Department of Computer Science & Engineering
The University of Texas at Arlington

Project Charter
CSE 4316: Senior Design I
Fall 2023

Team Honeycomb
HoneyTrap

Jair Rea
Raed Ali
Hugo Mendoza
Praisy Daniel
Joshua Catalan

Revision History

Revision	Date	Author(s)	Description
0.1	10.01.2021	GH	document creation
0.2	10.05.2021	AT, GH	complete draft
0.3	10.12.2021	AT, GH	release candidate 1
1.0	10.20.2021	AT, GH, CB	official release
1.1	10.31.2021	AL	added customer change requests

Contents

List of Figures

1 Problem Statement

As a team, we are addressing an issue of financial fraud in the digital landscape that has been ongoing for many years. Some examples of these fraudulent activities includes identity theft, and unauthorized transactions. As a result of these problems, our team believes that there is a critical need for an efficient solution to effectively detect and alert suspicious and potential fraud that may be happening. By solving this issue, We hope to help and support target groups, such as the elderly and any newcomer that may be experiencing the internet for the first time, not fall into any type of deceptive scheme.

2 Methodology

The team plans to build a solution to this overarching problem, which is to develop and implement a real-time anomaly detection system that can analyze and alert any suspicious patterns. The system/software will efficiently detect and defend against any type of fraudulent activity as well as mitigate loss of credentials or finances.

3 Value Proposition

With this project, we have many value points that will benefit our sponsor, instructor, and university. By supporting this project, our sponsor contributes to the development of software that will significantly enhance financial security. This will result in an increase of trust and preservation of brand reputation. Supporting our project also aligns with ethical values. This, as a result, demonstrates commitment to combating financial fraud and enhances cooperate image and strengthens customer trust. This also provides a base to educational growth as this partnership can lead to join research projects, internships, and recruitment. As a final addition, supporting this project also makes sponsors alike, participate in the creation of a sustainable solution to an ongoing problem. This means that the software will continue to evolve and adapt, providing lasting value and protection for our partners.

4 Development Milestones

List of milestones and completion dates:

- Project Charter first draft - III Quadrant, 2023
- System Requirements Specification - IV Quadrant, 2023
- Architectural Design Specification - IV Quadrant, 2023
- Detailed Design Specification - I Quadrant, 2024
- CoE Innovation Day poster presentation - IV Quadrant, 2024
- Final Project Demonstration - IV Quadrant, 2024

5 Background

The sponsor's motivation for pursuing the honeypot project is deeply rooted in a personal experience that highlights a pressing issue and a genuine concern for the vulnerable elderly population. The project's initial idea stems from the sponsor's Great Aunt's unfortunate experience. She found herself being targeted by scammers on social media after a certain phase of her life and the scammers preyed on her vulnerability, ultimately causing her to fall victim to their schemes, resulting in a significant financial loss of over $15,000.The project aims to tackle this problem by using technology to interact with scammers, learn their tricks, and raise awareness about the dangers of online scams. The project envisions leveraging artificial intelligence and other technological tools to engage with scammers, thereby shedding light on their tactics and offering a semblance of protection for potential victims. Here are some key points that justify undertaking the honeypot project:

- Early Warning System: The project can serve as an early warning system, alerting security teams to potential threats and attacks in real time. This enables the mitigation of potential damage.

- Rising Scam Epidemic: The project could raise awareness about the dangers of online scams and the tactics used by scammers. It may encourage people to be more cautious and vigilant online.

- Reputation and Trust: For social media platforms and online communities, the prevalence of scams can erode trust and reputation. Addressing this issue can lead to improved user trust and potentially attract more users, benefiting social media platforms from a business perspective.

- Data-Driven Insights: The project can generate valuable data and insights into scam tactics, which can be shared with relevant authorities and organizations to improve scam prevention measures.

