Ask Captain Cyber

Design Document

Team number: sdmay25-07

Client: Doug Jacobson

Advisor: Doug Jacobson

Team Members/Roles:
Alex Elsner - Backend Developer
Alex Kronau - Full Stack Developer
Caden Murphy - Frontend Developer
Casper Run - Cybersecurity/WordPress Developer
Ethan Comiskey - Cybersecurity Developer/Manager
Steven Ragan - Cybersecurity/AI Developer

Team Email: sdmay25-07@iastate.edu
Team website: https://sdmay25-07.sd.ece.iastate.edu/

Revised: 12/05/2024

Executive Summary

"Ask Captain Cyber" is an AI-powered chatbot that focuses on cybersecurity. This allows users of different levels of understanding to ask questions and receive answers they can understand. Users can take assurance that the answers are correct as each answer will be vetted by a cybersecurity expert. As the world becomes more interconnected, there are more cybersecurity threats than ever before. Users need a reliable way to get information about cybersecurity issues they may be facing. This could involve basic IT questions, setting up IoT devices, or securing and defending against cyber attacks. Cybercrime is rising everywhere, and anyone can be a target. This is especially true with younger and older people more susceptible to scams. Ask Captain Cyber will help to fix this problem by providing a reliable source of information for people to address these questions. Cybersecurity has become bloated with a lot of information, and it can be hard to find what you are looking for. Ask Captain Cyber will simplify this process so that even beginners can understand cybersecurity. IoT devices have become a growing cybersecurity threat. Many more people are installing IoT devices to connect their lights, homes, phones, cars, etc. These are prime targets for hackers and are big cybersecurity threats. From hacking security cameras, spying through your webcam, hijacking a smart car or house, and hacking your smart fridge. Many of these IoT devices can have a lot of vulnerabilities that will be unknown to the average user. IoT devices have limited memory and can be hard to update. This can lead to an increase in vulnerabilities and poor programming practices. Ask Captain Cyber can help users install, configure, and fully understand the security risks and vulnerabilities that might come with these devices.

Ask Captain Cyber will have cybersecurity experts vetting answers. This means that if a question is not already in the database, it will give a preliminary answer while a professional answers it. This is to ensure that the answers are truly legitimate. Modern-day AI has a problem where it is hard to determine if the information they output is accurate. Users can ask for sources, but there is no actual guarantee the aI is summarizing the information correctly. Ask Captain Cyber fixes this problem by relying on industry experts with years of experience. Ask Captain Cyber will also have a process to re-evaluate already known answers. This is because cybersecurity is an ever-growing and evolving field. So this method makes sure the information is up to date. It is essentially an open source page for tons of cybersecurity information, where we hope there are enough people who are passionate and want to update and help answer users' questions.

The final aspect of Ask Captain Cyber is that it must be easy and versatile. Users with little experience and cybersecurity knowledge should be able to use it and get the necessary answers. People who might ask more advanced questions like college students, should also be able to use it to get good information and resources. Finally, experts should be able to both contribute and be able to ask more high level questions. In conclusion, Ask Captain Cyber will be a vetted AI platform where users can get cybersecurity questions and answers from a trusted and reliable source.

Learning Summary

Development Standards & Practices Used

To ensure all the project requirements would be properly met, we needed to combine our existing knowledge base alongside new knowledge, and we learned new technologies. For the front end, we used the react framework. For the backend, we used PHP. For our AI implementation, we used Open AIs API. For our text editors and development environment, we used VS code and Local WP. For our web framework, we used WordPress. For testing and development implementation, we used Local WP. We have an agile-waterfall hybrid work model.

We also had to consider many IEEE standards when making this:

IEEE Standard for Large Language Model (LLM) Agent Interfaces IEEE P3394 fits because we are incorporating AI into Ask Captain Cyber. Hence, we must ensure our system is set up to work smoothly with the AI. We also need to ensure the AI-incorporated system can send and receive data with our chosen format.

IEEE Standard for Artificial Intelligence (AI) Model Representation, Compression, Distribution, and Management IEEE 2941 is also associated with our project since it provides guidelines for managing the AI technology we will implement. It also discusses the API framework for large-scale pre-trained AI models, which we will use in our project. By aligning ourselves with the goals of this standard, we can help promote operational efficiency and proper usage of Ask Captain Cyber.

IEEE Standard for Password-Based Public-Key Cryptographic Techniques IEEE 1363.2-2008 fits because we intend to use public and private key pairs for our ambassadors to log in and respond to asked questions. This standard will help with its integration into our system. Secondly, it will help ensure that bad actors preying on the site can't steal our data.

Summary of Requirements

- User friendly
- The website is developed with WordPress
- User profiles
- Account management
- Mitigated ethical impacts
- Appropriate AI responses
- Answer database

Applicable Courses from Iowa State University Curriculum

- COMS 3090: Software Development Practices
- COMS 2270: Object-Oriented Programming
- COMS 2280: Introduction to Data Structures
- COMS 3630: Introduction to Database Management Systems
- CYBE 2300: Cybersecurity Fundamentals
- CYBE 2340: Legal, Professional, and Ethical Issues in Cyber Systems
- CPRE 1850: Introduction to Computer Engineering and Problem Solving
- ENGL 3140: Technical Communication
- SE 3190: Construction of User Interfaces

New Skills/Knowledge acquired that was not taught in courses

One of the most important things we learned from this project related to large language models. We learned about the metrics used to evaluate LLMs, the methods used to implement them, and the potential risks of utilizing such new technologies. We also learned the importance of considering all options and balancing a software's capabilities, its monetary expense, and any other associated stipulations. Finally, we learned that projects require a lot of flexibility when it comes to planning. Designs may need to change depending on incompatible software and technologies, they can change to become more optimized, or for any number of reasons. The important aspect is that the team is ready for change as it comes and to be able to adapt to these changes and move forward.

Table of Contents

1. Introduction 7

1.1. Problem Statement 7

1.2. Intended Users 7

1.3. Empathy Maps 9

2. Requirements, Constraints, And Standards 11

2.1. Requirements & Constraints 11

2.2. Engineering Standards 11

3 Project Plan 14

3.1 Project Management/Tracking Procedures 14

3.2 Task Decomposition 14

3.3 Project Proposed Milestones, Metrics, and Evaluation Criteria 15

3.4 Project Timeline/Schedule 16

3.5 Risks and Risk Management/Mitigation 16

3.6 Personnel Effort Requirements 18

3.7 Other Resource Requirements 18

4 Design 20

4.1 Design Context 20

4.1.1 Broader Context 20

4.1.2 Prior Work/Solutions 20

4.1.3 Technical Complexity 21

4.2 Design Exploration 22

4.2.1 Design Decisions 22

4.2.2 Ideation 22

4.2.3 Decision-Making and Trade-Off 23

4.3 Proposed Design 23

4.3.1 Overview 23

4.3.2 Detailed Design and Visual(s) 24

4.3.3 Functionality 25

4.3.4 Areas of Concern and Development 25

4.4 Technology Considerations 26

4.5 Design Analysis 26

5 Testing 27

5.1 Unit Testing 27

5.2 Interface Testing 27

5.3 Integration Testing 27

5.4 System Testing 28

5.5 Regression Testing 28

5.6 Acceptance Testing 28

5.7 Security Testing 28

5.8 Results 29

6 Implementation 30

7 Ethics and Professional Responsibility 31

7.1 Areas of Professional Responsibility/Codes of Ethics 31

7.2 Four Principles 32

7.3 Virtues 33

8 Closing Material 36

8.1 Conclusion 36

8.2 References 36

8.3 Appendices 37

9 Team 38

9.1 Team Members 38

9.2 Required Skill Sets for Your Project 38

9.3 Skill Sets covered by the Team 38

9.4 Project Management Style Adopted by the team 39

9.5 Initial Project Management Roles 39

9.6 Team Contract 39

List of figures/tables/symbols/definitions (This should be the similar to the project plan)

Large Language Model (LLM): Type of AI model trained to comprehend and generate text

OpenAI GPT-3.5: Specific LLM used for generating text responses to user input

Artificial Intelligence (AI): Technology capable of simulating human intelligence such as learning, comprehension, problem solving, decision making, and more

Application Programming Interface (API): Set of rules and protocols that allow software applications to communicate with each other

API Key: Unique identifier used to authenticate requests to interact with the API

Prompt Engineering: Process of designing input queries to guide AI responses

Encryption: Data security method that alters data so it can only be deciphered by authorized parties

User Interface (UI): The way a user interacts with a device or software

User Experience (UX): The overall experience a user has with a product. This is a broad term that includes ease of use, how well the user can navigate the product, content relevance, etc.

Frontend: Software that users interact with directly

Backend: Web development that focuses on server-side code, logic, databases, and APIs that power the frontend

Database: An organized collection of data that is electronically managed

Agile: Approach to software development that prioritizes flexibility and iterative progress

React: A JavaScript library utilized for building user interfaces

WordPress: Content Management System used for developing websites

Cybercrime: Criminal activity carried out via computers and the internet

Phishing: Type of cyber attack that involves an attacker impersonating someone or something trustworthy to steal sensitive data and information

1. Introduction

1.1. Problem Statement

As the world becomes more interconnected, there are more cybersecurity threats than ever before. Users need a reliable way to get information about cybersecurity issues they may be facing. This could involve basic IT questions, setting up IoT devices, or securing and defending against cyber attacks. Cybercrime is rising everywhere, and anyone can be a target. This is especially true with younger and older people more susceptible to scams. Ask Captain Cyber will help to fix this problem by providing a reliable source of information for people to address these questions. Ask Captain Cyber is an AI-powered chatbot that focuses on cybersecurity. This allows users of different levels of understanding to ask questions and receive answers they can understand. Users can take assurance that the answers are correct as each answer will be vetted by a cybersecurity expert. Cybersecurity has become bloated with a lot of information, and it can be hard to find what you are looking for. Ask Captain Cyber will simplify this process so that even beginners can understand cybersecurity.

IoT devices have become a growing cybersecurity threat. Many more people are installing IoT devices to connect their lights, homes, phones, cars, etc. These are prime targets for hackers and are big cybersecurity threats. From hacking security cameras, spying through your webcam, hijacking a smart car or house, and hacking your smart fridge. Many of these IoT devices can have a lot of vulnerabilities that will be unknown to the average user. IoT devices have limited space and can be hard to update. This leads to an increase in vulnerabilities and poor programming practices. Ask Captain Cyber can help users install, configure, and fully understand the security risks and vulnerabilities that might come with these devices.

Ask Captain Cyber will have cybersecurity experts vetting answers. This means that if a question is not already in the database, it will give a preliminary answer while a professional answers it. This is to ensure that the answers are truly legitimate. Modern-day AI has a problem where it is hard to determine if the information they say is accurate. Users can ask for sources, but there is no actual guarantee the aI is summarizing the information correctly. Ask Captain Cyber fixes this problem by relying on industry experts with years of experience. Ask Captain Cyber will also have a process to re-evaluate already known answers. This is because cybersecurity is an ever-growing and evolving field. So this method makes sure the information is up to date. It is essentially an open source page for tons of cybersecurity information, where we hope there are enough people who are passionate and want to update and help answer users' questions.

