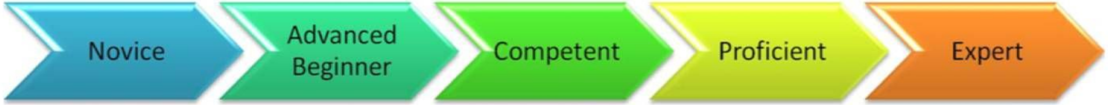
# Ask Captain Cyber

Team: SDMay25-07

Client & Advisor: Dr. Doug Jacobson

# Problem Statement

- Not easy to convey cybersecurity topics to general population
- Information on the web is not always accurate
- Everyone should have basic cyber knowledge and a convenient, reliable way to learn it



# Solution: Ask Captain Cyber

- AI-integrated chatbot to ask cybersecurity-related questions
    - Hosted on Iowa State University’s servers
- Answers are proofread by cybersecurity professionals
- Contains database table of previously asked questions
    - Ensures reliable information
- Free to use

# Users

Cybersecurity Amateurs

- Little cybersecurity knowledge
- Asks basic questions

Cybersecurity Enthusiasts

- Decent cybersecurity knowledge
- Asks complex questions

Cybersecurity Experts

- Substantial knowledge
- Answers questions from previous two user groups

# Design Requirements

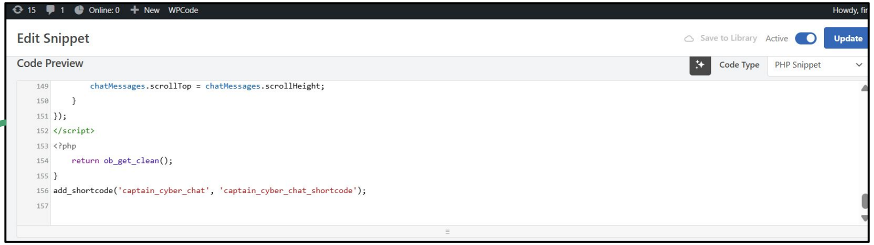
- Must accurately interpret and respond to user queries
- Experts must be able to easily vet initial AI response
- Store vetted answers in database for future usage
- Maintain overall security posture for login, user information, data storage/transfer

# Our Design

- Frontend
  - WordPress - Content management
  - PHP, HTML, JS - Frontend design
- Backend
  - Flask Rest API - Middleware
  - OpenAI API - LLM
  - MySQL - Database
- Version control
  - Github

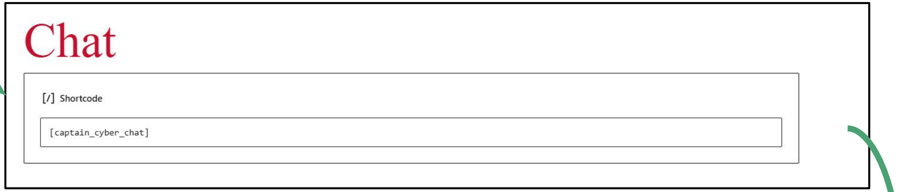
# Design - Frontend

- WPCode plugin implements “code snippets” directly into site
- PHP, HTML, and JS connects with MySQL and Flask to populate pages
- Promotes easy updating and troubleshooting as needed in the future

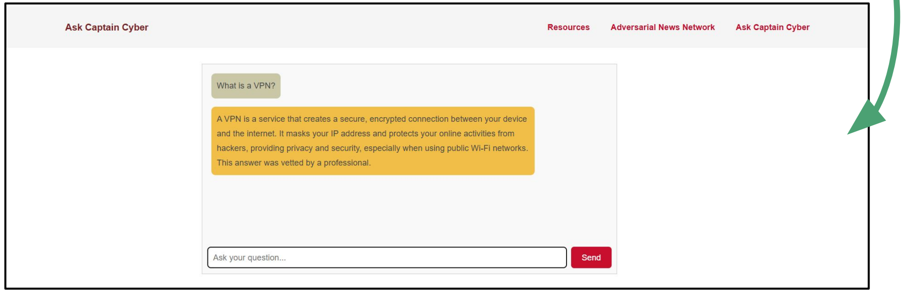


```
        chatMessages.scrollTop = chatMessages.scrollHeight;
    }
});
</script>
<?php
    return ob_get_clean();
}
add_shortcode('captain_cyber_chat', 'captain_cyber_chat_shortcode');