6 Related Work

A few related works that are currently available are as follows:

- Facebook's Deep Entity Classification (DEC) technology - The DEC combats the proliferation of fake accounts on the platform by differentiating fake and real users by their connection patterns across the network. DEC uses hand-coded rules and machine learning to identify fake accounts either before they are created or before they become active. This stage aims to prevent fake accounts from causing harm to real users [1]. Although the system aims to identify fake profiles accurately, DEC primarily relies on analyzing connection patterns and deep features of accounts. Scammers can be highly adaptive and use persuasive language and social engineering tactics in their messages. DEC may not fully grasp the nuances of these interactions and could potentially miss certain scam attempts.

- Natural Language Processing (NLP) - NLP algorithms are trained to recognize specific language patterns commonly associated with fraudulent activities. These patterns may include persuasive or coercive language, urgency, and emotional manipulation. By scanning text-based messages, NLP can flag conversations that exhibit these patterns as potentially suspicious. However, if the scammers target the elderly population who primarily communicate using non-textual methods, such as voice calls or images, NLP may not be able to analyze and detect their fraudulent activities effectively [2].

- Bitdefender TrafficLight - It is a browser extension designed to enhance online security and protect users against malicious threats. It works in real-time to identify and block phishing attempts, malware distribution, and fraudulent activities. While Bitdefender TrafficLight is a valuable tool to enhance online security, it does not have the capability to analyze the content of social media posts or messages comprehensively like NLP.

7 System Overview

Elderly people are often targeted by scammers attempting to defraud them. This poses a threat to their financial security and well-being. To address this problem, we propose to use an integrated solution that capitalizes on an LLM-based system. A Large Language Model (LLM) is a type of artificial intelligence model that is trained to take in a vast amount of text data to understand and generate human like text. This adaptable approach will utilize a chatbot, automated case reporting, and social media integration. The chatbot, powered by an LLM, assumes the role of an elderly person, and will effectively emulate a potential victim for a scam. The chatbot will engage with scammers with the intent of eliciting contact information such as payment destinations, name, addresses, and other important information. By posing as a vulnerable target, the chatbot will aim to lure in scammers into inadvertently revealing information that identifies them and stops their fraudulent activities. In the event that a scammer has made their move, the system will initiate the process of automatic case reporting. This automation will involve the generation of a comprehensive case report that includes a detailed history of the scam attempt and all the evidence that supports this claim. The report will then be sent to law enforcement which will provide them the necessary information to investigate the scam attempt. For the social media integration, the chatbot should be connected to platforms commonly used by the elderly. This entails setting up a Facebook account to mimic the online presence of an elderly individual. This approach requires a gradual posting strategy where pictures and other posts are shared over time. The chatbot should generate text-based posts that might reflect sentiments such as a missing spouse, loneliness, depression, or missing their children/grandchildren. The account should also be diversified by sharing memes or articles instead of just photos and text-based posts. To make the account more authentic, the chatbot will engage in comments offering explanations for the relatively young age of the profile. The explanations will be along the lines of, my other Facebook got hacked so I had to start a new one or I finally got on Facebook after a friend told me to.

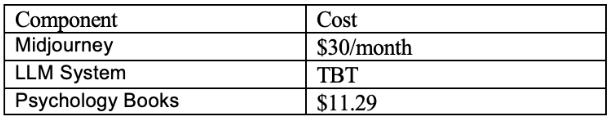
8 Roles & Responsibilities

The stake holder for this project is Michael Magnus, Sr. SEO Specialist, from Mouser Electronics. Michael has provided us with the necessary background information about the scam that happened to his great aunt along with a list of specifications/requirements that should be included in the project. He also provided us with a list of conversation notes that were said to his great aunt to lure her into the scam. This information is to help assist in identifying scammers and the approach they use when trying to contact or lure in potential victims. We have established our point of contact to be Jair Rea who will keep in touch with Mr. Magnus. He will relay any important information that Mr. Magnus wishes to be a part of the project along with asking clarification on how certain features or functionalities should be utilized. The team on this project consists of Jair Rea, Raed Ali, Hugo Mendoza, Praisy Daniel, and Joshua Catalan. We will divide and conquer to make sure every aspect of the project is making progress. We will also host meetings to keep track of where everyone is so that we can divert our attention to the parts of the project that need more work as a whole. Throughout this project we will maintain the product owner and we will rotate the scrum master if need be. Rotating the scrum master when needed will help elevate some responsibilities from people and allow for all of us to give our own strategies for tasks.

