Automated Social Engineering Reconnaissance with LLM Agents

[author names removed]

Rochester Institute of Technology

Golisano College of Computing and Information Sciences

Department of Cybersecurity

CSEC.490 Capstone in Cybersecurity

Project 12: Automated Social Engineering Reconnaissance with LLM Agents

[faculty sponsor names removed]

Friday, December 13, 2024

Table of Contents

Table of Figures

1 Introduction

In the digital age, the vast reservoir of information on the internet presents significant opportunities and risks to personal privacy and security. The widespread availability of public data has enabled the development of advanced tools and methodologies that streamline the aggregation, analysis, and exploitation of personal information on a large scale. In particular, open-source intelligence (OSINT) tools extract sensitive details from diverse sources such as social media platforms, government databases, employment directories, and online data brokers. Malicious actors exploit this aggregated information to craft sophisticated social engineering attacks, relying on psychological manipulation rather than technical vulnerabilities. Spear phishing (i.e. targeted phishing attacks) has emerged as a particularly effective method for compromising organizations. Due to its dual potential for legitimate and malicious use, OSINT has become a critical resource across various domains.

The efficacy of social engineering stems from inherent human vulnerabilities. Scholars emphasize that social engineering often surpasses most traditional hacking techniques in breaching even the most secure systems, as users represent the weakest link in the security chain (Krombholz et al., 2014). Despite significant advancements in security technologies, human susceptibility remains a consistent liability. Consequently, attackers increasingly focus on exploiting psychological weaknesses to breach otherwise secure systems and environments. This reality underscores the urgent need for continued research into the mechanisms and defenses against such attacks.

Parallel to the rising threat of social engineering, advancements in artificial intelligence (AI) and large language models (LLMs) have accelerated. These technologies have revolutionized task automation, enhanced productivity, and elevated proficiency across industries, often improving output quality (Cambon et al., 2023). This project investigates the intersection of OSINT-based social engineering and AI, focusing on how LLMs can act as a catalyst to automate and enhance information collection and create high-quality, effective phishing materials.

1.1 Motivation

Traditional OSINT processes are labor-intensive and time-consuming, requiring the manual identification and analysis of numerous data sources. This limitation restricts the scalability of spear phishing campaigns, typically reserved for high-value targets such as C-suite executives. These individuals, however, often receive extensive training on recognizing phishing attempts and exercise caution when interacting with email communications (Ball et al., 2012, p. 275).

The integration of AI into OSINT workflows transforms this dynamic. AI-assisted OSINT enables spear phishing email generation automation, significantly broadening the pool of potential victims. Unlike high-value targets, these newly accessible individuals often lack the training to identify phishing attempts, rendering them more susceptible to manipulation. Furthermore, traditional phishing indicators, such as grammatical errors or poorly constructed language, have become less reliable (Blythe et al., 2011). With the advent of LLMs, attackers can generate phishing emails that are grammatically accurate and linguistically sophisticated, and they can do it faster than ever before.

The ability of AI to influence the efficiency and effectiveness of social engineering attacks is a critical area of concern. Therefore, examining the mechanics and implications of these developments is imperative. This project investigates how LLMs can amplify the offensive and defensive capabilities of AI-driven social engineering attacks. Specifically, it will explore how LLMs can process data harvested by web scrapers to generate actionable intelligence for attackers.

1.2 LLMs in Social Engineering

This project explores how LLMs and generative AI can be leveraged to facilitate the collection of OSINT, conduct social engineering, and create phishing materials such as emails. By incorporating AI into various aspects of the process, the project aims to illustrate how AI and LLMs can streamline the collection, analysis, and creation of actionable results. LLMs will be applied in this tool's key functionalities via:

- Data Aggregation and Analysis: Organizing the collected intelligence into a structured format and analyzing it to derive actionable insights.
- Phishing Material Generation: Producing customized phishing emails and other materials aligned with the selected attack vector.

This project aims to raise awareness of the associated risks and provide insights into potential mitigation strategies for organizations and individuals by integrating AI and LLMs into social engineering campaigns.

2 Background

2.1 Traditional Social Engineering Methods

Traditional social engineering requires attackers to invest significant manual effort in gathering sensitive information from diverse sources such as social media platforms, government databases, employment directories, and online data brokers (Hadnagy, 2018). Attackers then analyzed this information to identify exploitable details and tailor their manipulative tactics. The labor-intensive nature of this process restricted the scalability of such attacks.

2.2 The Role of AI in Enhancing Social Engineering

Artificial intelligence and LLMs have transformed this landscape. AI enables the automation of social engineering attacks, reducing the time required for execution while increasing the frequency and effectiveness of such operations. Language models like OpenAI's GPT-4o, Anthropic Claude 3.5, and Google DeepMind's Gemini 1.5 are particularly adept at language recognition and generation, making them powerful tools for creating convincing social engineering content.