The final aspect of Ask Captain Cyber is that it must be easy and versatile. Users with little experience and cybersecurity knowledge should be able to use it and get the necessary answers. People who might ask more advanced questions, like college students, should also be able to use it to get good information and resources. Finally, experts should be able to both contribute and be able to ask more high level questions. In conclusion, Ask Captain Cyber will be a vetted AI platform where users can get cybersecurity questions and answers from a trusted and reliable source.

1.2. Intended Users

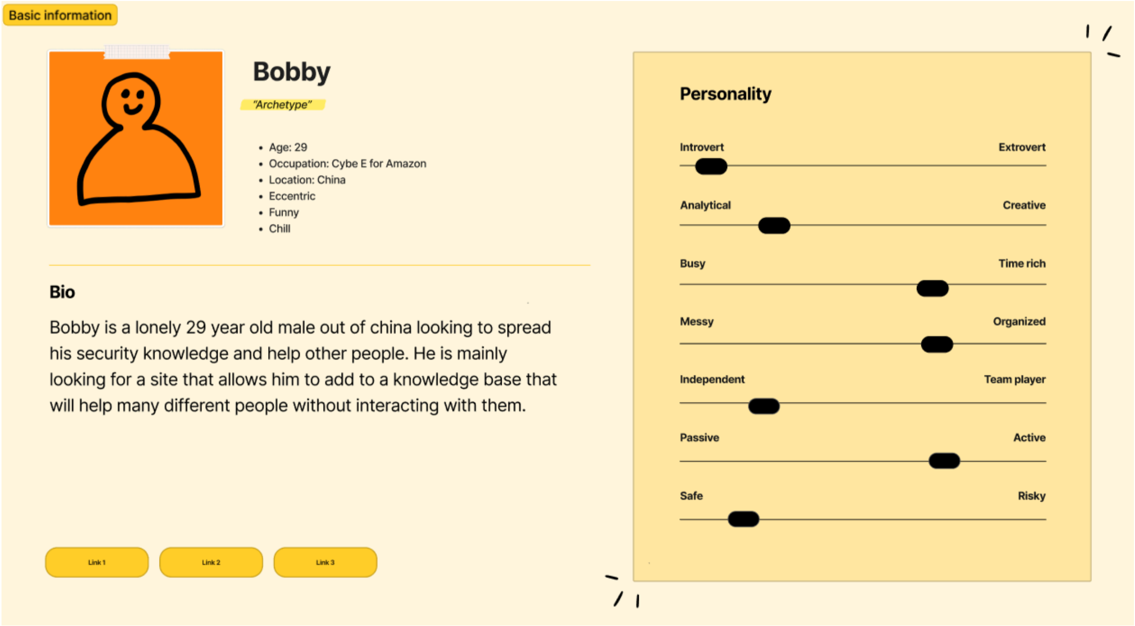
Any person with an internet connection and an interest in cybersecurity should be able to utilize Ask Captain Cyber. We have divided all of these users into three general groups: absolute beginners, cyber enthusiasts with above-average knowledge and interest, and experts with specialized cybersecurity knowledge. The beginners and enthusiasts will benefit from gaining knowledge by submitting questions to Ask Captain Cyber, who should promptly reply with accurate information. Anyone on the internet can access Ask Captain Cyber, considering it is a public-facing website hosted on ISU's servers.

Absolute beginners will have little to zero knowledge regarding anything cybersecurity-related. This would describe the average person with a relatively small interest in cybersecurity. Their need for this project would be to ask very simple questions to Ask Captain Cyber. Ease of use for the website would likely also be their top need for this project. Making this project intuitive and easy to use would encourage the user to ask more questions, whereas having a poorly designed website would likely frustrate the user and turn them away from the website. This product would give them the basic cybersecurity knowledge they desire and can actually utilize in their daily life (i.e., how complex a password should be, what a VPN is, and how they can use one, etc). This directly correlates to the problem statement by providing a vehicle to deliver cybersecurity knowledge to those who do not know much about cybersecurity but want to learn more.

Cybersecurity enthusiasts will have more knowledge regarding cybersecurity than the average person. This would include cybersecurity students, those just entering the industry, etc. Their need for this website would be to ask more detailed cybersecurity questions and obtain more advanced responses. The responses would involve more detailed terms and concepts as well as including resources to scientific/research articles. This group of users would likely need the answers to be very accurate, which is the purpose of the answer vetting done by the experts. Accuracy of more detailed responses is a high priority of the project. This correlates to the problem statement since the project provides knowledge to all cybersecurity users, including those with more intricate questions they may utilize in their studies or even the workspace.

Cybersecurity experts will be the subject matter experts and the highest authority regarding cybersecurity issues. This includes senior professionals in the industry and principal security engineers who serve on Cybersecurity standard committees. Experts will serve as ambassadors for Ask Captain Cyber and vet questions and responses from the tool. Experts will strive to ensure responses are relevant and technically accurate enough to satisfy the enthusiasts and ensure answers are accessible to beginners/average users.

1.3. Empathy Maps



Basic information

Bobby

"Archetype"

- Age: 29
- Occupation: Cybe E for Amazon
- Location: China
- Eccentric
- Funny
- Chill

Bio

Bobby is a lonely 29 year old male out of china looking to spread his security knowledge and help other people. He is mainly looking for a site that allows him to add to a knowledge base that will help many different people without interacting with them.

Personality

Introvert — Extrovert

Analytical — Creative

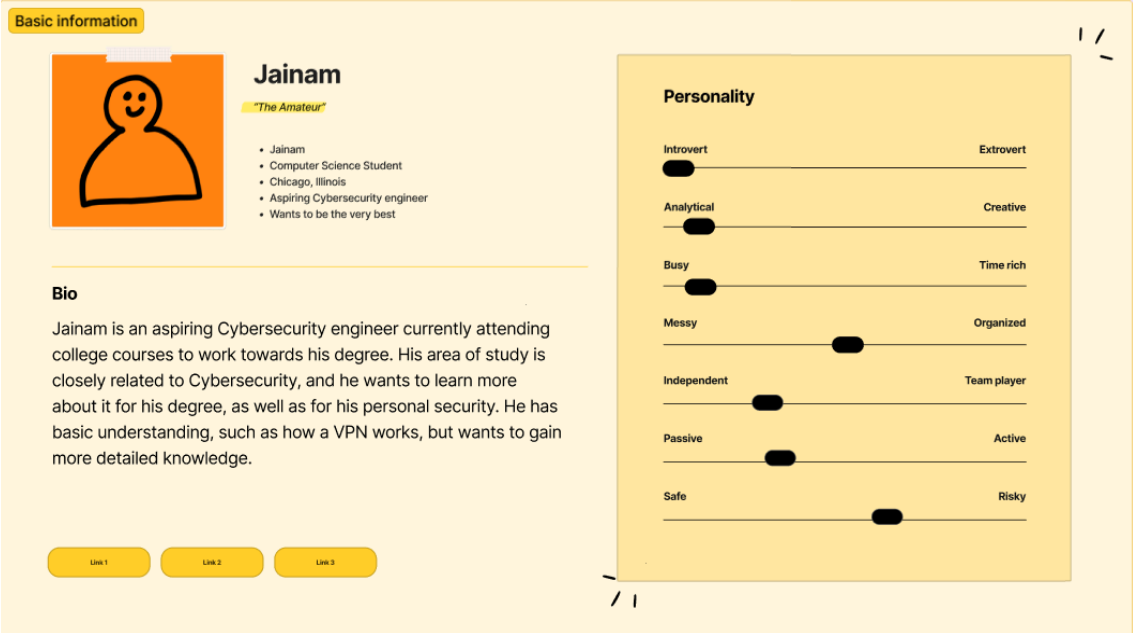
Busy — Time rich

Messy — Organized

Independent — Team player

Passive — Active

Safe — Risky



Basic information

Jainam

"The Amateur"

- Jainam
- Computer Science Student
- Chicago, Illinois
- Aspiring Cybersecurity engineer
- Wants to be the very best

Bio

Jainam is an aspiring Cybersecurity engineer currently attending college courses to work towards his degree. His area of study is closely related to Cybersecurity, and he wants to learn more about it for his degree, as well as for his personal security. He has basic understanding, such as how a VPN works, but wants to gain more detailed knowledge.

Personality

Introvert — Extrovert

Analytical — Creative

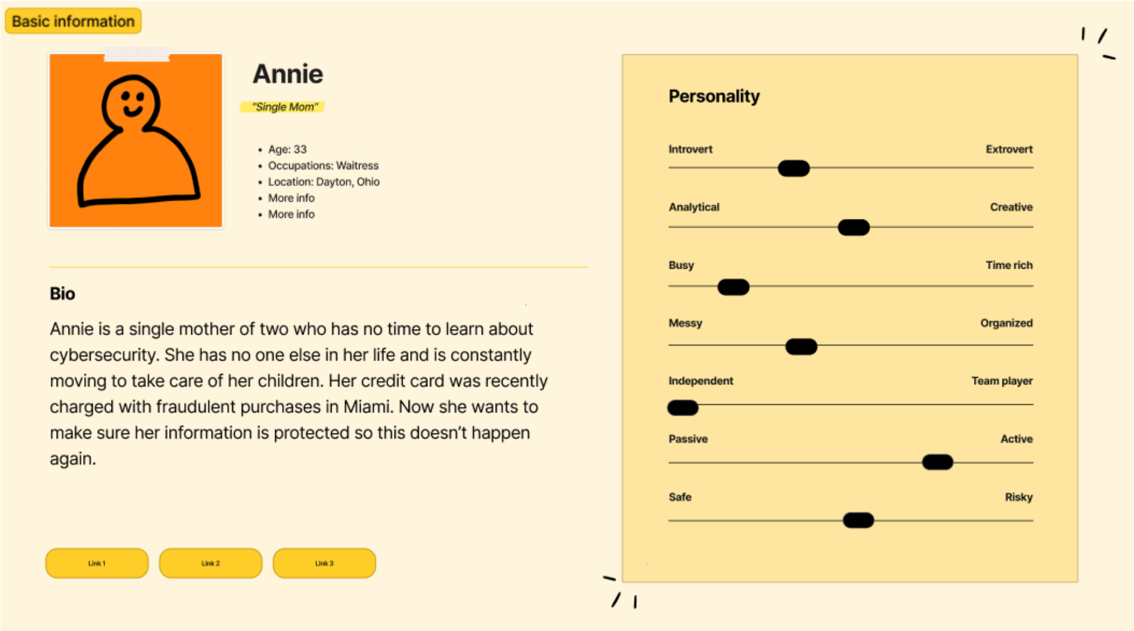
Busy — Time rich

Messy — Organized

Independent — Team player

Passive — Active

Safe — Risky



Basic information

Annie

"Single Mom"

- Age: 33
- Occupations: Waitress
- Location: Dayton, Ohio
- More info
- More info

Bio

Annie is a single mother of two who has no time to learn about cybersecurity. She has no one else in her life and is constantly moving to take care of her children. Her credit card was recently charged with fraudulent purchases in Miami. Now she wants to make sure her information is protected so this doesn't happen again.