9 Cost Proposal

For this project the costs will revolve around any software licenses that may be needed when we enter the development stage. The budget would have to cover data acquisition and infrastructure. Building the chatbot using an LLM-based system would require most of the cost.

9.1 Preliminary Budget



Component	Cost
Midjourney	$30/month
LLM System	TBT
Psychology Books	$11.29

9.2 Current & Pending Support

For this project, the funding will be provided by UTA CSE department and Michael Magnus from Mouser Electronics. The CSE department will provide an 800 dollar budget any other expenses will be covered by Mouser Electronics through Michael Magnus.

10 Facilities & Equipment

When it comes to physical equipment we won't need much except for our computers. This is a heavily software-based project after all. However, on the software side, we need many different tools and websites to complete this project. We will need to use Facebook and a couple of Facebook accounts to test out our product. For these Facebook accounts, we will need convincing profile pictures to represent the scammer's targets so we will use the AI image generation website Midjourney to generate a series of older women to use for our fake Facebook accounts. We are doing this since we could get in some legal trouble if we used real people. Now we will use Visual Studio Code to write our code. So our team can properly collaborate with each other we will be using GitHub to share and manage our files. We will be using Python to code our chatbot and we need the appropriate Python libraries to do many of the required features. We need to connect the chatbot to Facebook and use the appropriate LLM libraries to help make our chatbot act like an elderly woman. Plus we would need any documentation on these libraries and some YouTube tutorials to help with the development of the LLM part of our bot. The same thing goes for the other libraries we are using, we would need the documentation of the functions within the libraries plus any YouTube tutorials to help better understand how they work. We would also need the appropriate libraries for filing police reports after we have detected a scammer. We would also need some web-scraping libraries to analyze the profiles of other Facebook accounts so we can target and analyze the scammer's "friends" too so we can find more scammers to target. And finally, we would need our test subjects, actual scammers to test the product's capabilities. Now from the email from our sponsor, he recommended that we do some research in psychology to help make a convincing chatbot and even recommended a few books, like âNever Split the Differenceâ by Chris Voss. So we may need to buy a few books on psychology or at the very least watch some videos on psychology on YouTube to help guide us.

11 Assumptions

The following list contains critical assumptions related to the implementation and testing of the project.

- That we will have enough time to both set up these fake Facebook accounts and implement the code for the chatbots and to do both in a convincing and natural manner
- Scammers on Facebook will be able to find our fake accounts and start chatting with our bot
- The scammers that do chat with our bot won't realize the account is a bot.
- We will find sufficient chat logs/ data to feed our chatbot to create a convincing widow
- We will be able to provide the appropriate online resources to make a chatbot to mimic an old widow

- If a police report is filed it won't be filed on an innocent person

12 Constraints

The following list contains key constraints related to the implementation and testing of the project.

- Final prototype demonstration must be completed by May 1st, 2024
- We are full-time UTA students with other classes, exams, and projects, and some of us even have jobs so we would have to balance our schedules to work on this project.
- To truly test the bot it will take a while since we need to make several Facebook accounts and make them seem natural and gradually grow the accounts while also working on the code for the chat bot
- Total development costs must not exceed $800
- The project requires us to use an LLM approach to make our bot. Now any limitations there may be in the machine learning libraries/ algorithms we will be using, these limitations may make it seem like the scammer is talking to a bot rather than a real person. We would have to overcome these limitations and code our bot with these in mind.