Research has begun to elucidate how AI affects the speed and frequency of social engineering. Some discuss how LLMs facilitate the rapid scaling of social engineering attacks, presenting a risk quantification framework and emphasizing the need for proactive, adaptive defense strategies (Yu et al., 2024). Others examine the transformative role of generative AI in these attacks, proposing a framework for understanding AI-driven social engineering threats and outlining countermeasures (Schmitt & Flechais, 2023). They identify three key ways generative AI aids social engineering:

1. **Believable Content Creation:** LLMs can produce highly credible phishing messages, often mimicking a target's writing style or an organization’s communication patterns.
2. **Personalized Targeting:** AI sifts through vast data sets to identify personal details, enhancing the relevance and persuasiveness of attacks.
3. **Automated Execution:** AI automates message creation and scheduling, allowing attackers to scale their efforts and launch simultaneous attacks.

AI is also increasingly employed in social engineering to generate sophisticated voice and audio deepfakes. For instance, in 2019, a UK energy company CEO was deceived into transferring $243,000 to a Hungarian bank account in part because of deepfaked audio (Stupp, 2019). More recently, in 2024, a Chinese finance employee was manipulated into disbursing $25 million following a video call with a deepfake impersonation of the company’s chief financial officer (Chen & Magramo, 2024).

Most current research focuses on AI’s theoretical risks rather than its practical implementations. This project aims to bridge this gap by integrating AI and LLMs into a workflow for executing social engineering attacks.

2.3 Existing AI Tools for Social Engineering

Integrating AI into social engineering has revolutionized traditional methods, automating previously complex tasks requiring manual labor. Below is an exploration of several AI tools currently employed in social engineering, each contributing to different stages of attack development, from reconnaissance to content generation.

2.3.1 Primary Tools for Comparison

- **Maltego with GPT Plugin:** Maltego, an OSINT and data visualization tool, is enhanced with a GPT-based plugin to gather information and generate targeted phishing messages.

Its data visualization capabilities facilitate the design of sophisticated social engineering campaigns (Maltego, n.d.).

- **Recon-ng:** An OSINT framework that automates the collection of public information, creating detailed profiles of targets. LLMs can then use this intelligence to generate personalized attack content (lanmaster53, n.d.).
- **GoPhish**: An open-source toolkit for automating phishing campaigns. While not inherently AI-driven, GoPhish can be integrated with LLMs to craft personalized phishing emails, enhancing their efficacy (gophish, n.d.).

2.3.2 Complementary Tools

- **DeepFaceLab:** An AI tool used for creating deepfake videos. Social engineers utilize it to generate videos impersonating individuals, which can be leveraged in coercive or manipulative scenarios (iperov, n.d.).
- **Deep Voice Tools (Lyrebird AI)**: Lyrebird AI generates realistic audio deepfakes by cloning voices. Attackers use these audio samples to impersonate individuals, enabling voice phishing (vishing) attacks (Descript, n.d.).

2.3.3 Contextual Tools for Broader Insight

- **Power Pwn**: A tool that combines OSINT with AI-generated content to automate social engineering campaigns, collecting public data and generating targeted attack vectors (mbrg, n.d.).
- **Phishing Pipeline with AIaaS (GovTech Singapore)**: This system leverages AI as a Service (AIaaS) to create targeted phishing emails, combining personality analysis tools (e.g., Humantic AI) with GPT-3 for enhanced phishing capabilities (Lim et al., 2021).

2.4 Synthesis of Existing Tools

The tools discussed above share commonalities in their role in automating the social engineering process. Tools like DeepFaceLab, Lyrebird AI, and GPT-based plugins enhance the credibility of generated content, whether video, audio, or text, producing outputs previously regarded as authoritative. Meanwhile, Maltego, Recon-ng, and Power Pwn focus on automating the reconnaissance and data collection phases, significantly improving the efficiency of these processes and providing users with well-organized, actionable intelligence. In contrast, tools such as GoPhish and GovTech Singapore's AIaaS platform prioritize streamlining the deployment of attacks and the personalizing of attack vectors, enabling highly targeted and effective delivery methods.

2.5 Our Tool vs Primary Tools for Comparison

Feature	Our Tool	Maltego with GPT	Recon-ng	GoPhish
Data Collection	Automated web scraping (LinkedIn, etc.)	Preconfigured OSINT plugins	Automated collection	No data collection features
LLM Utilization	Analysis + phishing material creation	Phishing message generation	No native LLM integration	Requires external tools for LLM
Output	Dashboard/reports + actionable insights	Graph-based visualization	Raw data/profiles	Campaign management tools
Scope	End-to-end OSINT + social engineering	OSINT and relationship mapping	OSINT collection	Phishing campaign execution

2.6 Attack Framework for AI-Driven Social Engineering

AI integration into social engineering can be analyzed through an enhanced attack framework based on the model proposed by Mouton et al. (2014). This framework builds upon the Mitnick and Simon (2002) attack cycle by incorporating detailed phases and utilizing an ontological model to address existing weaknesses, enabling the generation of standardized attack scenarios for training, awareness, and countermeasure development. The phases are as follows:

1. **Attack Formulation**: Define the attack's goal and identify the optimal target to achieve this objective.
2. **Information Gathering**: Identify potential sources of information and collect relevant details about the target. Assess whether sufficient information has been gathered to proceed.
3. **Preparation**: Analyze the collected information to understand the target comprehensively. Develop the attack vector, including creating a convincing pretext or scenario to manipulate the target effectively.
4. **Develop a Relationship**: Use the gathered information and planned techniques to initiate communication with the target, building rapport and trust.
5. **Exploit the Relationship**:
   - **Priming the Target**: Manipulate the target's emotional state to make them more susceptible to exploitation.