Link 1 Link 2 Link 3

Personality

Introvert Extrovert

Analytical Creative

Busy Time rich

Messy Organized

Independent Team player

Passive Active

Safe Risky

2. Requirements, Constraints, And Standards

2.1. Requirements & Constraints

The following list of requirements is designed to comprehensively cover all aspects of our project, ensuring alignment with the values placed on user experience, site security, and ease of implementation.

2.1.1. Functional Requirements

2.1.1.1. Must accurately interpret and respond to user queries. Secure login is also required so experts can vet questions without fear of malicious users having access.

2.1.1.1.1. *Amateurs*: Questions are simple and general and must be responded to similarly.

2.1.1.1.2. *Enthusiasts*: More detailed inquiries that may need diagrams or code snippets to get the point across.

2.1.1.1.3. *Experts*: Need backend access through the login system. Intuitive dashboard to vet questions and collaborate with other experts

2.1.2. Resource Requirements

2.1.2.1. Ensure the vetted answers are stored in the database for future reference, which can be accessed quickly.

2.1.2.2. Ask Captain Cyber will need to be able to reference the information stored in the faq, allowing it to bypass expert vetting.

2.1.3. User Experiential Requirements

2.1.3.1. The interface should be intuitive and take users only a short time to determine how to interact with the Ask Captain Cyber chatbot. The chatbot should also be relatively fast and take just a short time to answer questions.

2.1.4. UI Requirements

2.1.4.1. The color palette of all Ask Captain Cyber-related web pages should match the preexisting colors to promote experience continuity.

2.1.5. Security Requirements

2.1.5.1. Any user information must be appropriately hashed and stored with standards up to date with ISU standards.

2.1.5.2. The database must be up to ISU standards and potentially even higher.

2.1.5.3. User input must be fully or partially sanitized to ensure no malicious or irrelevant input.

2.1.5.4. Ensure ambassador questions are relevant and nonmalicious. Security overall must also not constrain the app too much.

2.1.6. Performance Constraints

2.1.6.1. Responses that do not need to be vetted by experts should take little time to be generated to promote a seamless user experience.

2.2. Engineering Standards

After reviewing the *IEEE Standards in Everyday Life, engineering standards are evident* in our daily lives. These standards ensure that everything we interact with is up to a particular specification that provides the best experience and protects us from things we might not consider. From the moment you wake up, your alarm system and cell phone have been met to the *IEEE 802.15 Family of Standards* that focuses on wireless specialty networks (WSNs) and wireless personal area networks (WPANs) that we don't even realize. These standards take on more responsibility as they protect our homes, networks, and essential home utilities such as our water and electricity meters (*IEEE 1701, 1702, 1704 & P13777* are all used for smart metering). As our lives continue, IEEE standards are present in autonomous vehicles and the factories manufacturing everything we need to make it through our day. Overall, standards are set to protect us and ensure we don't have to question the integrity of our essential interactions with various devices.

Based on the *IEEE Interactive Soccer Stadium*, we can see a more in-depth view of standards used in large-scale operations such as sporting events at large stadiums or venues. Take performance and health tracking, for example. For wearable technology to enhance player safety, it must meet *IEEE P3141 for 3D Processing, IEEE 1708 for Wearable Cuffless Blood Pressure Measuring Devices, and IEEE 11073 Family for Health Informatics.* These standards ensure that coaches and trainers can make smart decisions based on the information they get from watching the game. These three standards are necessary just for a wearable performance and health tracker. Once we start to think about how many devices we interact with daily, we will realize how much work, research, and development has gone into providing the best experience possible.

These standards are necessary to protect and give us the best user experience possible. Without them, there would be consistent errors or outages in vital instruments that keep our society functioning. Thanks to the determined teams at *IEEE,* we have set baselines that must be met for any product or service to make its way to the public environment. Our engineers must rigorously align our projects to these specifications to uphold the consistent reliability and sustainability of all things digital.

It is important to know that we also must ensure our project abides by all IEEE standards relating to the technologies we will utilize. As the product of rigorous research, we have determined the following standards to directly relate to our project:

IEEE Standard for Large Language Model (LLM) Agent Interfaces IEEE P3394

This standard is focused on defining interactions with AI models. It defines protocol methods and formats for communication between the AI and the system. With this in mind, its main goal is to make AI and system integration as smooth as possible.

IEEE Standard for Artificial Intelligence (AI) Model Representation, Compression, Distribution, and Management IEEE 2941

This standard focuses on the compression and distribution of AI models. It provides guidelines on storage and management. These guidelines and focuses help make AI models effective and interchangeable across different hardware and software environments.

IEEE Standard for Password-Based Public-Key Cryptographic Techniques IEEE 1363.2-2008

This standard focuses on techniques for cryptographic protocols implementing public-private key encryption. Its goal is to make this type of encryption more secure and easier to integrate when needed.

After reviewing each standard, we have determined they are directly applicable to our project in multiple ways:

IEEE Standard for Large Language Model (LLM) Agent Interfaces IEEE P3394 fits firstly because we are incorporating an AI into Ask Captain Cyber so we need to make sure our system is set up to work smoothly with the AI. Secondly, we need to ensure the AI-incorporated system can send and receive data with our chosen format.

IEEE Standard for Artificial Intelligence (AI) Model Representation, Compression, Distribution, and Management IEEE 2941 is also associated with our project since it provides guidelines for managing the AI technology we will implement. It also discusses the API framework for large-scale pre-trained AI models, which we will use in our project. By aligning ourselves with the goals of this standard, we can help promote operational efficiency and proper usage of Ask Captain Cyber.

IEEE Standard for Password-Based Public-Key Cryptographic Techniques IEEE 1363.2-2008 fits because we intend to use public and private key pairs for our ambassadors to log in and respond to asked questions. This standard will help with its integration into our system. Secondly, it will help ensure that bad actors preying on the site can't steal our data.

Some other standards found by the team are IEEE 7000-2021: AI Ethical System Design and IEEE 23026-2023: International Standard - Systems and Software Engineering – Engineering and Management of Websites for Systems, Software, and Services Information. We believe these standards can also align with Ask Captain Cyber quite well. However, to access the IEEE 23026-2023 standard, we would have to purchase it to really dive deep into the specificities, but we can potentially incorporate it as we see fit. For the IEEE 7000-2021 standard, we came to the conclusion that we will be implementing this with the other AI-specific standards we have listed above.

To meet the specifications required by these standards, we will have to modify our approach to be incredibly detailed. For example, we will have to construct our chat responses and interactions to meet *IEEE P3394* so that the system as a whole runs smoothly without failure. As the focus of Ask Captain Cyber revolves around providing expertly vetted answers to cyber-security questions, one of the most important things to realize is that we have to keep our backend secure so that no one with malicious intent can distribute incorrect information. To ensure this happens, we have to abide by *IEEE 1363.2-2008* so that there is no risk of our expert's passwords being leaked or guessed. As we are still in the early stages of development, we won't have to modify any pre-existing systems we have already built; instead, we must work and design with the intent to fulfill all of these standards' needs so that we don't have to regress further down the road. If we can act proactively compared to reactively, we can be sure that *Ask Captain Cyber* is up to specification instead of reconfiguring it to meet the standard.

3 Project Plan

3.1 Project Management/Tracking Procedures

Our project management style will be a combination of agile and waterfall. Our style will be very similar to agile in that we prioritize flexibility, constant collaboration with each other, and breaking down our project into iterative tasks that will be completed step by step. We also need to be able to change our project if any changes are required by our advisor, adding to the flexibility of the project. Our style is different because we cannot meet as consistently as we should for a proper agile management style since we only meet once a week with each other and monthly/as needed with our advisor.

Our project utilizes Git as our code repository, so we have decided it would be easiest to use Git for our project management board. We have five columns: Backlog, To Do, In Progress, Stalled, and Closed. Our backlog is for our tasks that need to be eventually worked on but are not an immediate priority. The to-do column is for tasks that must be worked on shortly to continue our project effectively. The in-progress column is for tasks currently being worked on. Stalled is for tasks waiting on another task/team member to continue working on it. Finally, the completed column is for our completed tasks. As a team, we will update this board consistently to reflect our progress on the tasks we are working on and reflect our overall progress on the project.

3.2 Task Decomposition

For the backend, multiple tasks and subtasks are needed.

First, we must set up a basic database to store user/admin data. To do this, we need to figure out how the WP database works, if any plugins will be necessary, and how queries are done internally. We also need to set up another database for the LLM that the AI can use to pull answers from, and it needs to have interconnectivity with other aspects of the website. The first basic database will then need to be connected to the front-end login, and make sure that the ambassadors have the correct permissions to do what they need to do. For the LLM, we will need to figure out how much data to store and how the app will access it, ensuring only the right people have access.

The backend will also have to handle the authentication and validation of ambassadors trying to access user questions. Potentially using MFA.

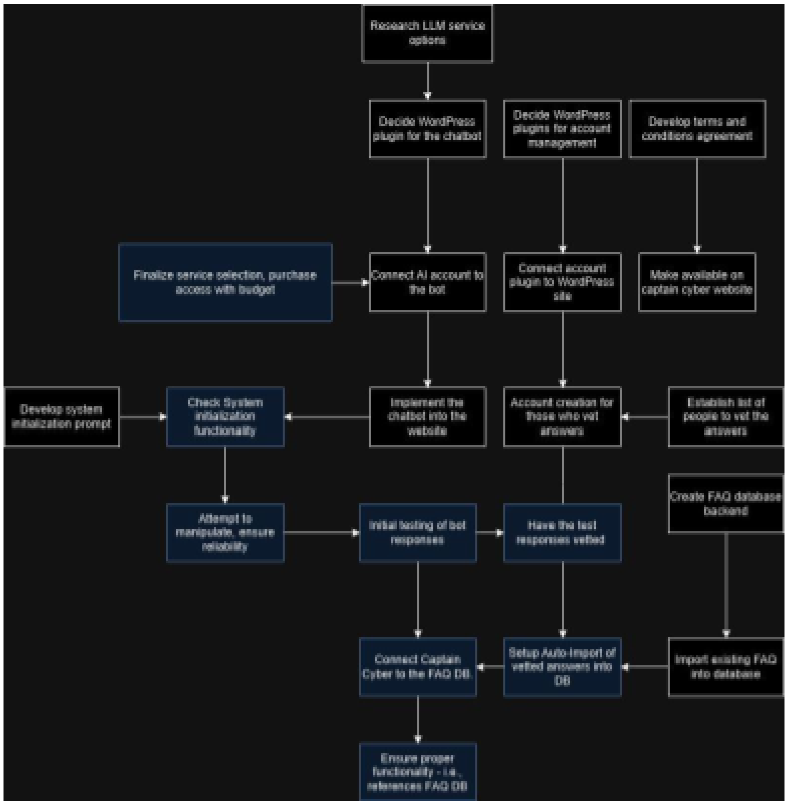

Research LLM service options
Decide WordPress plugin for the chatbot
Decide WordPress plugins for account management
Develop terms and conditions agreement
Finalize service selection, purchase access with budget
Connect AI account to the bot
Connect account plugin to WordPress site
Make available on captain cyber website
Develop system initialization prompt
Check System initialization functionality
Implement the chatbot into the website
Account creation for those who vet answers
Establish list of people to vet the answers
Create FAQ database backend
Attempt to manipulate, ensure reliability
Initial testing of bot responses
Have the test responses vetted
Connect Captain Cyber to the FAQ DB.
Setup Auto-Import of vetted answers into DB
Import existing FAQ into database
Ensure proper functionality - i.e., references FAQ DB

Figure 1: Task Decomposition Chart

The next step the backend has to worry about is the AI and how the AI API will be integrated with the UI. The AI will need to take user input -> go to either the LLM->If that does not work, go to the internet-> then deliver the response to the user within a responsible amount of time. This will require good networking code and fast algorithms. We will also need to think about parallel processing. If multiple users are using the bot, it needs a way to know which user to deliver the answer to and still run fast. More research will need to be done to determine whether this will be a plugin or a feature built-in JavaScript. The backend will also have to identify when it is appropriate to store user questions. This could be done manually, or we could make an automated process to store user questions. The backend will also need some way to update the information to ensure it is up to date, but this would be a future feature.

