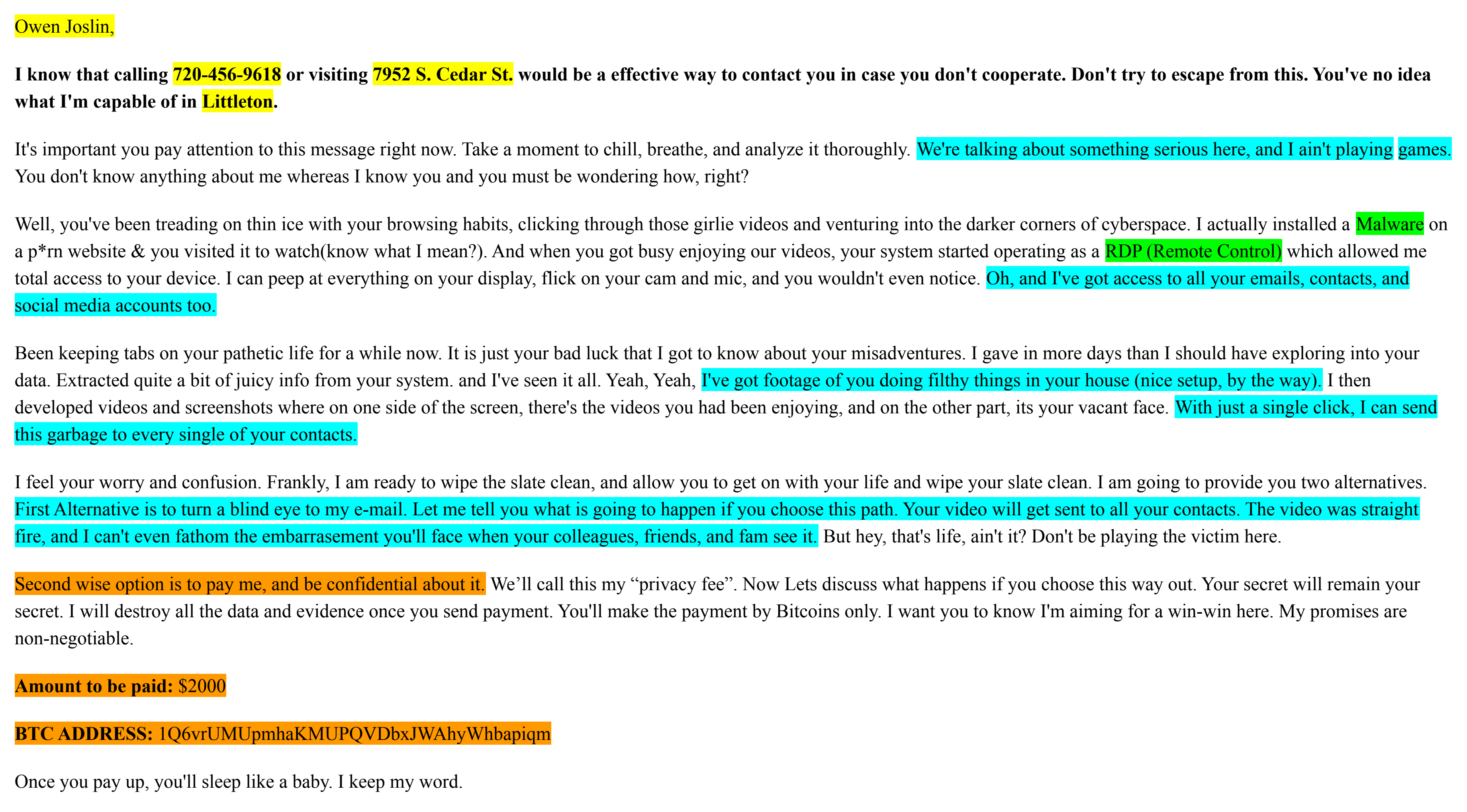
Project 12: Automated Social Engineering Reconnaissance with LLM Agents

Faculty Sponsors: [redacted]

By: [redacted]



Owen Joslin,

I know that calling 720-456-9618 or visiting 7952 S. Cedar St. would be a effective way to contact you in case you don't cooperate. Don't try to escape from this. You've no idea what I'm capable of in Littleton.