- ○ **Elicitation**: Use the established relationship to elicit the desired information or prompt the target to take a specific action.

6. **Debrief**: To prevent suspicion, restore the target's emotional state to normal. If necessary, gather additional information or conclude the attack after meeting determined goals.
7. **Goal Satisfaction**: Finalize the attack upon successfully achieving the intended objective.

Integrating AI into these phases enhances the efficiency and scalability of social engineering attacks. Automated OSINT collection and advanced content generation offer significant improvements over traditional methods, allowing for highly personalized and large-scale attacks.

2.7 Tools in Kevin Mitnick's Attack Cycle

Existing tools address specific stages of the attack cycle. For example, during the reconnaissance phase, Maltego with GPT Plugin collects and maps target data, while Recon-ng compiles publicly available information into detailed profiles. DeepFaceLab generates deepfake videos for impersonation in the weaponization phase, and Deep Voice Tools creates voice replicas for vishing attacks (Lyrebird AI, n.d). Tools like GoPhish and Phishing Pipeline with AIaaS contribute to the delivery phase by automating phishing emails tailored to specific targets. Finally, Power Pwn aids the exploitation phase by using AI to identify weaknesses and recommend attack strategies.

Our tool differs by combining multiple stages of the attack cycle into a single workflow. It automates reconnaissance through a web scraper that gathers information from platforms like LinkedIn, Instagram, and Twitter. It processes and analyzes this data during the weaponization phase using LLMs and generates phishing emails and other materials to simulate the delivery phase. Additionally, it generates reports highlighting vulnerabilities, providing insights to help organizations identify and address risks. This integration allows our tool to tackle offensive attack simulation and defensive risk assessment, making it uniquely versatile compared to existing options.

3 Project Description

3.1 Overview

Contemporary cybersecurity defenses rarely simulate the highly personalized tactics that real attackers employ during social engineering campaigns. Our project addresses this gap by developing a tool that emulates targeted spear-phishing at scale. Specifically, it integrates automated OSINT gathering with a LLM to produce realistic, context-rich phishing emails. The system seeks to mirror adversaries who leverage publicly available information, from social media profiles to professional networking sites, to craft compelling narratives and launch convincing phishing attacks.

3.2 Objectives

The primary objective is to create a reliable, scalable tool that red and blue teams can use to understand and counter sophisticated phishing threats. Rather than generating generic "shotgun" phishing content, the tool aims to:

- **Identify Relevant Profiles:** Given a target's name or known attributes, the tool locates associated public profiles (e.g., LinkedIn, Twitter, Instagram) and aggregates open data such as employment history, endorsements, interests, and affiliations.

- **Extract Actionable Intelligence:** The tool compiles a target's digital footprint into a comprehensive profile using structured scraping and parsing. This profile highlights personal and professional details that attackers could exploit.

- **Produce Tailored Phishing Content:** Using an LLM to analyze the compiled data, the tool generates realistic phishing emails designed to appeal to the target's interests, leverage their professional context, or reference recent posts. These emails mimic adversarial tactics, often involving subtle details to build trust and urgency.

- **Enable Security Training:** By producing believable, data-driven phishing scenarios, the tool serves as a training asset. Organizations can use these simulations to test employee response, refine incident response protocols, and strengthen organizational awareness.

3.3 System Components

The system consists of four integrated steps that work together to produce actionable outcomes:

1. **Profile Identification:** The user provides baseline data about a target, such as a name and general profession or location. The system then queries known public platforms and search engines to locate candidate profiles corresponding to the target.

2. **Data Collection:** Once candidate profiles are identified, a custom web scraper, built with Python and Selenium, collects publicly accessible details. These details include job titles,

posts, comments, listed affiliations, interests, known connections, and geographical information. The scraping code employs targeted selectors to parse profile pages, handles pagination where possible, and stores raw results.

3. **Data Aggregation and Analysis (LLM Integration):** The aggregated data is passed to an LLM (OpenAI's GPT-3.5-turbo) via a structured prompt. The LLM identifies relevant patterns, such as recurring professional themes, personal interests, or recent travel announcements, and determines how these might be exploited. The model then generates an internal representation of potential attack vectors, focusing on what would make a phishing email convincing. For example, if the target frequently discusses new tech tools, the LLM might propose impersonating a well-known vendor offering a product demo.
4. **Phishing Simulation Generation:** Using the identified patterns, the LLM produces a fully formed phishing email. This email might reference the target's professional role, recent social media posts, or known affiliations. The email is designed to create trust, urgency, or fear, mimicking real-world attacker strategies. The tool can produce multiple variants, allowing testers to assess resilience against different phishing narratives.

