RAMAIAH
Institute of Technology

GenAI Laboratory

Project Title Submission Report

Project Title: PhishNet" - An AI-Powered Social Engineering Simulation Game

Team Details

NAME	USN
Bhaskar Datta	1MS22CY014
Dhanush J	1MS22CY021
Dushyanth S	1MS22CY025

Abstract

The increasing sophistication of social engineering attacks, particularly phishing, poses a significant threat to digitally-native individuals like students. Traditional cybersecurity education often fails to provide practical, engaging training, leaving students vulnerable to real-world threats. This gap highlights the need for an interactive learning tool that can simulate realistic attack scenarios in a safe environment.

To address this challenge, we propose "PhishNet," a Retrieval-Augmented Generation (RAG) based simulation game. This project leverages the dynamic dialogue capabilities of Large Language Models (LLMs) to create an immersive experience where players must identify and navigate complex social engineering scenarios presented by AI-driven characters. The system features a "Digital Guardian" assistant, powered by a RAG mechanism, allowing players to query for real-time, context-aware advice grounded in a curated cybersecurity knowledge base. By combining interactive gameplay with factually accurate educational support, PhishNet aims to move beyond passive learning and actively develop critical threat-recognition skills. This solution not only makes cybersecurity education engaging and memorable but also provides a scalable platform for practical, hands-on digital safety training.

Problem Definition

Students and young adults are prime targets for cybercriminals due to their extensive online presence across social media, email, and messaging platforms. They constantly face a barrage of sophisticated social engineering attacks, from deceptive phishing emails offering fake internships to smishing (SMS phishing) attacks with malicious links. However, conventional methods of cybersecurity training, such as presentations or static multiple-choice quizzes, are often perceived as unengaging and fail to prepare individuals for the psychological manipulation used in real attacks. These methods lack the dynamism and contextual nuance of modern threats, creating a dangerous gap between theoretical knowledge and practical application. Therefore, there is a critical need for an intelligent system that can simulate realistic attack vectors in a dynamic, interactive format, allowing users to safely learn by doing and build resilient security habits.

Motivation

The motivation for this project stems from the desire to transform cybersecurity education from a passive requirement into an active, engaging experience. Students learn best when they are immersed and challenged. Existing tools often fall short, delivering dry content that is quickly forgotten. With the recent advancements in Generative AI, it is now possible to create highly realistic and unpredictable simulations that mimic the evolving tactics of cyber-attackers. We are motivated to leverage these technologies—specifically LLMs, RAG, and LangChain—to build a "digital playground" where users can make mistakes without real-world consequences. The goal is to develop a smart, interactive game that not only teaches the "what" and "how" of social engineering attacks but also helps users develop the critical thinking skills needed to protect their digital lives, thereby fostering a stronger, more intuitive security mindset.

GenAI Concepts

Large Language Models (LLMs): Generate dynamic, persuasive, and context-aware dialogue for the in-game non-player characters (NPCs) who simulate the social engineering attackers. This ensures each playthrough is unique and challenging.

Retrieval-Augmented Generation (RAG): Powers the in-game "Digital Guardian" assistant. It combines the retrieval of factual information from a curated cybersecurity knowledge base with an LLM's generative capabilities to provide players with accurate, context-specific explanations and advice about potential threats they encounter.

LangChain: Acts as the central orchestrator for the game's logic. It manages the conversational flow between the player and multiple AI agents (attackers and the

assistant), chains together different prompts and tool uses, and maintains the state of the game to create a coherent and interactive narrative experience.

Document Embeddings: Represent the text from the cybersecurity knowledge base (articles, blogs, threat definitions) as numerical vectors. This enables fast and effective semantic search, allowing the RAG system to find the most relevant information related to a player's query.

Vector Databases (e.g., ChromaDB): Store and retrieve the document embeddings efficiently. This serves as the long-term memory for the RAG system, enabling near-instantaneous lookup of relevant security information during gameplay.

Prompt Engineering: Craft sophisticated prompts that guide the LLMs to act as specific personas (e.g., a convincing phisher, a helpful but neutral assistant) and integrate the retrieved context seamlessly into their responses to create a believable and educational simulation.