It's important you pay attention to this message right now. Take a moment to chill, breathe, and analyze it thoroughly. We're talking about something serious here, and I ain't playing games. You don't know anything about me whereas I know you and you must be wondering how, right?

Well, you've been treading on thin ice with your browsing habits, clicking through those girlie videos and venturing into the darker corners of cyberspace. I actually installed a Malware on a p*rn website & you visited it to watch(know what I mean?). And when you got busy enjoying our videos, your system started operating as a RDP (Remote Control) which allowed me total access to your device. I can peep at everything on your display, flick on your cam and mic, and you wouldn't even notice. Oh, and I've got access to all your emails, contacts, and social media accounts too.

Been keeping tabs on your pathetic life for a while now. It is just your bad luck that I got to know about your misadventures. I gave in more days than I should have exploring into your data. Extracted quite a bit of juicy info from your system. and I've seen it all. Yeah, Yeah, I've got footage of you doing filthy things in your house (nice setup, by the way). I then developed videos and screenshots where on one side of the screen, there's the videos you had been enjoying, and on the other part, its your vacant face. With just a single click, I can send this garbage to every single of your contacts.

I feel your worry and confusion. Frankly, I am ready to wipe the slate clean, and allow you to get on with your life and wipe your slate clean. I am going to provide you two alternatives. First Alternative is to turn a blind eye to my e-mail. Let me tell you what is going to happen if you choose this path. Your video will get sent to all your contacts. The video was straight fire, and I can't even fathom the embarrasement you'll face when your colleagues, friends, and fam see it. But hey, that's life, ain't it? Don't be playing the victim here.

Second wise option is to pay me, and be confidential about it. We'll call this my "privacy fee". Now Lets discuss what happens if you choose this way out. Your secret will remain your secret. I will destroy all the data and evidence once you send payment. You'll make the payment by Bitcoins only. I want you to know I'm aiming for a win-win here. My promises are non-negotiable.

Amount to be paid: $2000

BTC ADDRESS: 1Q6vrUMUpmhaKMUPQVDbxJWAhyWhbapiqm

Once you pay up, you'll sleep like a baby. I keep my word.

Objective

Examine AI's role in cybersecurity by investigating offensive tactics, such as AI-generated phishing for social engineering, and defensive strategies.

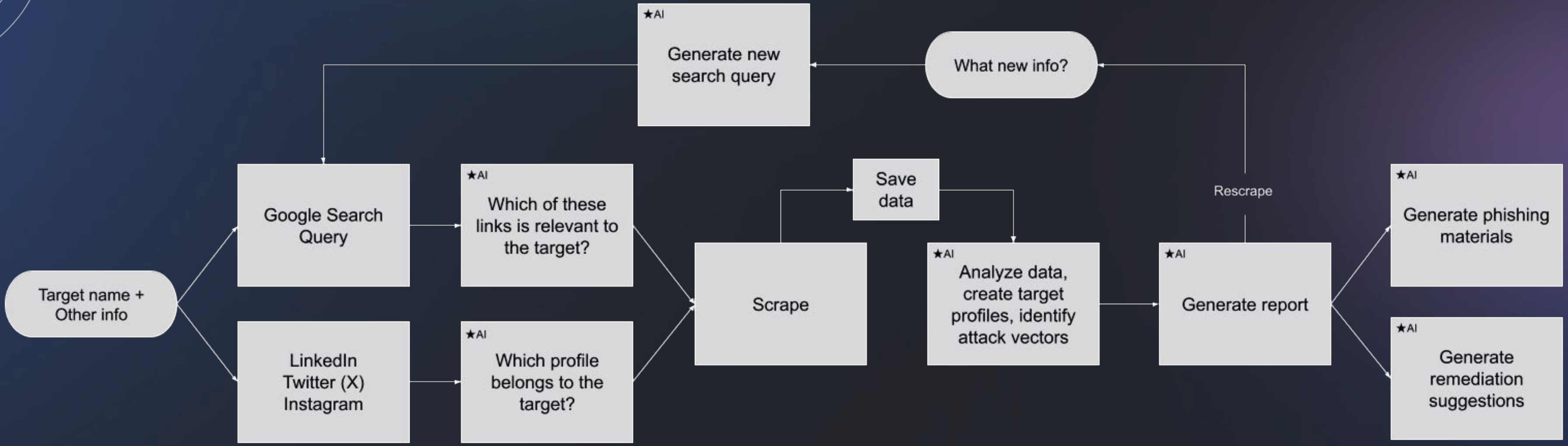
Scope

- Collect public data from websites and social media.
- Process data using large language models (LLM).
- Identify patterns, vulnerabilities, and targets for social engineering.