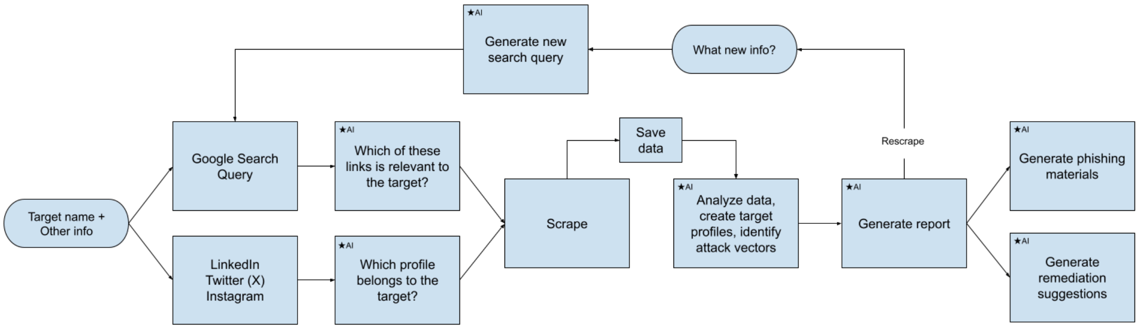

★AI
Generate new search query
What new info?
Google Search Query
★AI
Which of these links is relevant to the target?
Save data
Rescrape
★AI
Generate phishing materials
Target name + Other info
Scrape
★AI
Analyze data, create target profiles, identify attack vectors
★AI
Generate report
LinkedIn
Twitter (X)
Instagram
★AI
Which profile belongs to the target?
★AI
Generate remediation suggestions

Figure 1. Workflow Diagram

3.4 Ethical Considerations and Challenges

Ethical considerations informed every stage of development. While personalized phishing simulations have a legitimate training purpose, they risk infringing on privacy if misused. To address this, the tool incorporates:

- **Restricted Use Cases:** It operates only on authorized targets under the guidance of security professionals.

- **Data Anonymization:** Sensitive details are either replaced with placeholders or summarized in a manner that retains training value without disclosing private information.
- **Disclaimers and Permissions:** The tool's output includes statements clarifying that the content is for training and simulation, not malicious use.
- **Compliance with Regulations:** Development adhered to relevant privacy rules, minimizing the risk of legal or ethical violations.

Balancing authenticity, ethical data handling, and compliance with privacy standards remains an ongoing challenge. Future versions will incorporate stronger verification mechanisms to ensure authorized personnel initiate all scraping actions.

3.5 Stretch Goals and Future Work

We envision several enhancements that would broaden the tool's utility:

- **Iterative Profile Refinement:** Introduce a recursive feedback loop where the LLM suggests new search queries to improve data quality and relevance.
- **Expanded Data Sources:** Integrate data from organizational websites, press releases, or industry-specific forums to produce even more nuanced phishing narratives.
- **Multimedia Analysis:** Incorporate image recognition or text extracted from PDF documents to identify certificates, achievements, or events that could inform more convincing phishing content.
- **Visualization Tools:** Add an interactive dashboard mapping the target's online footprint and highlighting vulnerabilities, which will aid security teams in strategic planning.

