional filters. By focusing on intent detection and regional language support, this study aims to strengthen enterprise email security and provide a practical, implementation-ready solution tailored to the Indian cybersecurity landscape.