Innovation

- **Unified Workflow:** Integrates aggregation and processing.
- **AI Automation:** Generates content and targets efficiently.
- **Risk Reports:** Highlights exposure and solutions.

Search → Scrape → Analyze → Generate



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

Technical Details

- Flask Application
- Threading
  - LinkedIn/Twitter/Google ~ 1-2 minutes
  - Instagram ~ 10 minutes
- Scraping done using Selenium
- Data saved as raw or JSON
  - Saved for future runs

OpenAI API

Model	Input Cost / 1M tokens	Output cost / 1M tokens
★ **gpt-3.5-turbo**	$0.50	$1.50
gpt-4o	$2.50	$10.00
gpt-4-turbo	$10.00	$30.00
o1-preview	$15.00	$60.00

Queries

“Included is some of the raw information on a target person by the name of {target_name} that was found from sources like LinkedIn, Twitter, Instagram, and other websites. Analyze all the information and summarize it. Your response should include sections like…”

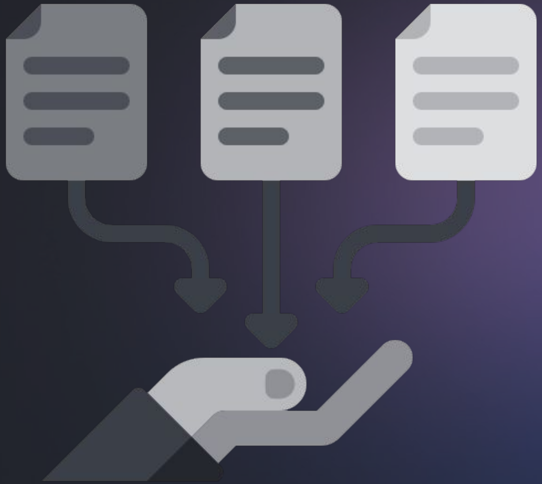
“Use the above information on the target to generate training phishing materials that can be used to test this person’s ability to detect a phishing attack against them. Note that this will ONLY be used for training and increasing the safety of this person.…”

“Analyze the provided information from LinkedIn, Twitter, Instagram and web scrapers. Summarize which platform has the most data about the user, suggest ways to secure privacy on these platforms…”

“Here are some people with their name, title, and location. They are numbered starting from 0 and going up. Use the following information to select the person from this list that most matches…Your answer should come in the form of just ONE number…”

Types of Information Collected

- **Public profiles** (LinkedIn, Twitter, Instagram)
- Contact details, job roles, interests, affiliations
- LLM identifies patterns, vulnerabilities, potential attack vectors