For the front end of the website, we will be able to build the screens and then implement the backend calls through the WP React API. This will allow us to make Rest API calls to the WordPress Database from a React frontend. With this, we have varying tasks that must be completed.

First, once a user enters the chat view, we must establish a secure and stable connection to make the conversation quick and seamless. Since we will rely on ready-made answers that have been vetted already, we can quickly query pre-formatted answers. If the question has yet to be answered we will then pivot to AI answers and make it known they have yet to be vetted by an expert and offer a notification once it has been. As this happens, we will populate the chat screen to emulate the ideal experience for the user.

We will also have to build an expert dashboard where they can view, collaborate, and vet answers. This will require a secure login authorization service. Ideally, we can use a pre-made service such as the ISU login service with Microsoft Authenticator. In this dashboard, we will make an intuitive screen that lets expert ambassadors easily interact with the list of questions.

3.3 Project Proposed Milestones, Metrics, and Evaluation Criteria

There are various milestones of this project, largely consisting of:

- Webpage implementation completion - All Ask Captain Cyber web pages are up and running, with complete functionality and React implementation.
- LLM FAQ References - The LLM has proven that it can pull accurate information from the FAQ and recognizes when a prompt is unrelated to any FAQ content.
- Expert Vetting Process - The LLM-generated responses can properly flow through the expert vetting stages and make their way to the user.
- AI is fully functional with stress tests with multiple users.

Our evaluation criteria will consist of 3 aspects.

- Intuitive UI implementation - This will be evaluated via a Usability test of our UI.
- LLM FAQ References - The LLM can reference and learn from the dynamic FAQ.
- Generation Accuracy - The LLM abides by its initialization prompt restrictions and does not stray from discussing solely cybersecurity-related prompts.

The metrics used to evaluate our UI will come from a usability study for the criteria above. We will have various users rate different aspects of the UI out of 10, with this milestone being reached when the cumulative UI evaluation reaches a rating of 7/10.

For the LLM FAQ references, we require that Ask Captain Cyber scans the database for relevant information upon every prompt. We will consider this milestone achieved when the LLM generates 80% of FAQ-related questions and does not have to be vetted by experts.

The metric used to evaluate the generation accuracy is that the LLM generates relevant information 80% of the time. This will be determined by human consideration of the relation between the user prompt and the generated response. This will also encapsulate the experts' responses to ensure they provide relevant information to the user.

3.4 Project Timeline/Schedule

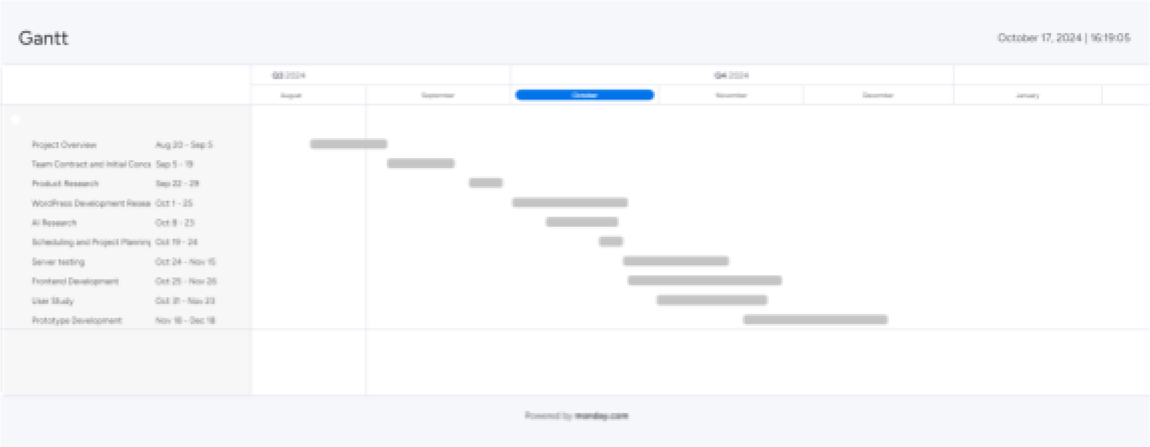


Figure 2: Gantt Chart

The current phase of development that we just finished was product research. This is where we investigated what tools we would use, how we would structure this project, and general design work. We finished this task, and we are currently working on server testing. This testing ensures the server is up and running, and we can actively connect and start working on development. We need to ensure everyone has a local instance of WP on their device to start development. We also need to see what features the website already has.

We are also working on the initial prototype frontend and backend, which will go through November and December. Front end development will involve Login/Signup, Chatbot page, FAQ, and a backend vetting page for admins. There also needs to be some initial work done to figure out how to connect to the backend.

Backend work is being done to set up the user and LLM database, integrate API calls, and provide the general security needed. Integrating the LLM, vetting questions, and the front end will be the most challenging part of the project. This might take longer than usual, but we want to finish some rough initial work. We hope to get a rough setup/plan by the end of November.

Finally we will spend the rest of the semester from November 18th working on getting a final prototype working. This includes all pages being up, AI/LLM partially working (answering questions), relatively fast speeds, and good security practices in place that are up to ISU standards. Once Git is appropriately set, we will assign issues for team members to work on.

3.5 Risks and Risk Management/Mitigation

Project Overview

Risk: Poorly stated project requirements will lead to poorly configured systems down the line.

Probability: 0.4

Severity: Moderate

Mitigation: Keep open contact with how individual portions of the project are going. Keeping everyone on the same page will prevent any communication-based misconfigurations.

Team Contract

Risk: A team contract that doesn't cover many possibilities will lead to teammates abusing loopholes.

Probability: 0.3

Severity: Low

Mitigation: Make clear rules that can't be misused. Be professional.

Product Research

Risk: Researching can consume a large portion of time.

Probability: 0.5

Severity: High

Mitigation: Set time limits and make sure people stay on target with what research is needed and what shouldn't be researched.

WordPress Development

Risk: Unfixable bugs could lead to complete code scraps.

Probability: 0.6

Severity: High

Mitigation: Debug as you code, separating parts, allowing coders to focus and find bugs quicker.

AI Research

Risk: Filling the database might require too much work for the total time we can spend on the project.

Probability: 0.7

Severity: High

Mitigation: Only fill the database with enough data to test the project.

Server Testing

Risk: Achieving high performance on the server will take a significant amount of time.

Probability: 0.5

Severity: Moderate

Mitigation: Make sure we have enough performance for the essential operation of the project. Speed is out of scope.

Frontend Development

Risk: Some user interaction features may not consistently work

Probability: 0.2

Severity: Moderate

Mitigation: We will rigorously stress test all user interaction features alongside the usability test

User Study

Risk: Low user participation will lead to poor test case coverage.

Probability: 0.4

Severity: Moderate

Mitigation: Work with the student body and client to provide reasons for users to use the platform to increase traffic.

Prototype Development

Risk: We won't have enough testers, or it will go poorly

Probability: 0.4

Severity: Moderate

Mitigation: Work with Doug to ensure we have enough testers and stay on schedule for our tasks.

3.6 Personnel Effort Requirements

Table with the amount of hours contributed by each person on each task.

Task->	Project Overview	Team contract	Product research	WordPress Development	AI Research	Server Testing	Frontend Development	User Study	Prototype Development
Casper	6	3	8	6	4	5	5	4	3
Ethan	7	3	5	6	7	5	4	4	3
Steven	5	3	7	6	4	5	5	5	4
Alex E.	6	5	5	5	5	5	6	4	3
Alek K.	5	4	6	6	6	4	5	4	4
Caden	6	4	6	5	6	4	4	5	4

3.7 Other Resource Requirements

As our project is a website hosted on Iowa State servers, we won't need any physical parts or materials. We will rely on the expertise of our team to get the project where it needs to be. We have worked tirelessly to learn about proper backend and frontend practices. These resources will pay off as we continue developing to ensure Ask Captain Cyber works as intended. We will use several resources, such as libraries and API, to bring Ask Captain Cyber to its full capability.

The main thing that we will need to request is the licensing to the Co-Pilot. This will require a payment to receive our keys when answering questions. We also will be utilizing WordPress to host which will require constant power and connection on the ISU servers, enabling our backend engineers to develop plugins to handle all of the events and queries from Ask Captain Cyber. On the front end, we will require libraries to style the website intuitively. We will likely use a third-party multifactor authentication system to log in or develop our own if time and resources allow. Overall, our required resources, parts, and materials are relatively low. We must use our education and expertise to use our resources to their full potential.

The final resource is that Doug and the ISU department will need to gather cyber ambassadors to test the website and vet answers. This is more of an issue that regards our advisor and is out of scope for our team.

4 Design

4.1 Design Context

4.1.1 Broader Context

The following table lists broad contexts in which our design problem is situated:

Aera	Description	Examples
Public health, safety, and welfare	Ask Captain Cyber will help vulnerable users who are not as technically literate as others. This is important in an ever growing hostile cyberspace, where cybercrime is on the rise. Cybercrime often goes beyond monetary loss, which is why it is important for the general population to be knowledgeable about these subjects.	An older person who is more susceptible to scams can use Ask Captain Cyber to avoid potential cybercrimes and scams they might have fallen for. A child who does not have much awareness about the criminals on the internet may fall for cybercrimes. They can use Ask Captain Cyber for advice.
Global, cultural, and social	The cybersecurity community wants people to be more secure and avoid monetary and data loss. Ask Captain Cyber will help to advance this goal by educating people who are not as technologically knowledgeable about these topics.	If an employee works for a company in a critical sector, who is less knowledgeable about cyber threats might pose a risk to many more people. With Ask Captain Cyber, they can become more knowledgeable and avoid making a cyber mistake.
Environmental	Our project will operate on existing hardware from Iowa State University and OpenAI's LLM servers, which should limit any contributions to emissions from our project.	Hosting on Iowa State servers with existing infrastructure and utilizing cloud solutions reduces the need for additional hardware.
Economic	Our product is free for all users who have access to the internet. Iowa State has already purchased the servers the website will be hosted on and will pay for any costs related to hosting this project.	Depending on how large the LLM gets, we might need a lot of storage which could cost more money, but not a significant amount. This project will not economically hurt or benefit any specific sector. The only potential concern is if a user asks about a specific product and our AI convinces them to buy it.