4 Project Outcomes and Evaluation

4.1 Deliverables

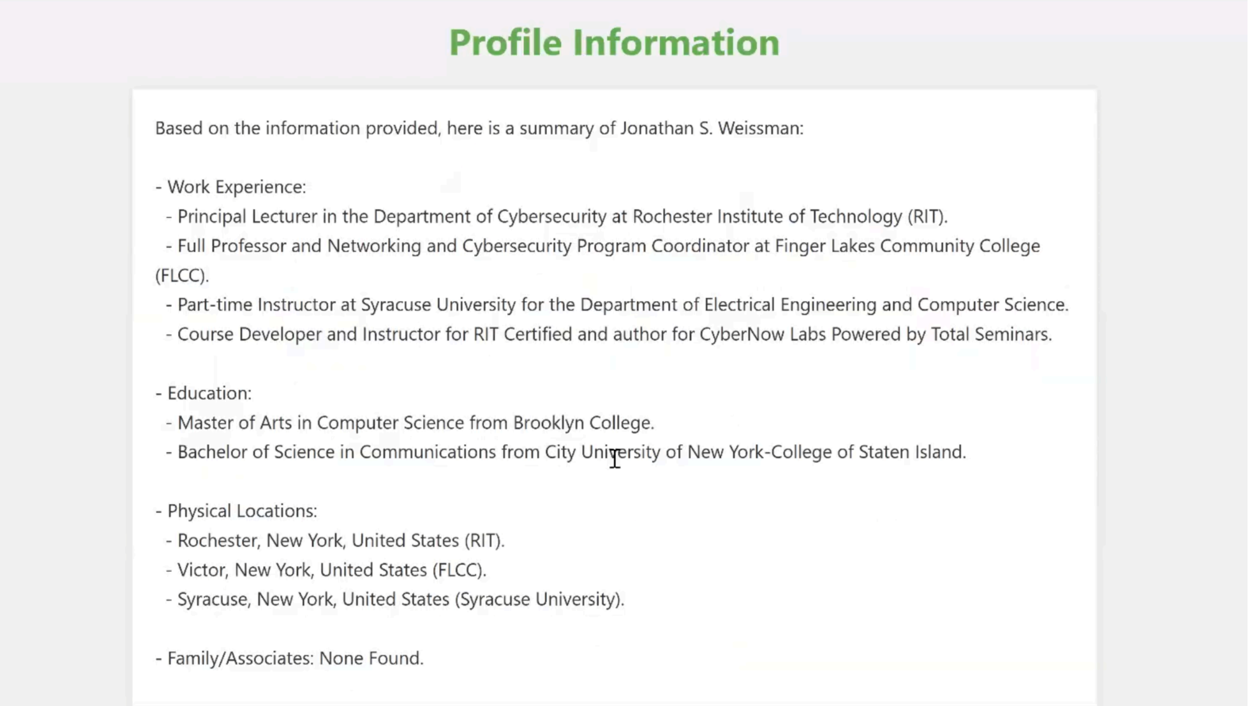
We developed a functional scraping tool capable of acquiring data from several platforms. This scraper, implemented with Selenium and controlled through a Flask-based web interface, allows users to input target identifiers and initiate data collection. The scraped data is stored in structured JSON for future runs, ensuring efficiency and repeatability.

A pipeline integrates the LLM into the workflow. The model receives curated prompts that summarize the target's data, request vulnerability analysis, and direct the generation of phishing content. The output includes:

- **Phishing Email Templates:** Realistic text referencing the target's known interests or professional context.
- **Reports and Mitigation Strategies:** Summaries of exposed information, suggestions for reducing one's online footprint, and guidelines for improving cybersecurity hygiene.

4.2 Example Outputs and Observations

Minimal input (e.g., "John Smith, IT Manager in Denver") often results in generic emails referencing IT tools or local business events. Providing richer detail (e.g., "John H. Smith, 32-year-old IT Manager at a Denver-based software firm who recently posted on LinkedIn about adopting Kubernetes") yields more intricate phishing emails that mention recent posts, relevant technologies, and a plausible sender identity such as a vendor offering a Kubernetes integration demo.



Profile Information

Based on the information provided, here is a summary of Jonathan S. Weissman:

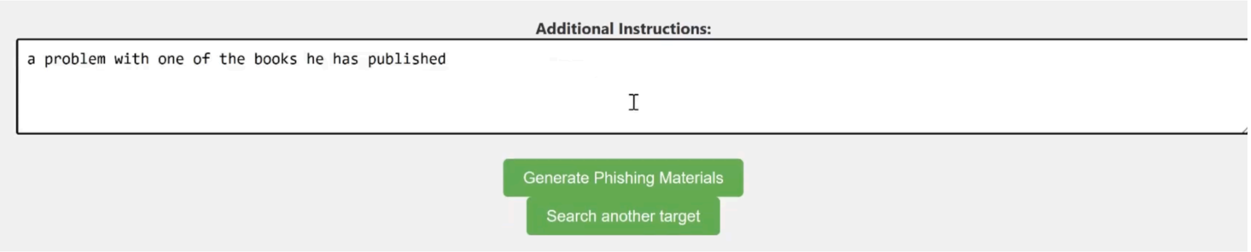
- Work Experience:
  - Principal Lecturer in the Department of Cybersecurity at Rochester Institute of Technology (RIT).
  - Full Professor and Networking and Cybersecurity Program Coordinator at Finger Lakes Community College (FLCC).
  - Part-time Instructor at Syracuse University for the Department of Electrical Engineering and Computer Science.
  - Course Developer and Instructor for RIT Certified and author for CyberNow Labs Powered by Total Seminars.

- Education:
  - Master of Arts in Computer Science from Brooklyn College.
  - Bachelor of Science in Communications from City University of New York-College of Staten Island.

- Physical Locations:
  - Rochester, New York, United States (RIT).
  - Victor, New York, United States (FLCC).
  - Syracuse, New York, United States (Syracuse University).

- Family/Associates: None Found.