13 Risks

This section should contain a list of at least 5 of the most critical risks related to your project. Additionally, the probability of occurrence, size of loss, and risk exposure should be listed. For size of loss, express units as the number of days by which the project schedule would be delayed. For risk exposure, multiply the size of loss by the probability of occurrence to obtain the exposure in days. For example:

The following high-level risk census contains identified project risks with the highest exposure. Mitigation strategies will be discussed in future planning sessions.

Risk description	**Probability**	**Loss (days)**	**Exposure (days)**
failure to meet important deadlines	0.50	20	10
Delay with working through other potential problems that may arise	0.30	20	6
failure to create a good LLM for chatbot (chance it does not sound human)	0.05	10	.50
Could potentially detect/flag someone wrongfully	0.05	5	.25
May run into issues libraries to build LLM	0.10	5	1

Table 1: Overview of highest exposure project risks

14 Documentation & Reporting

14.1 Major Documentation Deliverables

14.1.1 Project Charter

This document will be maintained as needed on a regular basis. We will set deadlines for ourselves and make sure to update everything as needed. It will range from a daily basis to weekly basis. The initial charter will be delivered 9/26/2023. Our final charter will be delivered 5/1/2024.

14.1.2 System Requirements Specification

This document will be maintained on overleaf/LATEX. Every member of the group has access and will make updates as needed. For system requirement specification, we will first discuss how to approach this within our group and then document it when we reach a consensus for how to best handle this. We can expect an initial version to be made in the next month or two. And the final version will be done 5/1/2024.

14.1.3 Architectural Design Specification

In the beginning there will be many changes as the team decides the best approach for the project. It will be updated at least on a weekly basis in the beginning. Once the project gets started, it will also be updated frequently if we see flaws in the architectural design. The initial version will be delivered within the first month or two and the final version will be delivered 5/1/2024.

14.1.4 Detailed Design Specification

The detailed design specification may take more time than the system and architectural specifications. This will be maintained when the team is able to determine the best direction for the project to go in, and update accordingly. We will approach this as a group and have discussions and then update the design specification as we go. The initial specification will be complete in 2-3 months and the final specification will be ready by 5/1/2024.

14.2 Recurring Sprint Items

14.2.1 Product Backlog

Items will be added to the product backlog when our team can decide what is needed. The decision will be based on a team vote. Our team will determine which software is best to maintain this data at a future time.

14.2.2 Sprint Planning

We will create an outline of the whole project with an idea of how long each task will take to complete the whole project. From there we will divide the work into 8 sprints with a goal of completing the goal by the 8th sprint.

14.2.3 Sprint Goal

We will decide as a team what our sprint goal is. We will be in constant communication with the customer and ask for input when needed.

14.2.4 Sprint Backlog

Our team will discuss this as well go with our project. We will maintain the software when we pick one for our project nd maintain it based on our teams performance. We may have a SCRUM board we update, or we may go into another direction.

14.2.5 Task Breakdown

The product owner will work with the team but the team will discuss what tasks are given to what team member. Tasks may be given or members can claim them if they feel they are best suited for the task. Each team member will have to document time spent on their own where we will then update it on a document that holds the data.

14.2.6 Sprint Burn Down Charts

The burn down chart will be a team effort. We will rotate unless we can pick a team member who will be in charge for this. We will have some type of document where everyone can update it with their time as we go along with our project. We will use a linear chart along with a table below to display our expected vs actual deadlines to see if we are meeting expectations.