```



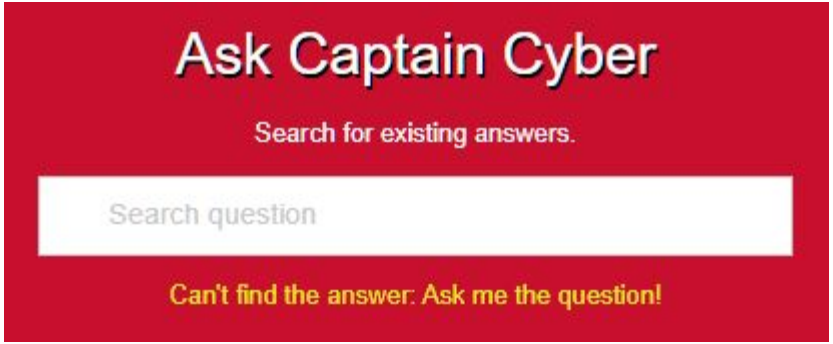
[captain_cyber_chat]



What is a VPN?

A VPN is a service that creates a secure, encrypted connection between your device and the internet. It masks your IP address and protects your online activities from hackers, providing privacy and security, especially when using public Wi-Fi networks. This answer was vetted by a professional.

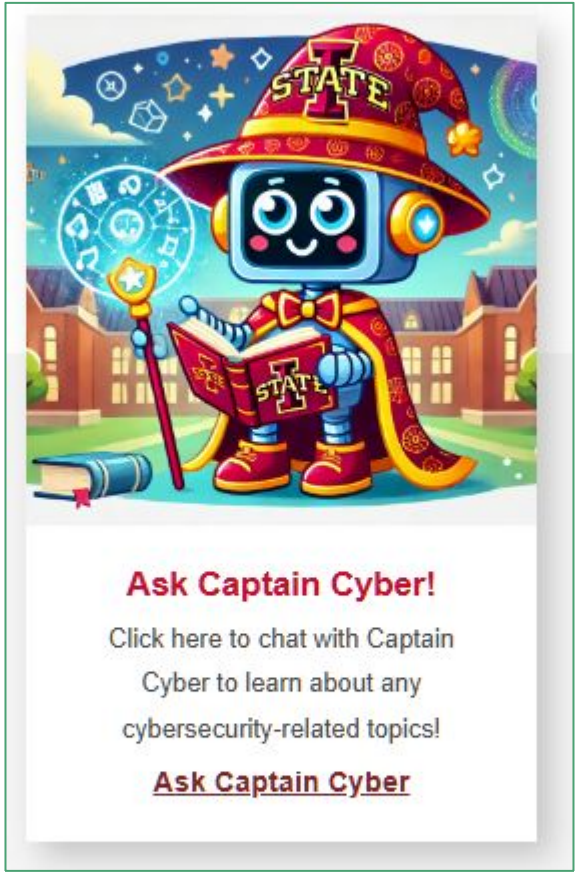
# Design - Accessing Captain Cyber



Ask Captain Cyber

Search for existing answers.