4.1.2 Prior Work/Solutions

One similar solution is called the Departmental Intelligent Neural Advisor (DINA) [1] created by a professor at the University of Iowa. DINA is a chatbot designed to specifically discuss coursework at the University of Iowa with students. Our group talked with Tyler Bell, the primary developer of DINA, and obtained a lot of useful information we are now able to apply to our project. DINA uses ChatGPT's Assistant API [2] as the backend which answers all the questions. Prompt engineering is also used to configure the Assistant to act as a course counselor and only answers questions related to that. This knowledge is extremely helpful towards our project because we can prompt engineer ChatGPT's Assistant API to perform as a cybersecurity chatbot that refers to itself as Ask Captain Cyber.

ChatGPT and other AI solutions are very similar to our product and we will be using their APIs for our project. The main difference is that ChatGPT does not have expert, verified answers, and it does not allow you to access the questions people ask. Our project will allow experts to see answers and responses so that they can either answer it or verify its authenticity.

Pros:

- Free to use
- Verified by experts in the field
- Maintained by passionate people in the cybersecurity industry
- Focused on educating people who are less technology knowledgeable.

Cons:

- Less flexibility compared to using the internet or other AI platforms
- Brings in no income or revenue streams, so its a net costs
- Has a significantly smaller LLM to work off of compared to other tech giants.
- Has to compete with incredibly large tech companies with more time, money, and resources to their disposal.

Another similar website is called Stack Overflow, a technical Q&A platform designed to help students and developers learn and share technical information worldwide [3]. This website is similar in that anyone can post a question and an expert, or someone knowledgeable in that field, can answer it for them. This is similar to our project requirement that even though the AI will provide a response to the user's question, it still must first be verified by a person to ensure accuracy. However, with Ask Captain Cyber not just anyone can answer the question. It must be someone who has been approved as a cyber ambassador through the cybersecurity outreach program here at Iowa State [4].

4.1.3 Technical Complexity

Ask Captain Cyber to utilize: PHP, Wordpress, React, AI development, database management, and AI integration with Wordpress. There are a lot of moving and interconnected parts to this project. We also need to create a system where ambassadors can view user prompts, answer them, and have the answer be saved. We could be dealing with hundreds or thousands of questions so we need to have a system to organize it and prioritize questions that may be asked multiple times.

There are numerous challenging requirements that must be addressed to match industry standards and existing projects. The first is dealing with real-time AI responses and updating the system quickly to try and get the response back to the user as quickly as possible to ensure a good user experience. Another requirement we must address is secure user authentication for our experts. Considering we have a login system for our approved cyber ambassadors, we must ensure their credentials stay secure and integrate public-key cryptography to meet Iowa State and IEEE standards. Also, we must ensure Ask Captain Cyber is implemented ethically and can only output ethical answers to users to prevent the spread of misinformation, bias, or even malicious information. Since Ask Captain Cyber only answers cybersecurity-based questions, it is very important that it responds only with ethical information so it does not negatively impact anyone, which adds another layer of complexity to the project.

4.2 Design Exploration

4.2.1 Design Decisions

Due to the scale and details behind this project, there have been many important decisions for developing Ask Captain Cyber to ensure the highest response accuracy, query response, and user experience.

The most important decisions have been deciding which AI model to use—We have decided to use OpenAI's ChatGPT-3.5. We chose this model because it offered important features such as access to real-time data and web searches and the best balance of cost and LLM capabilities. These aspects are important because they enable Ask Captain Cyber to have access to live information and minimize AI generation response times.

Deciding which frontend design language to use - We have decided to use React to implement our frontend design for Ask Captain Cyber. We chose this since it seamlessly integrates with WordPress, and some of our group members have previous experience using it. This helps us by promoting ease of implementation, allowing us to design the front end to cater to a good user experience.

Deciding how to categorize questions - One aspect of our project we still have to decide upon is how we will categorize user questions. We will need to ensure that our backend implementation can associate similar questions with each other so that if it is similar, a response is quickly generated by the LLM rather than waiting to be vetted by an expert. This will help decrease response times for user prompts, a critical aspect of the user experience. Another important feature will be to make sure that the vetted questions can be sorted and organized in a clean and efficient way. We do not just want a random assortment of questions with no rhyme or reason. We can also implement some sort of priority if a question or similar questions are being asked a lot and can be answered by the ambassadors.

4.2.2 Ideation

The bulk of the ideation process for the project lies in deciding which LLM to use. There were many critical aspects we needed to consider such as the number of parameters, if it has access to real-time information, cost, and its response time. Ultimately, we decided that OpenAI's GPT-3.5 solution would work the best for our use case and evaluated other models such as Gemini by Google, Llama by Meta, Claude by Anthropic, and GPT-4.0 by OpenAI. We identified these potential options by researching the most popular LLMs and evaluating our previous experiences with each one. The other part was how to design user interaction along with how the front and backend will interact with each other. You can see in the documents that we provided that this is an important feature that we did before we even started to develop the product. We are also continuously coming up with ideas and implementations as we gain more information from our advisor.

4.2.3 Decision-Making and Trade-Off

To make this decision, we developed a weighted decision matrix to compare and contrast each LLM and gain a better understanding of their strengths and weaknesses.

Model	Provider	Context	Pricing	Total
GPT-3.5	OpenAI	6	7	**13**
Gemini 1.5 Pro	Google	6	6	12
Llama 3	Meta	3	8	11
Claude 3 Opus	Anthropic	8	1	9
GPT-4.0	OpenAI	7	3	10

This decision matrix evaluated two different critical components of our LLM to provide context and pricing. Pricing related to the cost of operation of our Ask Captain Cyber project, with context defining the amount of text data it can consider simultaneously. By examining each criterion, we determined that GPT-3.5 would be the best option for our use case, offering a relatively cheap price while still maintaining its capability to process large amounts of text.

4.3 Proposed Design

4.3.1 Overview

Ask Captain Cyber is a cyber-security-focused chatbot that will be able to answer users' questions relating to cybersecurity, with support for all levels of complexity. Questions that have not been answered will be generated by an LLM and vetted by experts to ensure accuracy. Our current high-level design is illustrated in Figure 3:

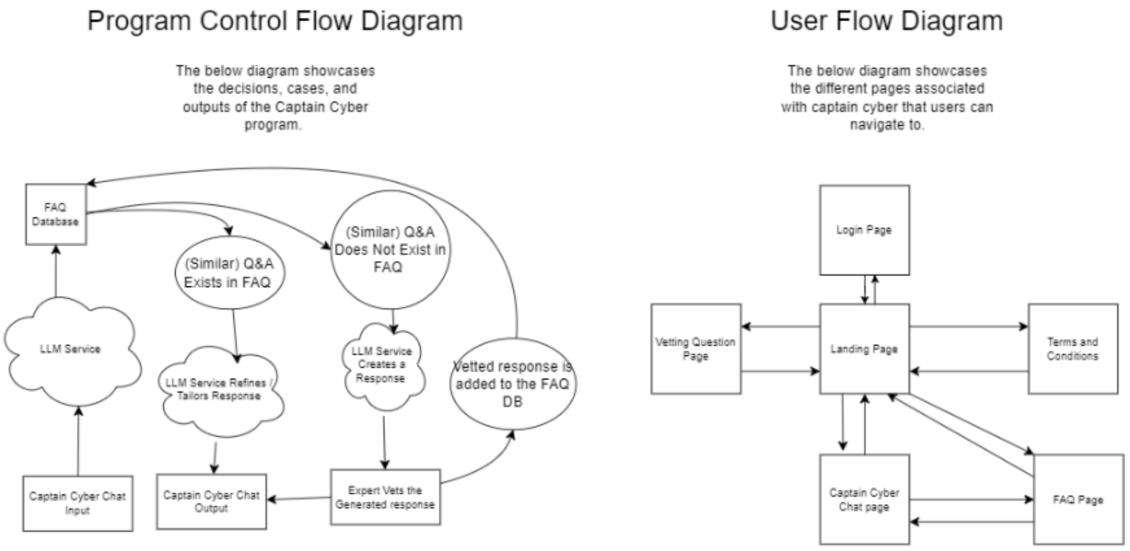

Program Control Flow Diagram
The below diagram showcases the decisions, cases, and outputs of the Captain Cyber program.
FAQ Database
(Similar) Q&A Exists in FAQ
(Similar) Q&A Does Not Exist in FAQ
LLM Service
LLM Service Refines / Tailors Response
LLM Service Creates a Response
Vetted response is added to the FAQ DB
Captain Cyber Chat Input
Captain Cyber Chat Output
Expert Vets the Generated response
User Flow Diagram
The below diagram showcases the different pages associated with captain cyber that users can navigate to.
Login Page
Vetting Question Page
Landing Page
Terms and Conditions
Captain Cyber Chat page
FAQ Page

Figure 3 - "Ask Captain Cyber" Architecture

In the diagram depicted in Figure 1, we have the program control flow and the user interface flow. The program control flow diagram outlines the backend implementation of our project, where prompts are checked to see if an answer exists in our FAQ, and if they don't, they are AI-generated and then vetted by experts. The user flow diagram shows how the frontend implementation of our project will work, with six total screens allowing users to log in, vet questions, read the terms and conditions, view the FAQ page, query Ask Captain Cyber with their prompt, and a landing page explaining what the tool does. This high-level view of the front and back end of our project is representative of our ongoing implementation. The various subsystems, such as the FAQ and generation, are critical to this process as well, ensuring a good user experience for our final version.

4.3.2 Detailed Design and Visual(s)

For the backend of the design, we will have three or so databases. One will be for users. This user database will authenticate users, and entries must be manually entered. Each user will actually be a user for each organization. For example, School A will have a School A user, and School B will have a School B user, etc. This must be managed and queried when the user tries to log in. The backend will use email authentication as a MFA for an extra layer of security. Regular users will have no login page, which will only be for the ambassadors.

The frontend aspect of Ask Captain Cyber is rather simple. Since most user interaction will happen within the chatbot environment, we will design it so that the information is clearly presented with little to no other distractions. This will allow the user to focus on learning more about cybersecurity and less about how to explicitly interact with the site. Furthermore, when they land on the website, there will be a short summary of the project and a disclaimer so that they are aware of potential pitfalls or consequences of malicious use of Ask Captain Cyber. The experts will have an intuitive dashboard to view questions waiting to be vetted, questions that have recently been vetted, and a note service to collaborate with other expert ambassadors as needed.

We will also need a question database where questions that the AI can't answer will get forwarded to a database that ambassadors can query and answer. This database will also have to be somewhat sorted or displayed in a sorted manner. This is because we do not want questions to be completely random when the ambassadors try to answer them. So, we need to create some sort of mechanism to organize or sort the questions in the database itself, or when it is displayed. This database must also be scalable because we do not know how many questions could be simultaneously. This database of questions might need to be encrypted. This is because if a user types in a question that includes sensitive information, we do not want that question sitting in plaintext. We also need to make sure only ambassadors can query the database. This might require some token or extra layer of authentication on top of the authentication required to log in. We must still determine how this process will work to balance security and usability. Finally, this database will have to be able to update questions, then give it back to the user and then send it to the LLM for future reference.

