5ta Jornada de Ciberseguridad de la Universidad Nacional de Colombia

CodeRonin S-AI: Developing Social Engineering Attacks using Artificial Intelligence

Abstract

This project proposes **Code Ronin S-IA**, a solution that uses Artificial Intelligence (AI) to carry out social engineering attacks by combining voice cloning, phishing, and Open Source Intelligence (OSINT) techniques. Additionally, **Code Ronin S-IA** integrates various tools such as *Maltego* for gathering public information about a victim, *GoPhish* for email delivery, and *MaxPhisher* for capturing credentials and biometric data.

Introduction

Code Ronin S-IA is an AI-powered tool that generates personalized social engineering attacks, allowing organizations to assess their response capabilities to such attacks and identify weaknesses in the cybersecurity training of their employees. Additionally, **Code Ronin S-IA** will serve to train both practitioners and cybersecurity students, exposing them to social engineering tactics and enhancing their ability to detect and mitigate threats in real-world environments.

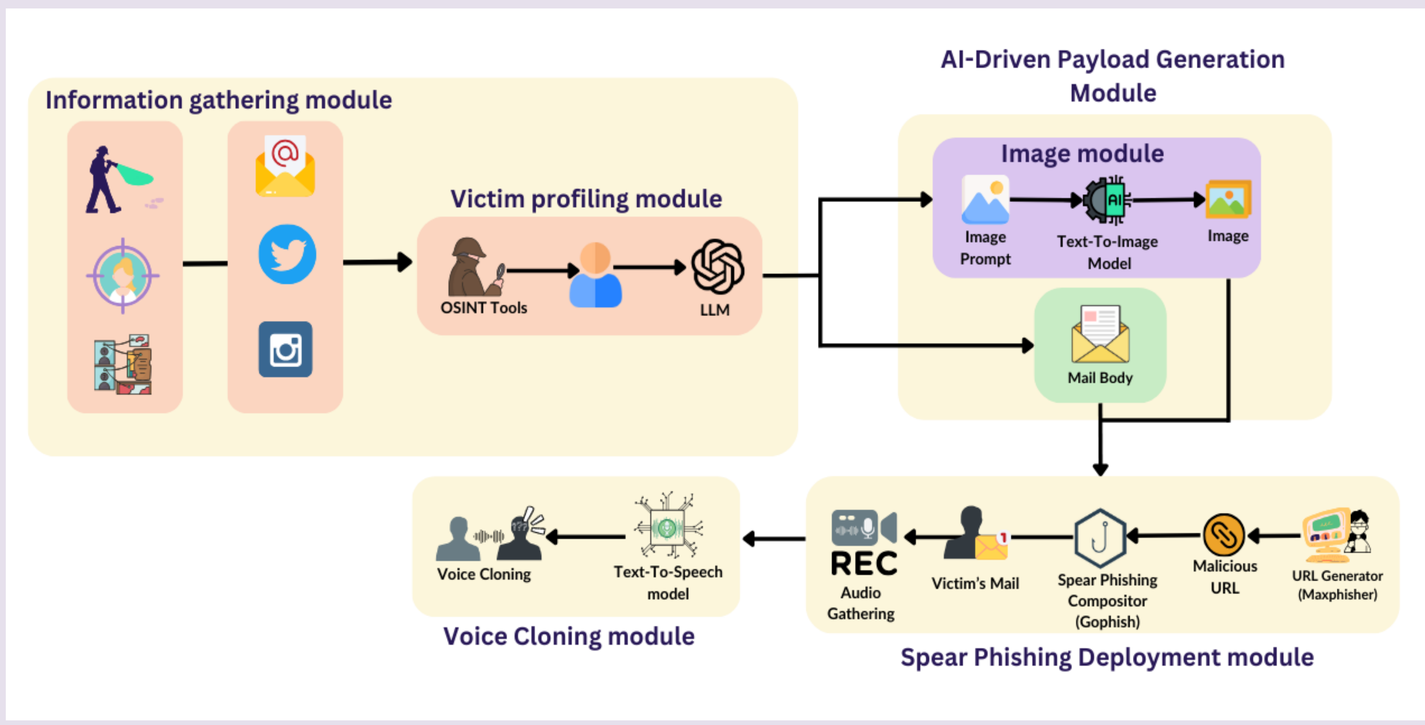
Background

Recent studies [1] show that employees remain a weak link in security, and attack simulators can help mitigate risks associated with this vulnerability by identifying susceptible human actors. On the other hand, AI has advanced to the point where attackers can leverage more sophisticated social engineering techniques, such as deepfakes[2], phishing, and jailbreak-as-a-service[3]. These AI-enabled tactics are increasingly accessible and harder to detect, posing a greater challenge for cybersecurity teams.

Proposal

OSINT Module: Code Ronin S-IA integrates an OSINT module that utilizes tools such as *Maltego* and *Sherlock* to gather detailed information about a target, along with conducting a manual information search using OSINT techniques.

Mail Distribution and MaxPhisher: Code Ronin S-IA includes a module dedicated to email distribution, implemented using the *GoPhish* tool. It incorporates AI-generated images designed to be highly appealing to the victim, as well as a link to a cloned webpage created with *MaxPhisher*.



Experiments

LinkedIn, Facebook, TikTok, Steam, and TryHackMe profiles, as well as other accounts registered on various CTF platforms, were successfully obtained. Additionally, the victim's personal webpage, healthcare provider (EPS), and voting location were identified. By leading victims to believe they were accessing a Zoom meeting, detailed information was captured, including audio recordings.

Pruebas de funcionamiento del modelo de reconocimiento de hablantes	
Escenarios de prueba	**Puntaje de similitud [−1, 1]**
Misma voz, textos diferentes	0.895
Voces diferentes, mismos textos	0.03
Voz original vs Voz clonada, textos diferentes	0.74
Misma voz, mismo texto	1.00

Resultados del modelo en voces clonadas con modulo de voz	
Escenarios de prueba	**Puntaje de similitud [−1, 1]**
Voz clonada vs Voz original 8 (hombre)	0.63
Voz clonada vs Voz original 9 (mujer)	0.69
Voz clonada vs Voz original 10 (mujer)	0.76

Conclusions

Code Ronin S-IA demonstrates remarkable versatility, particularly excelling in generating audio that replicates the victim's voice with high precision. In conducted tests, the similarity reached a level of 0.76 on a scale from -1 to 1, reflecting a highly accurate cloning. However, achieving better results requires extensive and high-quality recordings of the original voice. Additionally, the quality of prompts for generating images and texts directly depends on the detailed information collected about the victim, highlighting the need to integrate more OSINT tools to enhance target profiling.

References

1. Kaspersky: Investigación de kaspersky confirma que tales violaciones son casi tan comunes como las brechas de ciberseguridad (2023)
2. BBVA: La IA, en los dos lados de la ciberseguridad: aliada y amenaza en el mundo digital (2024)
3. Desplanques, B., Thienpondt, J., Demuynck, K.: ECAPA-TDNN: emphasized channel attention, propagation and aggregation in TDNN based speaker verification (2020)

Autor(es)

Nicolas Rogers, Carolina Rojano, Laura Salazar, Diryon Mora, Fabio Rizo, Juan Casanova, Daniel Díaz-López
Matemáticas Aplicadas y Ciencias de la Computación (MACC)
Escuela de Ingeniería, Ciencia y Tecnología (EICT)
Universidad del Rosario

Organiza

Universidad Nacional de Colombia
Grupo de investigación UNsecureLab
Semillero de investigación Uqbar
Grupo de investigacion Tlön

UNIVERSIDAD NACIONAL DE COLOMBIA