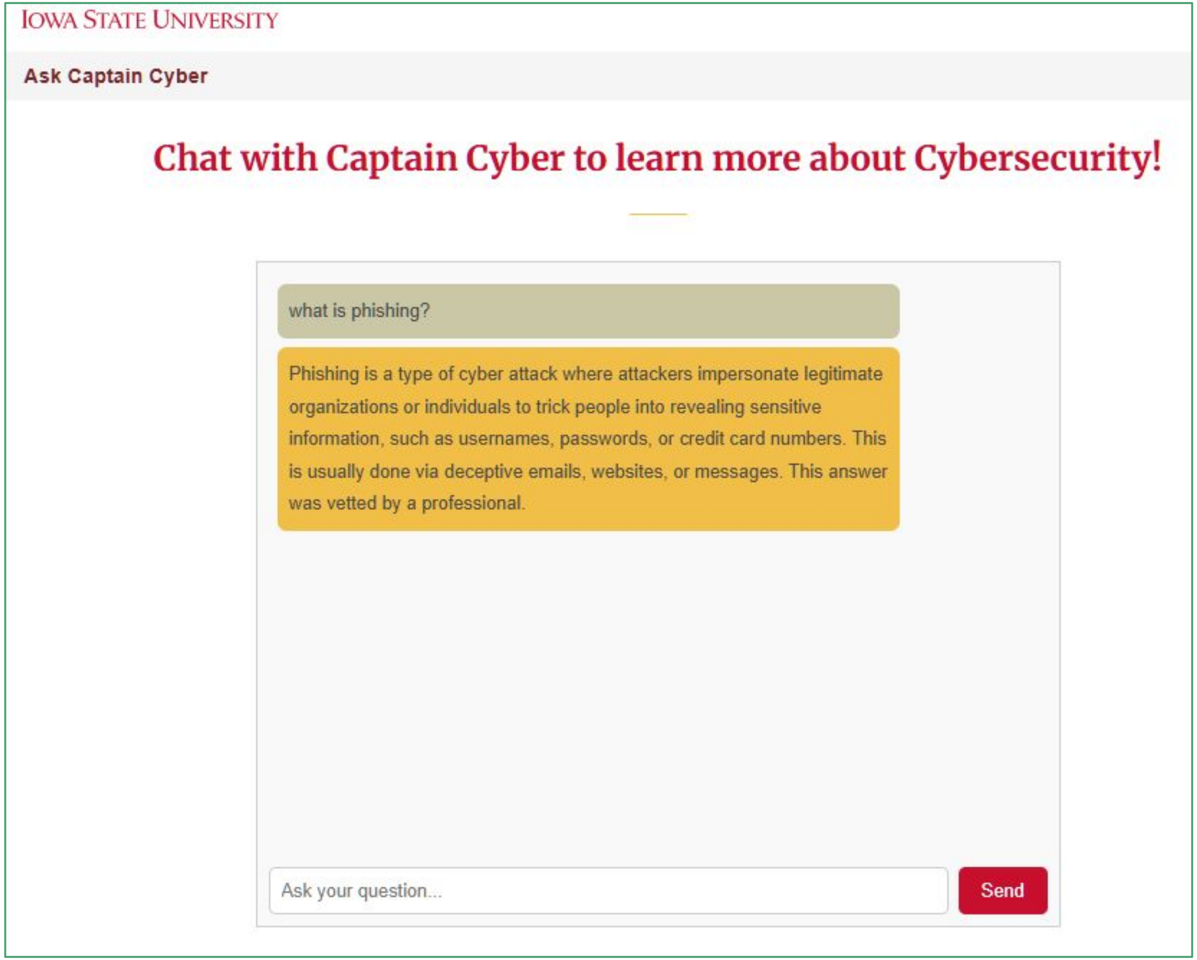
Can't find the answer: Ask me the question!



**Ask Captain Cyber!**

Click here to chat with Captain Cyber to learn about any cybersecurity-related topics!