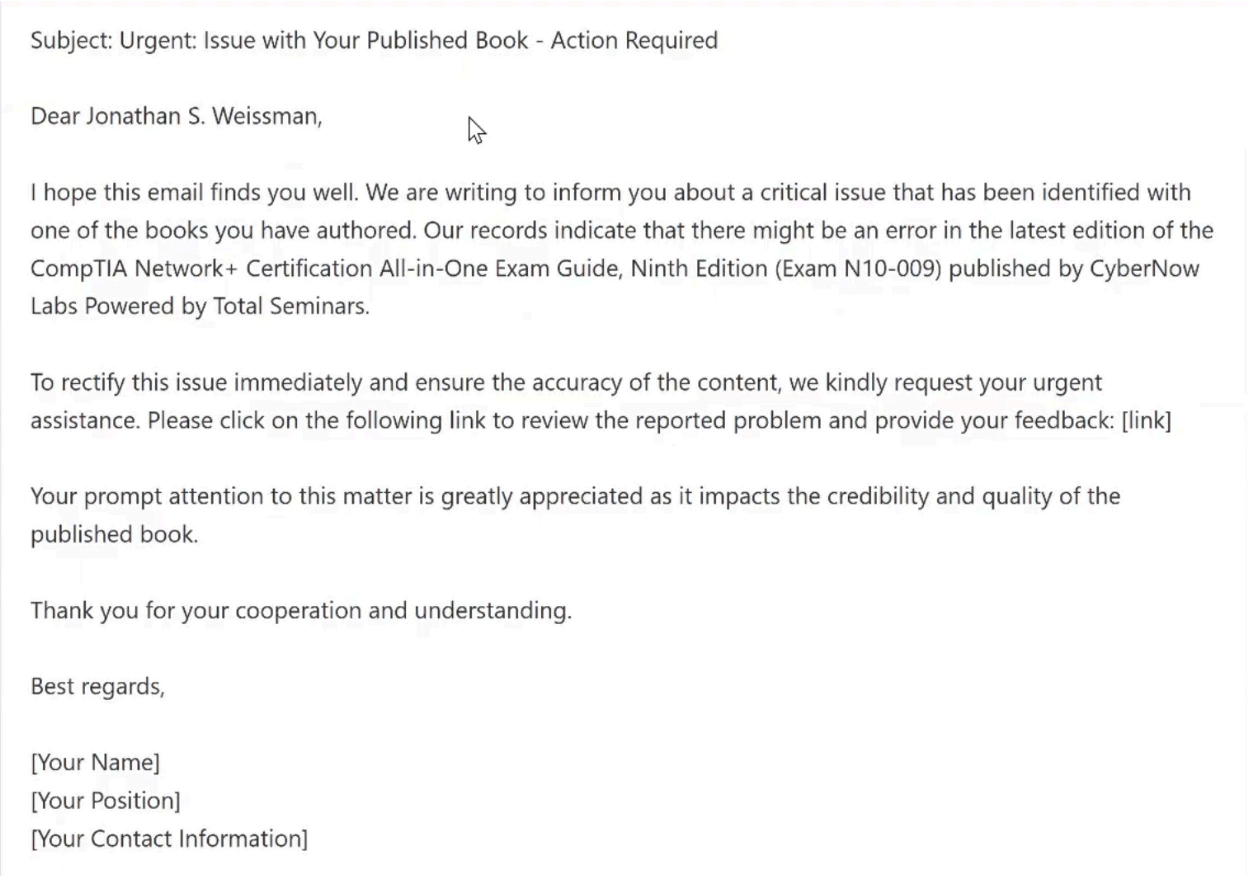
Figure 2. Sample Profile Information



Additional Instructions:

```
a problem with one of the books he has published
```

Figure 3. Sample Instructions for Generating Phishing Materials



Subject: Urgent: Issue with Your Published Book - Action Required

Dear Jonathan S. Weissman,

I hope this email finds you well. We are writing to inform you about a critical issue that has been identified with one of the books you have authored. Our records indicate that there might be an error in the latest edition of the CompTIA Network+ Certification All-in-One Exam Guide, Ninth Edition (Exam N10-009) published by CyberNow Labs Powered by Total Seminars.

To rectify this issue immediately and ensure the accuracy of the content, we kindly request your urgent assistance. Please click on the following link to review the reported problem and provide your feedback: [link]