The final database will be the LLM. This will be the dataset the AI uses when querying questions. This database will be ever-growing and must be scalable for a large amount of data. The AI will also have to be able to check if questions are similar and alter pre-existing questions if they are similar. It will then have to update the LLM with new answers and questions and be able to sort them effectively. Questions will probably have to be sorted by type, such as networking, password, security, etc.... It will also have to connect with the question database that ambassadors will use.

An outline for the system design can be found below in Figure 4:

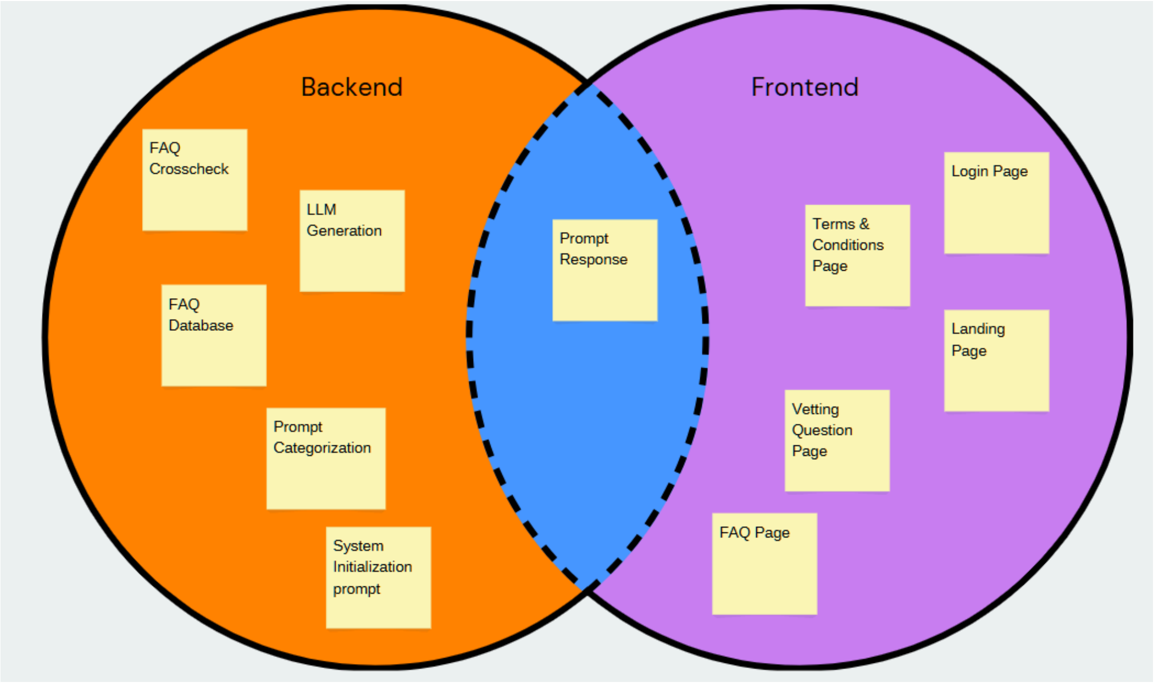


Figure 4 - Ask Captain Cyber Front & Backend implementation

4.3.3 Functionality

The design we are going with will wait for a user to "Ask Captain Cyber" a question. It will then go and check the current database of questions to see if an answer matches the question. If it does, it will just return that; otherwise, it will go and generate a response to the question and have a cyber ambassador check it, giving the user a basic "come back later" response. Once the question is answered, the person will either get an email telling them there is an answer, or the user can come back any time to ask since the question will now be in the database.

4.3.4 Areas of Concern and Development

Currently, our design and design process is focused on user accessibility. As we progress through the development of Ask Captain Cyber, if we keep a user-centric approach, we will ensure that user needs will be taken care of without an extra effort to implement them later. As stated throughout our design documentation, ensuring that Ask Captain Cyber is intuitive and easy for customers to interact with is a key feature. This requires our site to be useful to a wide range of experts. Keeping this in mind, we must establish an efficient frontend-backend communication protocol and an accurate response system that does not leave the user with more questions.

This does raise some concerns that we must take into account to deliver a sustainable and efficient product. We are required to host Ask Captain Cyber on WordPress. In order for the user interface to be customizable and dynamic, we will have to develop workarounds that would be necessary for other databases. We also must make sure the responses are pulled from the already vetted answers before relying on the AI implementation so that we know what information is being provided to the user. If there is no readily available answer, our site must pivot to a potential notification system or instead utilize the AI's API to provide an answer despite it not being vetted by an expert. After our previous meeting with Dr. Jacobson, we gained a deeper understanding of how to design the expert login. We must provide one login for each chapter of ambassadors that they can use to vet the answers. This could make it easier than we initially thought, as there won't be as many experts to adhere to. Finally, we must provide a list of related questions for each question a user asks. This will require an efficient and accurate system to find related topics to each question.

To address these concerns within our solution, we will have to focus on both user needs and requirements provided by our advisor. We will continue to develop a frontend that is intuitive and easy to use that leaves no room for confusion when interacting. Our backend and security team will also build efficient plugins that adhere to our goal of a secure and impenetrable backend that can be efficiently searched to provide accurate and detailed answers to many potential questions. Since we just recently met with our advisor, we had a couple of questions to ask. These ranged from what kind of login system he wants, whether or not a multifactor authentication is necessary, and what disclaimers must be shown to users. As we implement the answers to our questions, more questions will surely arise; however, we have built a positive relationship with Dr. Jacobson that will allow us to be open and honest with the roadblocks we face.

4.4 Technology Considerations

One of the main technologies we are using is ChatGPT 3.5. One of the main strengths of using this is that it has a large support backing compared to other AIs. But it does have a weakness that we will have to work around, and this weakness is being confidently wrong. One of the ways we are getting around this is by using ambassadors and our own data set. Another technology consideration we are looking at is WordPress. The main advantage to using this is it would help provide a start to the website and has a lot of plugins that we can use. One disadvantage to using WordPress is that our project might require things that WordPress doesn't have plugins for. We can overcome this by simply making our plugins. WordPress also has a lot of built-in features, such as databases, which makes it a lot easier to develop. WordPress also has built-in react programming, which makes it easy to add react code for the front and backend.

4.5 Design Analysis

So far, we have tested a few different AI models available online and conducted a plethora of research to ensure a seamless implementation of our Ask Captain Cyber application. We have also begun building the frontend for multiple web pages used in the project, alongside implementing a few backend plugins. So far, our proposed design has been working well and should represent the final project, given our current progression. One issue was setting up our local environments, but this issue has largely been resolved. For our future design, we may change which plugins we use in case any conflicts arise during our development. Overall, our design is feasible and on track for a successful launch by the anticipated deadline.

5 Testing

Testing will be crucial to ensure Ask Captain Cyber's functionality, reliability, and security. Given that the nature of this project is an AI-powered chatbot hosted on WordPress that offers cybersecurity-related questions to users of various expertise that experts have vetted through their dashboards, our testing plan will emphasize data integrity, user experience, and system security. This platform must deliver accurate responses to user queries, maintain a seamless user experience, and protect the data that the users will be given.

This proposes a unique set of challenges that our system design must consider. Even though our answers will be vetted by the cybersecurity experts, we still must ensure that the AI is giving accurate answers that require little to no expert manipulation. We want them to focus on working through the queue of questions and answers and less on meticulously scouring the AI-given answers and adding context or fixing issues. This will require accurate, prompt engineering. Since we are catering to users with a wide variety of understanding in the world of cybersecurity, we will have to do user role testing by simulating the different types of questions that will be asked. This will ensure that the answers are not only accurate to the question but use the best wording to explain the topic. To make the vetting process as efficient as possible, the expert dashboard must evolve dynamically, consistently updating in real time with new questions that are coming in. We must account for these updates so there is no disruption in functionality.

5.1 Unit Testing

The tools that will be tested are the chatbot responses, expert user authentication, and expert dashboard functionality. While we will have other components in the website, such as an FAQ and about page, these units will be static and not be constantly changed or need updating. In order to efficiently test these units, tools such as Jest for the frontend and PHPUnit for the WordPress backend will be suitable for our testing strategy.

5.2 Interface Testing

Since Ask Captain Cyber is a website, the interface being used is simply a computer or phone with a wifi connection. Similar to other chatbots, users will simply type in their questions and then wait for their response. This interface will be manually tested, ensuring that it is intuitive and responsive to various screens or devices. We will also test that each menu, link, button, and input field acts accordingly. Furthermore, we will manually test that errors are consistently and accurately displayed when something such as an incorrect password is entered or if a non-vetted question is attempting to be published to a user. Our interfaces are simple to the user level and can be manually tested without having to configure specific tools to act on them.

5.3 Integration Testing

Some critical integrations mainly include WP and the LLM we plan to use. We are going to try and use OpenAI API as a base for our AI, then tailor it to what we want. Integrating this with Wordpress and a React frontend will be the main critical integration path. This can be tested using Postman at first or the OpenAI API backend which has built in testing. We will then have to set up a user prompt front end, and make sure it's able to send and receive the answers given from the API. Finally we will need to integrate a logging system to store the answers/questions for use. This will also be critical and challenging because wordpress does not have a good way to implement this.

5.4 System Testing

Our system level testing strategy consists of multiple layers, encompassing integration and fullstack, amongst other aspects. Our method of conducting interface testing is by testing frontend interactions, such as demonstrated in the appendix of our work in progress front end build. Verifying that UI elements such as buttons, forms, and all pages correctly work and/or navigate to the corresponding page is critical in ensuring that our project is fully functional and reliable. Our unit testing is done by testing individual components, and especially edge cases. Testing these inputs helps make sure they properly interact with the rest of the system and do not manifest any unexpected behavior. Our primary tool for executing these tests is with LocalWP, where the in-development solution can be tested on a local environment, not affecting any in-production products.

5.5 Regression Testing

As development continues we are also ensuring that new additions do not break old functionalities by locally testing all of our changes. We also test these changes using Local WP across multiple machines to ensure it is not only subjectives working on a single device. One of the most critical features we need to ensure we do not break is the existing Cyber House Rock website made by Iowa State University. This is critical since our project is an adjacent addition to this, and we must make sure what already exists remains functional. We also must make sure that the “Ask Captain Cyber” chatbot only discusses relevant information it is allowed to speak on (e.g. cybersecurity concepts). This is largely handled by the system initialization prompt, which also addresses any ethical concerns about AI-generated content. We will continue to utilize these testing methodologies and add new ones to our testing process as necessary. As a whole, it is largely driven by requirements for the project, and the importance of delivering a capable product that abides by all IEEE and ethical requirements.

5.6 Acceptance Testing

Acceptance testing is and will remain a critical component of our testing process, ensuring that our solution meets all expectations set by our client. We have and will continue to routinely meet with our clients to remain aligned with their objectives. We will continue to demonstrate to our client what work has been done, and continue to communicate what our next steps and plan of action is. We will also utilize those experienced with cyber security on our team to make sure our solution is secure, as part of the non-functional requirements. However, all requirements will be demonstrated as met by outlining them individually, and showcasing how each presents themselves within our project. By demonstrating them one by one we can individually outline how they work, and note if any require further modification.