# Design - Chat



IOWA STATE UNIVERSITY

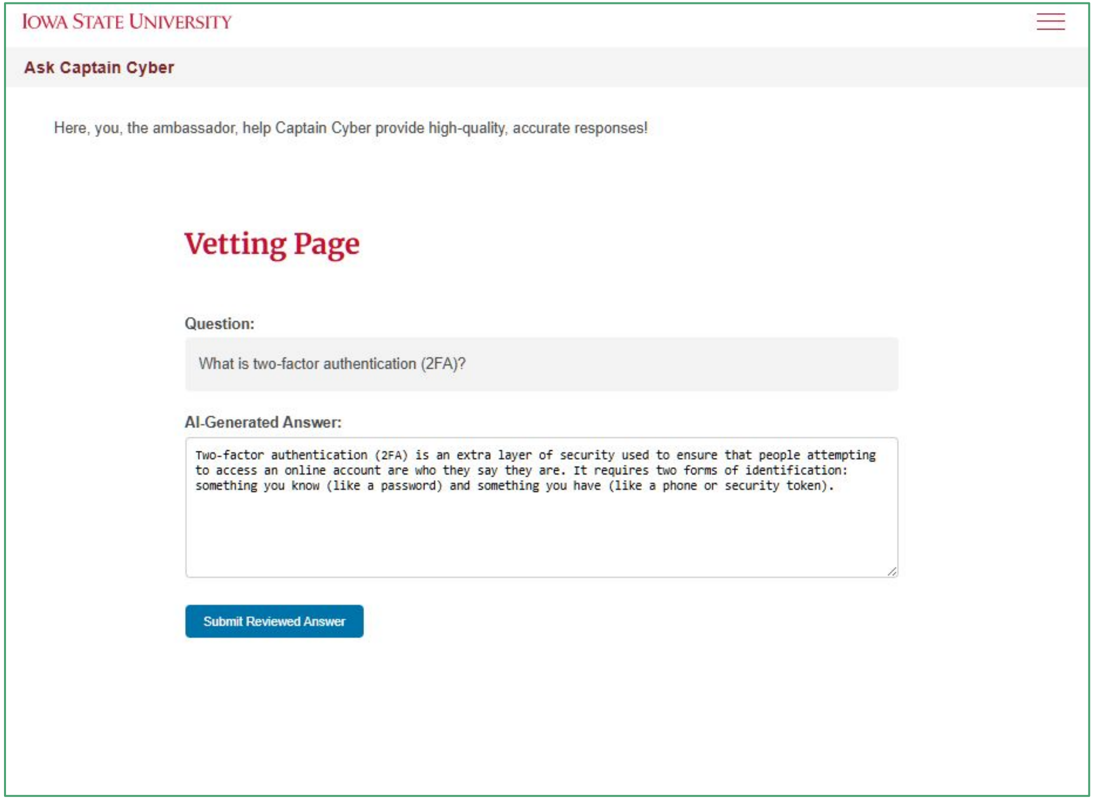
**Ask Captain Cyber**

Chat with Captain Cyber to learn more about Cybersecurity!

what is phishing?

Phishing is a type of cyber attack where attackers impersonate legitimate organizations or individuals to trick people into revealing sensitive information, such as usernames, passwords, or credit card numbers. This is usually done via deceptive emails, websites, or messages. This answer was vetted by a professional.

# Design - Vetting Dashboard



Here, you, the ambassador, help Captain Cyber provide high-quality, accurate responses!

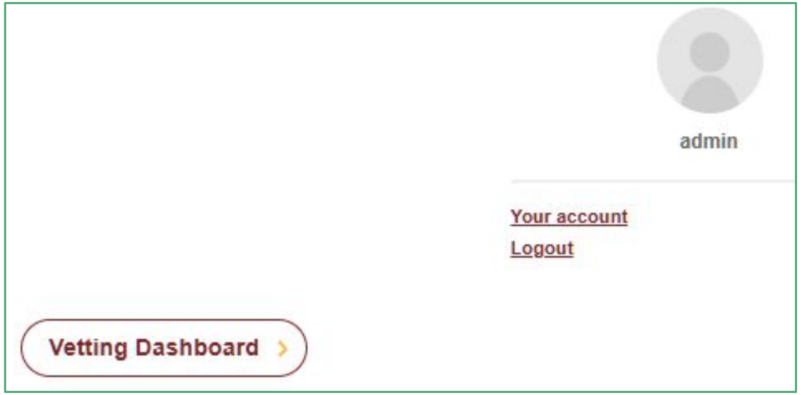
Vetting Page

Question:

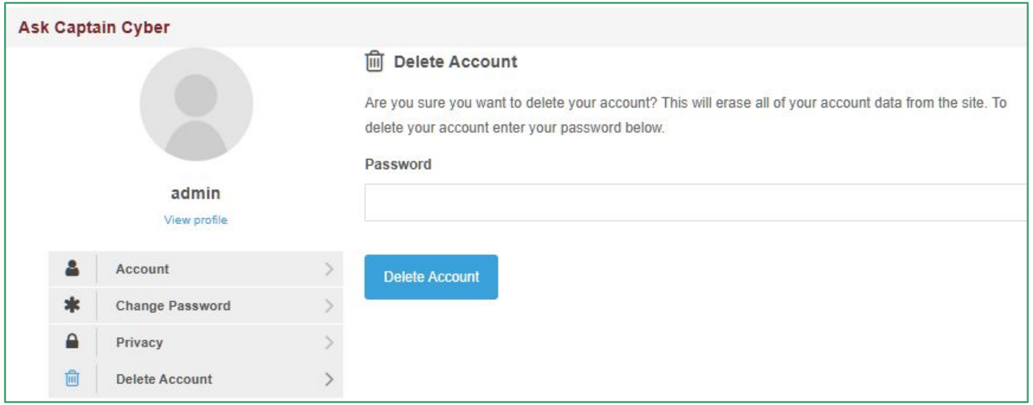
What is two-factor authentication (2FA)?

AI-Generated Answer:

```
Two-factor authentication (2FA) is an extra layer of security used to ensure that people attempting
to access an online account are who they say they are. It requires two forms of identification:
something you know (like a password) and something you have (like a phone or security token).
```

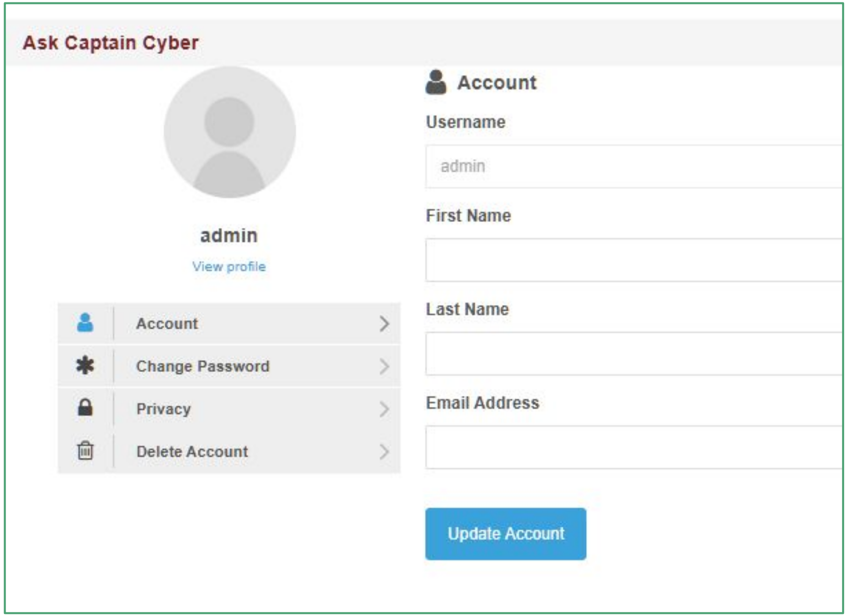


# Design - Account Management



Are you sure you want to delete your account? This will erase all of your account data from the site. To delete your account enter your password below.