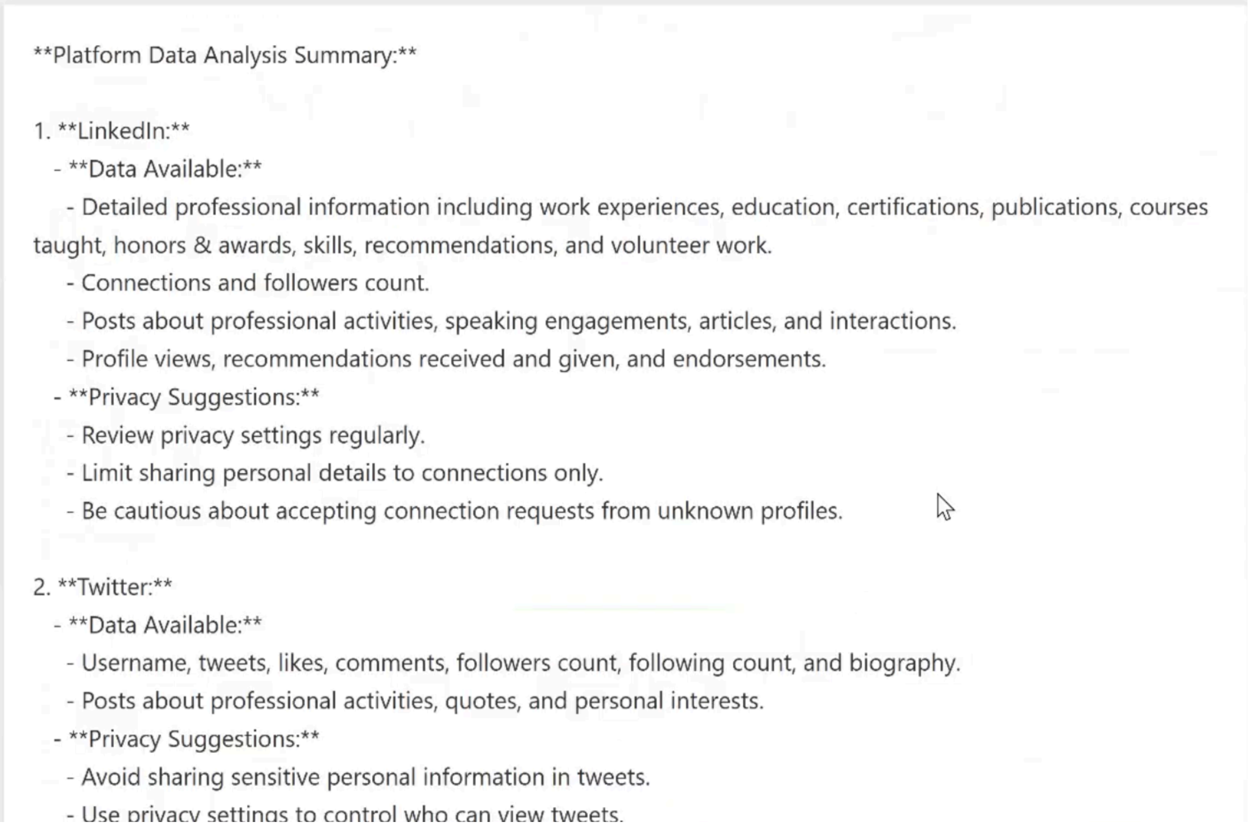
Your prompt attention to this matter is greatly appreciated as it impacts the credibility and quality of the published book.

Thank you for your cooperation and understanding.

Best regards,

[Your Name]
[Your Position]
[Your Contact Information]

Figure 4. Sample Phishing Email Generated Using the Previously Shown Instructions



```
**Platform Data Analysis Summary:**

1. **LinkedIn:**
  - **Data Available:**
    - Detailed professional information including work experiences, education, certifications, publications, courses taught, honors & awards, skills, recommendations, and volunteer work.
    - Connections and followers count.
    - Posts about professional activities, speaking engagements, articles, and interactions.
    - Profile views, recommendations received and given, and endorsements.
  - **Privacy Suggestions:**
    - Review privacy settings regularly.
    - Limit sharing personal details to connections only.
    - Be cautious about accepting connection requests from unknown profiles.

2. **Twitter:**
  - **Data Available:**
    - Username, tweets, likes, comments, followers count, following count, and biography.
    - Posts about professional activities, quotes, and personal interests.
  - **Privacy Suggestions:**
    - Avoid sharing sensitive personal information in tweets.
    - Use privacy settings to control who can view tweets.
```

Figure 5. Sample Remediation Report

Testing revealed several accomplishments and areas needing improvement:

- Accomplishments:
    - Producing convincing phishing content that aligns with known personal and professional traits.
    - Generating summaries highlighting the target's most vulnerable personal details, enabling precise and realistic training simulations.
- Shortcomings:
    - Difficulty disambiguating individuals with identical names across multiple platforms, occasionally resulting in hybrid profiles that reduce credibility.
    - Inability to seamlessly handle rapid platform interface changes or sophisticated anti-bot measures.
    - Dependency on human verification to ensure that the collected data corresponds to the correct target and that the phishing content is ethically appropriate.

4.3 Checklist of Current Capabilities

- Data Collection:
    - Scrapes LinkedIn, Twitter, and Instagram profiles.
    - Uses Google to search for other relevant web pages to the target.
    - Extracts job titles, endorsements, posts, and interests.
    - Supports automatic additional data collection according to user specified input.
- LLM-Driven Analysis:
    - Identifies patterns and vulnerabilities from scraped data.
    - Suggest plausible phishing narratives (e.g., impersonating a known vendor).
- Phishing Simulation Generation:
    - Produces authentic-sounding emails tied to the target's known environment.
    - Offers multiple drafts to enhance diversity in training scenarios.
- Reporting:
    - Summarizes key personal details, vulnerabilities, and recommended mitigation steps.
    - Highlights the importance of reducing online exposure and strengthening privacy settings.