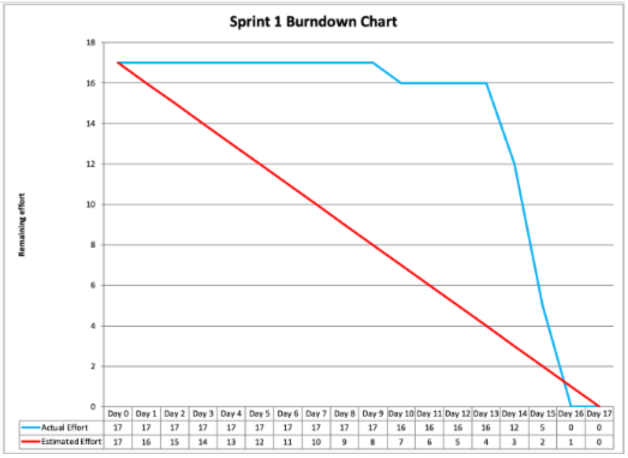


Figure 1: Example sprint burn down chart

14.2.7 Sprint Retrospective

We will schedule a meeting after every sprint where we assess our performance and see if we are ahead or behind our schedule. We will document our results as a group where we determine the best actions for the project. We will also have a document where we keep up with our performances and make actions based on our performances for the sprint. We will have a retrospective due right after each sprint.

14.2.8 Individual Status Reports

Everyone in the team will have a document where we report our status on the work we have been given. We will have topics to update such as task, time spent on task, whether the task is completed, as well as other details. We will use this data to determine if we can move ahead or see if a teammate can assist with a task. We will report after each sprint or as needed.

14.2.9 Engineering Notebooks

At a minimum the notebook will be updated on a bi-weekly to monthly basis. We will not set a minimum of pages required but we will have a minimum in terms of what we have completed in each interval. Our team leader will be the sign off witness in charge of holding ass accountable. We will determine the length of the interval as we go along.

14.3 Closeout Materials

14.3.1 System Prototype

For our final system prototype, we will provide all of the necessary documentation for our project. The demonstration will be 5/1/2023. We will have a prototype acceptance test.

14.3.2 Project Poster

Our poster will be delivered on 5/1/2024 and we will decide what information to post on it as we get close to finishing the project. We expect it to be a 3 foot by 3 foot poster.

14.3.3 Web Page

Since our project will work with fake facebook accounts we create, it will not be like an âactive webpage.â It will not be accessible to the public, only to the investor. It will be delivered on 5/1/2024. If our project is not completely available by closeout, we will have a video demo ready by then.

14.3.4 Demo Video

The demo video will contain an outline of the problem we are solving and the product. We will show how it is used and show the different functionalities of it. We are not sure of the length of the video yet.

14.3.5 Source Code

We will run and maintain it on Github. We are not sure whether the source code or the binaries only will be provided yet. We have not determined if we will let it be open source either.

14.3.6 Source Code Documentation

We will use comments as well as diagrams to show how the code works as well as to explain aspects of the code such as functions classes variable. For documentation we will use Doxygen to format our documentation. We will present our final documentation in PDF format.

14.3.7 Installation Scripts

We will determine the best way to install the program once we have completed it. We intend on writing scripts for the customer to download directly onto their machine and be ready to use. We will create a folder/package with everything necessary for the customer to use.

14.3.8 User Manual

We will try to provide a physical and digital user manual to make it as simple as possible for the customer to use and understand. We will determine if a video is necessary after creating the manual. We will also ask the customer if they would like video.

References

[1] Hao, Karen. MIT Technology Review, "How Facebook Uses Machine Learning to Detect Fake Accounts," MIT Technology Review, 04-Mar-2020.

[2] Putatunda, Jayeeta. Indellient, "Natural Language Processing (NLP) in Fraud Analytics," Indellient, 20-Jun-2020.

[3] Statt, Nick. "Facebook's AI Moderation with DEC: Detecting Fake Accounts and Scams," The Verge, 04-Mar-2020.

[4] "What Is Natural Language Processing (NLP)?", Statistical Analysis System Inc.

[5] Hutchinson, Andrew, "Facebook's Improved Fake Account Detection Tool Has Seen the Removal of 66 Billion Fake Profiles," Social Media Today, 06-Mar-2020.