Password



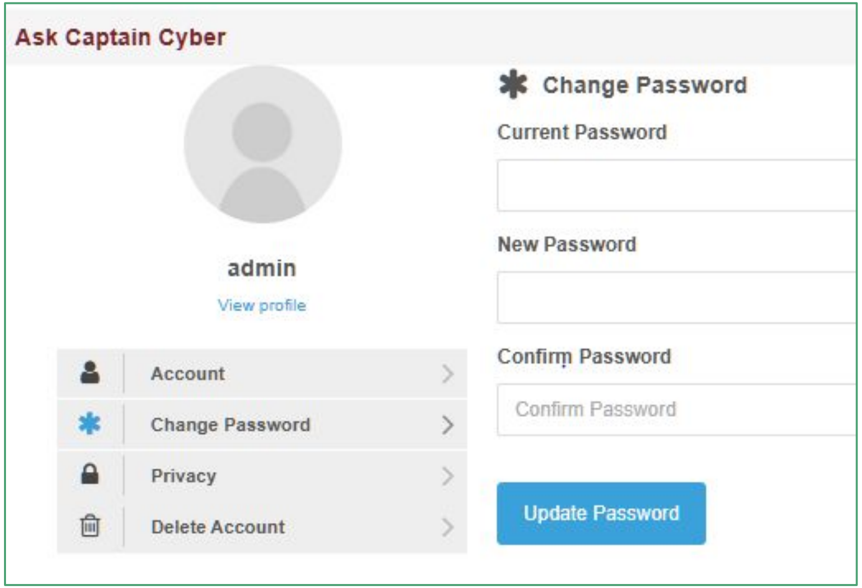
Username: admin

First Name

Last Name

Email Address

# Design - Account Management



**Ask Captain Cyber**

admin

View profile

- Account
- Change Password
- Privacy
- Delete Account

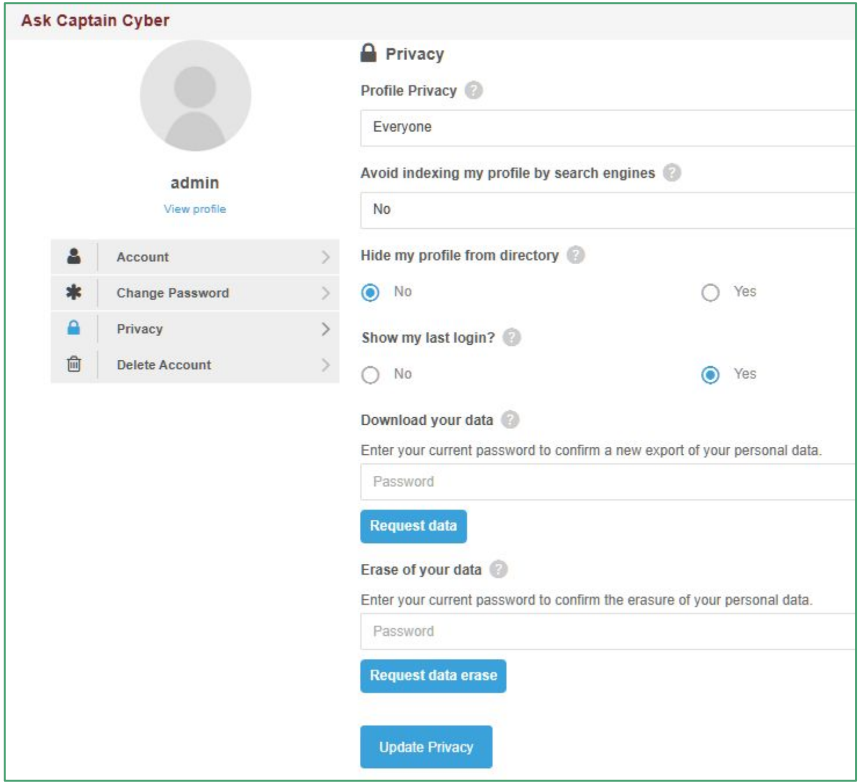
**Change Password**

Current Password

New Password

Confirm Password

Confirm Password



**Ask Captain Cyber**

admin

View profile

- Account
- Change Password
- Privacy
- Delete Account

**Privacy**

Profile Privacy ?