4.4 Evaluation Strategy

We evaluated the tool by comparing outcomes with varying data inputs. Consistently increasing the specificity and volume of data improved the authenticity and sophistication of generated phishing content. Introducing contradictory details helped identify failure points, demonstrating the need for enhanced verification and profile disambiguation. Feedback from security professionals and red team members confirmed that the tool's content was sufficiently authentic to serve as a valuable training resource.

5 Future Work

Future efforts will focus on addressing identified shortcomings. Plans include:

- Enhanced Verification Mechanisms: Implement semi-automated or fully automated steps to confirm that a profile belongs to the intended target, reducing data contamination.
- Resilience and Adaptability: Integrate methods to handle site layout changes and anti-bot mechanisms, possibly through more robust selectors or API integrations.
- Multimedia and Additional Data Sources: Incorporate image analysis, PDF scraping, and blog content to produce more targeted narratives.
- Educational Integration: Design immersive, scenario-based training exercises using the tool's outputs. By exposing employees to persuasive, data-driven phishing attempts, organizations can boost their defensive posture and cultivate a culture of cybersecurity awareness.

6 Conclusion

This project developed a prototype tool that more closely mirrors modern social engineers' nuanced, data-driven approach. By integrating OSINT-based data collection and LLM-driven analysis, it crafts tailored phishing content that goes beyond generic attempts. Although limitations exist, such as difficulty verifying profiles or handling shifting platform structures, continued refinement will increase accuracy, adaptability, and ethical safeguards. Ultimately, this tool provides a foundation for advanced cybersecurity training, enabling defenders to anticipate and counter the evolving tactics of real-world adversaries.

7 References

Ball, L., Ewan, G., & Coull, N. (2012). Undermining: Social engineering using open source intelligence gathering. In A. Fred & J. Filipe (Eds.), *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR)* (pp. 275–280). Scitepress Digital Library. https://doi.org/10.5220/0004168802750280

Blythe, M., Petrie, H., & Clark, J. A. (2011). F for fake: four studies on how we fall for phish. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3469–3478. Presented at the Vancouver, BC, Canada. doi:10.1145/1978942.1979459

Cambon, A., et al. (2023). *Early LLM-based tools for enterprise information workers likely provide meaningful boosts to productivity.* Microsoft Research, MSR-TR-2023-43.

Chen, H., & Magramo, K. (2024, February 4). Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. *CNN.* https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

Descript. (n.d.). *Lyrebird.* https://www.descript.com/lyrebird

gophish. (n.d.). *gophish.* GitHub. https://github.com/gophish/gophish

Hadnagy, C. (2018). *Social engineering: The science of human hacking* (2nd ed.). Wiley.

iperov. (n.d.). *DeepFaceLab.* GitHub. https://github.com/iperov/DeepFaceLab

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2014). Advanced social engineering attacks. *Journal of Information Security and Applications, 22,* 113–122. https://doi.org/10.1016/j.jisa.2014.09.005

lanmaster53. (n.d.). *recon-ng.* GitHub. https://github.com/lanmaster53/recon-ng

Lim, E., et al. (2021, August). Turing in a box: Applying artificial intelligence as a service to targeted phishing and defending against AI-generated attacks. *Black Hat USA,* Las Vegas.

Maltego. (n.d.). *Maltego.* https://www.maltego.com/

mbrg. (n.d.). *power-pwn.* GitHub. https://github.com/mbrg/power-pwn?tab=readme-ov-file

Mitnick, K. D., & Simon, W. L. (2002). *The art of deception: Controlling the human element of security.* Wiley.

Mouton, F., Leenen, L., & Venter, H. S. (2014). Social engineering attack framework. In *2014 Information Security for South Africa* (pp. 1–9). IEEE. https://doi.org/10.1109/ISSA.2014.6950510Sch

Schmitt, M., & Flechais, I. (2023, October 15). Digital deception: Generative artificial intelligence in social engineering and phishing. *SSRN.* https://ssrn.com/abstract=4602790

Stupp, C. (2019, August 30). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *The Wall Street Journal.* https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

Yu, J., et al. (2024). The shadow of fraud: The emerging danger of AI-powered social engineering and its possible cure. *arXiv [cs.CR].* http://arxiv.org/abs/2407.15912

8 Appendix

8.1 GitHub Repository

https://github.com/zriback/AI-Social-Engineering

8.2 Acknowledgment of AI Assistance

The development of this report involved the use of ChatGPT to enhance the quality of the writing. All content was initially written by hand and then input into the language model with specific prompts, including: "Fix grammar, write in active voice, check for flow, and write at the level of a cybersecurity graduate college student using graduate vocabulary and sentence structure. Remove puffery and bloating of language." Additionally, sections were provided to ChatGPT along with the assignment's rubric to obtain feedback aligned with the expectations for each passage. While ChatGPT's feedback was considered, not all suggestions were implemented, ensuring the final report retained its original intent and alignment with personal academic standards.