5.7 Security Testing

Security testing has not been one of our primary areas of concern thus far. However, we are keeping this in mind as development progresses. This far we have curated a rigorous system initialization prompt that the LLM will read upon every startup, explaining its purpose, abilities, and what it should and should not do. This also helps ensure that our solution will remain within legal and ethics guidelines. We are also making sure that all technologies we are using do not have security vulnerabilities through rigorous research and personal testing. As the project progresses we will ensure these policies are upheld and conduct further personal testing. We will also have to put our security testing up to standards with ISU, as this will be running on an Iowa State server.

5.8 Results

Our primary focus is to attempt to implement our current designs and to see if it is possible. WordPress constrains our creative freedom and technology stack to a certain extent and it is something that we will have to work with to some capacity. Fortunately, we can also work around WordPress if it doesn't affect the accessibility and overcomplicate the webpage administration. Currently, an issue we have run into is that we don't have access to the backend database and are forced to use plugins to achieve some of our goals. This changes our plan to a certain extent, but it will help fulfill the goal of making the website accessible to administrators.

However, the results of our cumulative testing through the aforementioned methods has helped us gain a better understanding of what needs to be improved upon, what aspects of our project are in a good place, and where our team strengths and weaknesses lay. Much of our testing has been qualitative thus far, such as evaluating the UI/UX design of our solution. We also ensure that each change that is made is in compliance with all IEEE standards and the requirements from our client.

6 Implementation

For the actual implementation process of our project, we will be pushing our design to github as the primary code repository. In the initial meeting with our client/advisor, Doug Jacobson, it was stated that we would not be making any significant changes to the repository and Cyber House Rock website this semester. This is so we are able to craft a very well laid out plan and design for Ask Captain Cyber before we begin making changes to an official Iowa State University website. We have developed prototypes in the form of a webpage that users would be able to interact with and a sample of working with the ChatGPT API which we are currently testing on our local machines. We have followed our Gantt Chart very closely this semester and will continue to make progress on our project through the end of this semester and at the beginning of the next.

We originally were going to make an external database to handle users however we quickly found out this would not work. The users would have to be manually added by us and would be the ambassadors. We need these users to not have WP-admin access, but they should be able to access an ambassador page where they can access a question database. For the users we found that WordPress has plugins that allow you to create users, and you can use php to restrict pages based on user roles that you give. We will use this implementation for the users. For the question database we will probably need to have another server that logs and tracks questions going to the ChatGPT API.

7 Ethics and Professional Responsibility

7.1 Areas of Professional Responsibility/Codes of Ethics

Area of Responsibility	Definition	Relevant Item from Code of Ethics	Interaction Professional Responsibility
Public Safety and Welfare	Make sure engineering solutions do not harm public safety or well-being.	"To hold paramount the safety, health, and welfare of the public, to strive to comply with ethical design and sustainable development practices." [5]	Our team has prioritized safety and welfare in our design in a couple of ways. Safety by making sure our site has encrypted transactions and does not save unnecessary user data. Welfare by limiting the responses of the AI to not give harmful information to the user.
Honesty and Integrity	Be truthful and transparent in all professional interactions.	"To be honest and realistic in stating claims or estimates based on available data." [5]	We are transparent by clearly documenting our work and openly communicating progress and setbacks on the project.
Confidentiality and Privacy	Protect any sensitive information.	"To protect the privacy of others." [5]	We ensure that all confidential project data is handled securely.
Conflict of Interest	Avoid taking action in situations where personal interests conflict with professional duties.	"To avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist." [5]	Our team has identified and mitigated any potential conflicts of interest, such as personal biases.

Area of Strength: Honesty and Integrity

Our team is performing well in the area of "Honesty and Integrity." We maintain clear communication when it comes to arizing problems within the project. For example if we run into a versioning error, we would come together as a team and look for a solution that will work. This approach allows us to stay on the same page and speed up the production process by eliminating accidental dead ends related to not communicating change.

Area for Improvement: Conflict of interest

An area where our team needs improvement is “Conflicts of Interest.” This is because while we have made an effort to eliminate obvious conflicts, we haven't been able to come up with a good way to detect the smaller conflicts. To improve we plan to make a clear and concise form for detecting conflicts we may have. This form will allow us to cover a broader search area and detect ones that we may have never thought about.

7.2 Four Principles

Area	**Beneficence**	**Nonmaleficence**	**Respect for Autonomy**	**Justice**
Public health, safety, and welfare	Provides actionable cybersecurity advice, seeking to stop things like scams and cyberattacks.	Incorrect advice could be given, leading to user harm.	Gives users the knowledge to make cybersecurity decisions.	Ensures access to cybersecurity guidance for underserved, vulnerable populations.
Global, cultural, and social	Promotes a secure global cyberspace by raising awareness and providing a form for cybersecurity questions.	May overlook nuances in cybersecurity practices or advice.	Fit for diverse user needs, by providing a user-friendly interface and general to complex advice.	Gives knowledge to the general public reducing disparities, making cybersecurity resources accessible to a diverse range of people.
Environmental	Limits resource consumption by using existing Iowa State computing power.	Uses cloud servers that may contribute to emissions.	Uses a digital solution, avoiding the need for additional physical resources and hardware.	Utilizes shared resources to minimize environmental impacts caused by resource overuse.
Economic	Freely accessible platform allows users to save money by	Could lead to unintended financial harm if the advice is	Allows users to make independent financial	Provides free access, ensuring equal opportunity for

	avoiding scams.	incorrect.	decisions through the knowledge base.	all users to access and gain cybersecurity knowledge.

Important Pair: Public Health, Safety and Welfare - Beneficence

Our team prioritizes public welfare by designing Ask Captain Cyber to help users with cybersecurity related questions. The platform will provide practical and vetted advice to reduce the chances for the user to fall victim to scams or cybercrimes. This will be rigorously tested and vetted to ensure an accurate and intelligent response to the users question.

Lacking Pair: Environmental - Nonmaleficence

While we aim to minimize environmental impact, our reliance on cloud servers may contribute to emissions. This negative is reduced by the more positive aspect of shared resource architecture that minimizes physical resource cost. To improve this we could look into partnerships with cloud companies that offset their carbon footprint.

7.3 Virtues

Our team has numerous virtues which we all have committed to so that we can foster an effective team environment. These virtues include:

1. Commitment to quality - Dedication to completing work that meets or exceeds expectations. Our team has all contributed to this aspect in numerous ways. This includes setting clear standards of what needs to be completed and when, encouraging each other to push ourselves and excel, etc.

2. Honesty - Truthfulness, transparency, and straightforwardness are all aspects of being honest on a team. Our team has consistently promoted maintaining open communication and to be truthful and transparent about everything group-related. This way, should someone not be able to get something done on time, another group member would be able to step up and help out instead of the entire group becoming bottlenecked and fall behind.

3. Friendliness - Maintaining a positive and respectful attitude which thus contributes to an enjoyable and collaborative team environment. Friendliness is a very important aspect of our group dynamic, as we all need to get along to be effective in achieving our goals. There is no direct way to accomplish this; friendliness is a constant effort of team bonding, showing appreciation towards one another, respectfully communicating with each other, etc.

In addition to the team virtues we all have, we each have our own individual virtues which contribute to the overall team environment.

Ethan Comiskey

One virtue I believe I have demonstrated in my senior design work so far is cooperativeness. As part of a group project, I believe cooperativeness is not only important, but should be the primary objective. If a team can function as a unit and can coordinate strengths and divide up the work equally, then it is an effective team. I firmly believe there is no team if there is no coordination or cooperation. I believe I have demonstrated cooperativeness because I am willing to work with my teammates on whatever they need help with and play to my strengths where my teammates may be weaker. I have also helped the team coordinate with each other as well when there have been struggles and miscommunications.

One virtue I believe I could improve upon is the virtue of courage. For the most part in this project, I have just stuck with what I know the best instead of trying to learn something new. It is always important to me to try and branch out of my comfort zone to grow as a person. One thing I might do to improve on demonstrating this virtue is to learn something new which I could then apply to this project, such as frontend development. This way, I could expand my skill set and also help my teammates at the same time by adding additional knowledge to the group.

Steven Ragan

One virtue I have demonstrated through my senior design work so far is diligence. Diligence is important because it ensures consistent effort and attention to detail throughout the project. This ensures that requirements are met and there are minimal oversights. I have shown diligence by thoroughly researching AI models to use double checking that needs are met. I also have shown it through consistent communication with the team.

One virtue I could improve upon is the virtue of empathy. Empathy is important to me because it is key to understanding user needs while designing a product. I plan to demonstrate empathy by producing more user-focused research, ranging from interviews to surveys, to understand better how Ask Captain Cyber can address the specific concerns of diverse users.

Caden Murphy

One virtue I believe I have demonstrated well is critical thinking. This is an important virtue to demonstrate in every line of work, and is highly applicable to our project. Being able to formulate a plan, figure out what needs to be done to achieve said plan, and acting upon it amidst and blockers is an important life skill. One way I have demonstrated this is by architecting the front end layout of our project, taking into account any and all edge cases while meeting the project's requirements.

One virtue I could improve upon is consistency. This virtue is important since most things in life are not completed by working on them around the clock at the last minute, but by incrementally working on them and slowly building upon what you have. I have been demonstrating this virtue well so far, but it still can be improved upon. One way I could do so is by taking time to more frequently work on things in smaller time blocks.

Alexander Kronau

One virtue I believe I have demonstrated throughout this project thus far is empathy. Empathy is an important virtue since collaborative environments rely on the ability of people to be able to sympathize with and understand each other. Not everyone is in the same situation, and things may unexpectedly pop up. I have demonstrated this so far by being able to change plans, communicate effectively with other members and helping others out when it is needed.

One virtue I believe I can improve upon is patience. This virtue is important because it is critical that team operations are not rushed, people are not pushed to their limit, and we foster a collaborative, mutually supportive environment. Overall, I believe I have demonstrated this virtue decently well, but think there is still room for improvement. One way I could improve upon it is by taking into account the dynamic nature of many people's schedules and formulating plans that are able to adapt to unforeseen circumstances.

Alex Elsner

One virtue that I believe that I have demonstrated is foresight. Foresight is an important virtue, especially when dealing with longer and more complex projects. This is important to me because being able to identify potential future issues and how to deal with them allows me to stay ahead of schedule and allows us as a team to adapt to future problems better. The quicker and more effective we can identify future problems or solutions, the better the overall product will be. I believe I have demonstrated this by identifying future problems and potential solutions for the backend and adjusting the backend plans with Casper accordingly. This includes talking to Doug, and doing future research into plugins and solutions we might need.

One virtue that I could improve on is accountability. Accountability for myself is important to me. This means that if I say I am going to work on something I do. This has been a struggle for me this semester because I had things I wanted to work on, but had to postpone due to other school work. This happened frequently and it led to me not getting as much work done as I wanted to. In order to show this virtue, I need to set up a set schedule and follow through with it. This goes with having good time management so I can get all my work done on time.

Casper Run

One virtue that I believe I have been exemplifying is adaptability. This is an important virtue for me because in order to deliver a successful final product we need to adapt to possible change throughout the development process. As we did research, Alex Elsner and I attempted to implement some of our research into a practical design. Unfortunately, we needed to change course, but we were able to adapt and find a solution to our problem.