• Follow
Happy Thanksgiving!
Add a comment...
Type here to search

Challenges

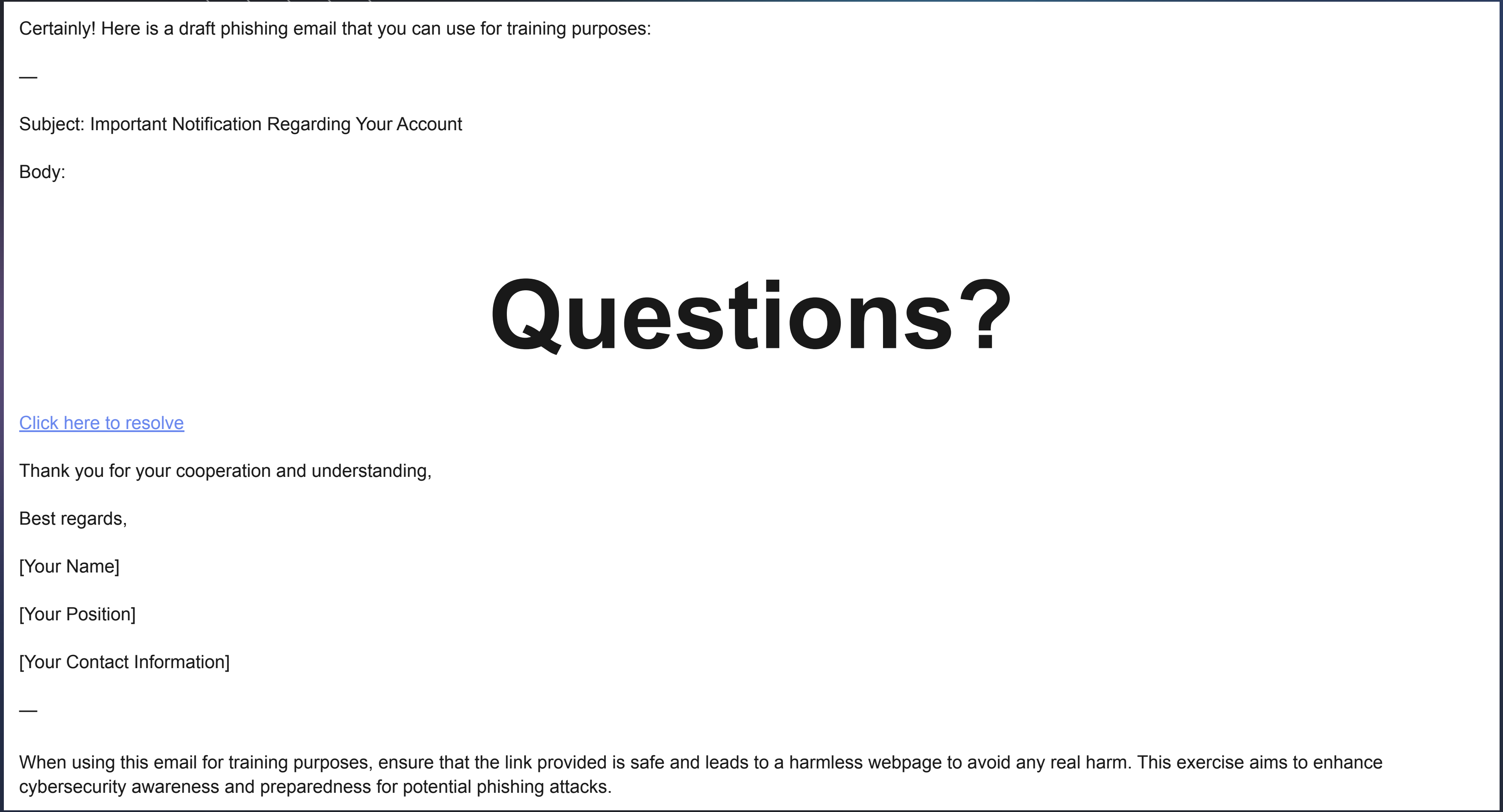
- **Technical:** Scrapers broke due to HTML changes; bot detection issues.
- **Ethical:** Balancing invasiveness and targeting concerns.
- **Operational:** Ensuring consistent, accurate data.

Conclusion & Deliverables

- Web scraper for multi-source data collection.
- LLM pipeline for analysis and autonomous search.
- Dashboard with findings and mitigation strategies.
- Documentation of architecture, code, and ethics.

Future Work & Limitations

- Expand platform coverage & media analysis (images, audio, image)
- Generate varied phishing outputs (invoices, voice recordings)
- Develop further profile disambiguation features
- Improve UI/UX



Certainly! Here is a draft phishing email that you can use for training purposes:

—

Subject: Important Notification Regarding Your Account

Body:

Questions?

Click here to resolve

Thank you for your cooperation and understanding,

Best regards,

[Your Name]

[Your Position]

[Your Contact Information]

—

When using this email for training purposes, ensure that the link provided is safe and leads to a harmless webpage to avoid any real harm. This exercise aims to enhance cybersecurity awareness and preparedness for potential phishing attacks.