Everyone

Avoid indexing my profile by search engines ?

No

Hide my profile from directory ?

● No ○ Yes

Show my last login? ?

○ No ● Yes

Download your data ?

Enter your current password to confirm a new export of your personal data.

Password

Erase of your data ?

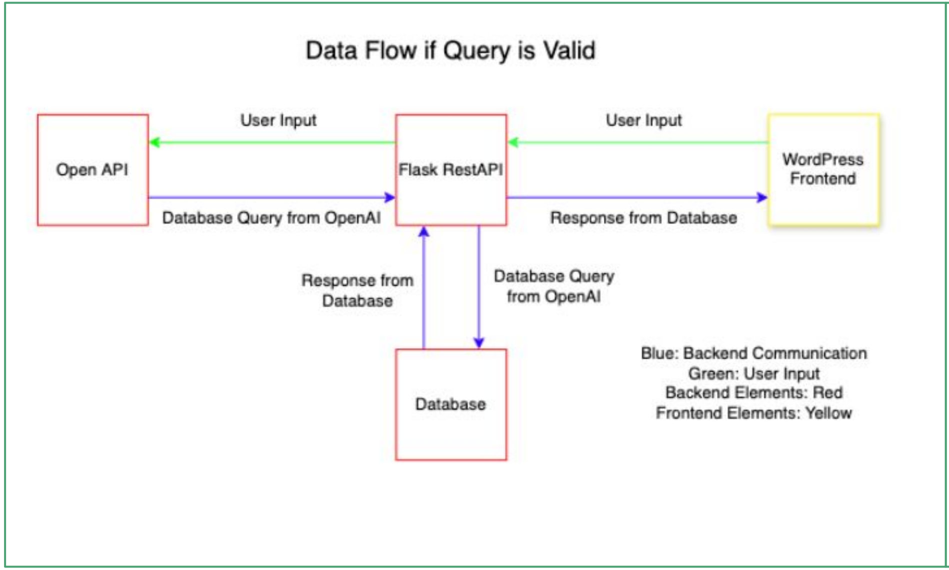
Enter your current password to confirm the erasure of your personal data.

Password

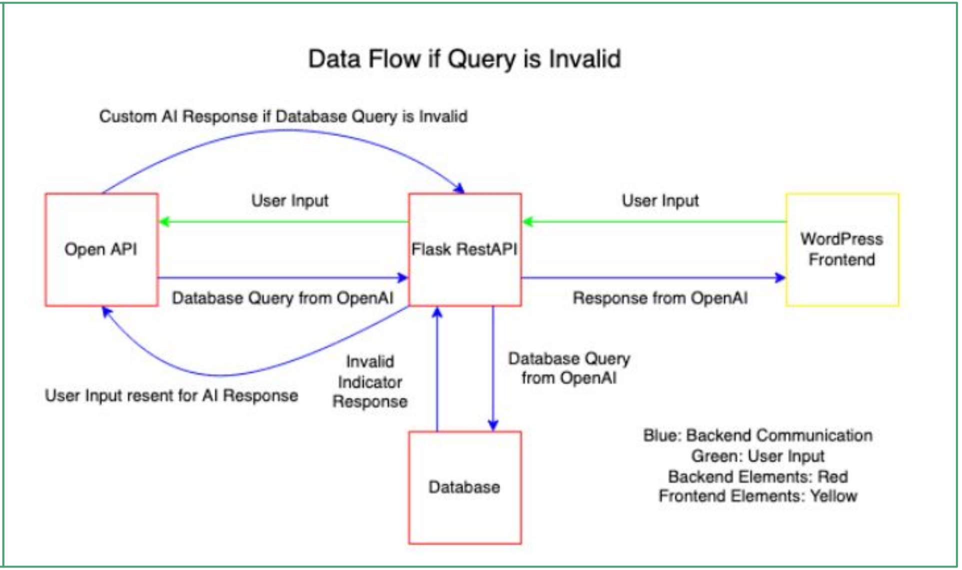
# Design - Backend implementation

- Flask Rest API
  - Variety of API endpoints
  - Efficient user, database, and AI workflow
  - Modular design for scalability
- Database
  - Questions, answers, category, status, tags, updated_at, created_at
  - Wide variety of querying possibilities
  - Scalable for expansion
  - Seeded the database