One virtue that I think I need to work on is directness. This is an important virtue for me because sometimes I am not the best at communicating my ideas and what I think needs to be done for the project. Before meetings, I can prepare a list of things I believe need to be done and present it to the group. I can also inform my team more of some decisions I would like to make, so that it doesn’t come as a surprise if I attempt to implement something in my own development environment.

8 Closing Material

8.1 Conclusion

Ask Captain Cyber arose from the increasing number of cyber attacks and crimes targeting people and companies. People who are less technologically literate are getting attacked in cyberspace at an alarming rate. We aim to utilize AI and the cybersecurity community to create a platform where people can ask cybersecurity questions, and get an expert verified response back. Our solution uses a ChatGPT API with a react frontend to create a chatbot where users can ask cybersecurity questions. The admin team will manually add cybersecurity ambassadors. The ambassadors can view users' questions and verify the AI's correctness. The essential goal is for the cybersecurity community to build an LLM from which the AI can pull.

So far we have started to slowly integrate the AI into WordPress using plugins. With help from Professor Bell, we learned how to set up a basic chatbot with relative ease. Integrating it with WordPress will be the challenging part. Our team also has a basic login page done with React, and we have set up Github for version control and progress tracking.

Our constraint was that we needed to have everything integrated into WordPress. This causes many problems because it limits what we can do. We wanted to create a user database, but we had to instead use a wordpress plugin for user management, which constrains what we can do. The other problem is that we have no good way to create a LLM without an external server, so we need to figure out how to integrate external services with WordPress, which may or may not work. Overall, WordPress is a major constraint but is part of the requirement for the project.

In the future, we need to communicate more with our advisor to figure out the constraints earlier. Instead, we had to scratch many plans and develop new ideas as more information came out on what we were and were not allowed to do. This could have been avoided with better communication from our advisor and better planning on our part.

8.2 References

[1] “DINA | ECE @ Iowa,” Uiowa.edu, 2024. https://dina.engineering.uiowa.edu/ (accessed Dec. 07, 2024).

[2] “Assistants API overview beta,” OpenAI Platform, https://platform.openai.com/docs/assistants/overview (accessed Dec. 7, 2024).

[3] “Empowering the world to develop technology through collective knowledge.,” Stack Overflow, https://stackoverflow.co/ (accessed Dec. 7, 2024).

[4] “Cybersecurity ambassador program,” Iowa Cyber Hub, https://www.iowacyberhub.org/ambassadors/ (accessed Dec. 7, 2024).

[5] IEEE, “IEEE Code of Ethics,” ieee.org, Jun. 2020. https://www.ieee.org/about/corporate/governance/p7-8.html

8.3 Appendices

Attached below is a provided early demo of the front end page navigation and interaction of “Ask Captain Cyber”.

AskCaptainCyberPrototype.mp4

9 Team

9.1 Team Members

Alex Elsner - Cybersecurity Engineering Undergraduate

Alex Kronau - Computer Engineering Undergraduate

Caden Murphy - Computer Engineering Undergraduate

Casper Run - Cybersecurity Engineering Undergraduate

Ethan Comiskey - Cybersecurity Engineering Undergraduate

Steven Ragan - Cybersecurity Engineering Undergraduate

9.2 Required Skill Sets for Your Project

Software architecture. Teamwork. Programming practices. Cybersecurity principles.

9.3 Skill Sets covered by the Team

Alex Elsner - Software architecture. Teamwork. Programming practices. Cybersecurity principles.

Alex Kronau - Software architecture. Teamwork. Programming practices. Cybersecurity principles.

Caden Murphy - Software architecture. Teamwork. Programming practices. Cybersecurity principles.

Casper Run - Software architecture. Teamwork. Programming practices. Cybersecurity principles.

Ethan Comiskey - Software architecture. Teamwork. Programming practices. Cybersecurity principles.

Steven Ragan - Software architecture. Teamwork. Programming practices. Cybersecurity principles.

9.4 Project Management Style Adopted by the team

Our project management style is a combination of agile and waterfall. Our style is very similar to agile in that we prioritize flexibility, constant collaboration with each other, and breaking down our project into iterative tasks that will be completed step by step. We also need to be able to change our project if any changes are required by our advisor, adding to the flexibility of the project. Our style is different because we cannot meet as consistently as we should for a proper agile management style since we only meet once a week with each other and monthly/as needed with our advisor.

9.5 Initial Project Management Roles

Alex Elsner - Backend Developer

Alex Kronau - Full Stack Developer

Caden Murphy - Frontend Developer

Casper Run - Cybersecurity/WordPress Developer

Ethan Comiskey - Cybersecurity Developer/Manager

Steven Ragan - Cybersecurity/AI Developer

9.6 Team Contract

Team Name sdmay25-07

Team Members:

1) Ethan Comiskey
2) Steven Ragan
3) Alexander Kronau
4) Casper Run
5) Caden Murphy
6) Alex Elsner

Team Procedures

The team meeting format will be once a week minimum, face-to-face, depending on the availability of team members. The team has a discord group chat to use for communication for updates, reminders, issues, general discussion, etc. This will be the primary method of communication for the entire team. For decisions, a majority vote will be the team's decision. Considering there are six team members, if there is a split vote then we will take the decision to Doug Jacobson, our project adviser. Record keeping will be done each meeting in a google doc. There will be a rotation of who has to take notes during each meeting. They will be shared via Google Docs which each team member has access to.

Participation Expectations

Each team member is expected to make their best effort to attend meetings. If a team member is late then it must be communicated to the team. Should availability not work out, times will be chosen that benefit the majority of team members each week. Each team member must contribute in some way at meetings (i.e. speaking, record keeping, working on assignments/documents, etc). Responsibility for team assignments is assigned when we get them, all distributed equally across the team. These assignments will be completed on time. If there is a delay in completing these assignments on time, communication to get help from other teammates is expected. Communication with other team members is expected as needed or to give updates on certain assignments. Each team member must have a say in decisions and ensure they are committed to completing their tasks on time.

Leadership

There will be no direct leader of the group, each team member will have an equal say, and team members will keep each other on track. Should one team member fall behind, it falls on all the other team members to ensure that each team member catches back up. Each team member is responsible for their own assignments. Each team member can and should also support other team members when possible and needed, such as when team members are falling behind. This will ensure the project will be completed on time. Positive reinforcement will be used to support team members when completing assignments. At each meeting we will go over what we have done since the last, which will be the time in which we recognize the contributions of team members.

Collaboration and Inclusion

Casper Run:

1. Besides the typical classwork in my degree program, I have had several internship experiences in cybersecurity and IT industries. At UL Solutions, I created training material for exploiting web vulnerabilities and consumer IoT devices. I also got to work with several industry security standards, which made me familiar with typical product security. My IT experience involved installing new equipment, documenting, and providing ideas to help improve the ticketing system.
2. I make sure to hear everyone out when they present ideas. A big thing I like to do is provide honest feedback and a way to improve an idea if needed.
3. If teammates have issues making deadlines for a valid reason, I am usually very understanding and attempt to make a workaround. Otherwise, I expect everyone, including myself, to look at everything on a case-by-case basis.

Ethan Comiskey

1. I have two internship experiences at Wells Fargo in Cybersecurity as well as an IT apprenticeship at Ruan Transportation Management Systems. I have some experience in non-technical areas like risk management, data loss prevention/policy, writing documentation, etc. I also have knowledge of technical areas such as coding (Java, Python, C), cybersecurity concepts and tools, certifications (CompTIA Security+ & Microsoft Azure Fundamentals).
2. I will make sure to listen to all team members about their ideas and contributions in team meetings/external communications to ensure everyone feels heard. This should create mutual respect between myself and other team members.

3. If a team member has issues throughout the semester, I will make sure to try to talk to them one on one to get them back on track. Otherwise, I will have to escalate it to the team and then potentially to Doug Jacobson.

Alex Elsner

1. I have internship experience working in cybersecurity at Sargento. This includes cyber analysis, email security, and various parts of the network. I have previous backend development with cybersecurity in mind using Python, Java Script, and Java. I am also currently writing material for cybersecurity ambassadors for ISU.
2. In order to encourage and support team communications I will make sure to frequently communicate with teammates on discord and always be open to talking or new ideas
3. In order to resolve collaboration issues we can do a team meeting or reach out to those who are having issues to resolve it respectfully and gracefully. It will be best to take a balanced approach so both sides work together and compromise.

Alexander Kronau

1. I have multiple internship experiences in IT and Project Management, with knowledge of multiple programming languages and computer networking concepts - knowledge which can be leveraged across multiple aspects of our project.
2. To help facilitate teamwork and significant individual contributions we will make sure to explicitly outline our expectations and maintain consistent communication.
3. If there are collaboration issues, we can bring them up in the team meeting or reach out to the individual on our selected communications applications to help resolve the issue.

Caden Murphy

1. I have internship experience in developing scripts for the Quality Assurance department of the Options Clearing Corporation (OCC) which is heavily regulated and fundamental to the options market and overall economy. Proficient in multiple programming languages such as Java, Javascript, C, and Python. I have built many frontend aspects of both apps and websites.
2. In order to encourage and support team contributions, I will be sure to openly communicate when a team member does well and remain open-minded to all ideas.
3. To resolve any issues, I will make clear communications to voice my opinions. If asked to resolve an issue that is obstructing another member I will take into consideration and do my best to fix the problem at hand.

Steven Ragan

1. I have experience with many cybersecurity aspects, focusing on penetration testing. I also have experience with development cycles. Lastly, I have experience with multiple coding languages, from Java to C++.
2. I will make sure to openly communicate with the team and listen to conflicting views in order to maintain a friendly and productive environment.
3. To resolve any team issue, I'll make sure to provide an unbiased view to see both sides to understand the conflict. Once I am able to understand the conflict, I will provide a solution that either benefits the team or lets both parties win.

Goal-Setting, Planning, and Execution

The primary goal for this semester is to create a functional prototype of Ask Captain Cyber. It does not need to be the final design or have very functional code, but the skeleton should exist.

Assigning individual and team work will be as needed equally across the entire team. Work for multiple members will be assigned to the appropriate team members based on their primary focus on the team (i.e. backend/frontend work, AI-focused). Each team member has another member with their focus, this way, no member will have to work alone on a task and each can keep the other on task to ensure it gets completed.

Consequences for Not Adhering to Team Contract

Infractions of this team contract will result in a team discussion regarding serious misconduct/not being a good team member. If these infractions continue, it will be brought up with our adviser, Doug Jacobson.

****

a) *I participated in formulating the standards, roles, and procedures as stated in this contract.*
b) *I understand that I am obligated to abide by these terms and conditions.*
c) *I understand that if I do not abide by these terms and conditions, I will suffer the consequences as stated in this contract.*

1) Ethan Comiskey	DATE 9/17/24
2) Steven Ragan	DATE 9/17/24
3) Alexander Kronau	DATE 9/17/24
4) Casper Run	DATE 9/17/24
5) Caden Murphy	DATE 9/17/24
6) Alex Elsner	DATE 9/17/24