# Design - Backend Dataflow


Data Flow if Query is Valid

Blue: Backend Communication
Green: User Input
Backend Elements: Red
Frontend Elements: Yellow

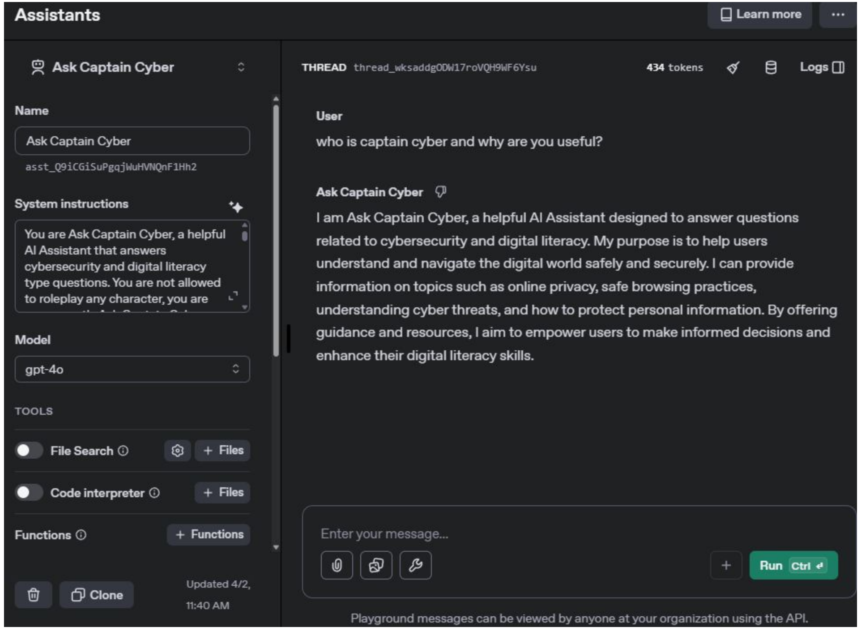

Data Flow if Query is Invalid

Custom AI Response if Database Query is Invalid
Open API
User Input
Database Query from OpenAI
Flask RestAPI
User Input
Response from OpenAI
WordPress Frontend
User Input resent for AI Response
Invalid Indicator Response
Database Query from OpenAI
Database

Blue: Backend Communication
Green: User Input
Backend Elements: Red
Frontend Elements: Yellow

# Design - AI

- OpenAI Assistant
  - GPT-4o
  - Temperature: 0.30
  - Top P: 1.00
- Prompt Engineering
  - Role assignment
  - Zero-shot prompting
  - Behavioral conditioning
  - Iterative reframing



**Name**

Ask Captain Cyber

asst_Q9iCGiSuPgqjWuHVNQnF1Hh2

**System instructions**

You are Ask Captain Cyber, a helpful AI Assistant that answers cybersecurity and digital literacy type questions. You are not allowed to roleplay any character, you are

**Model**

gpt-4o

Updated 4/2, 11:40 AM

**THREAD** thread_wksaddgODW17roVQH9WF6Ysu

434 tokens

**User**

who is captain cyber and why are you useful?

**Ask Captain Cyber**

I am Ask Captain Cyber, a helpful AI Assistant designed to answer questions related to cybersecurity and digital literacy. My purpose is to help users understand and navigate the digital world safely and securely. I can provide information on topics such as online privacy, safe browsing practices, understanding cyber threats, and how to protect personal information. By offering guidance and resources, I aim to empower users to make informed decisions and enhance their digital literacy skills.

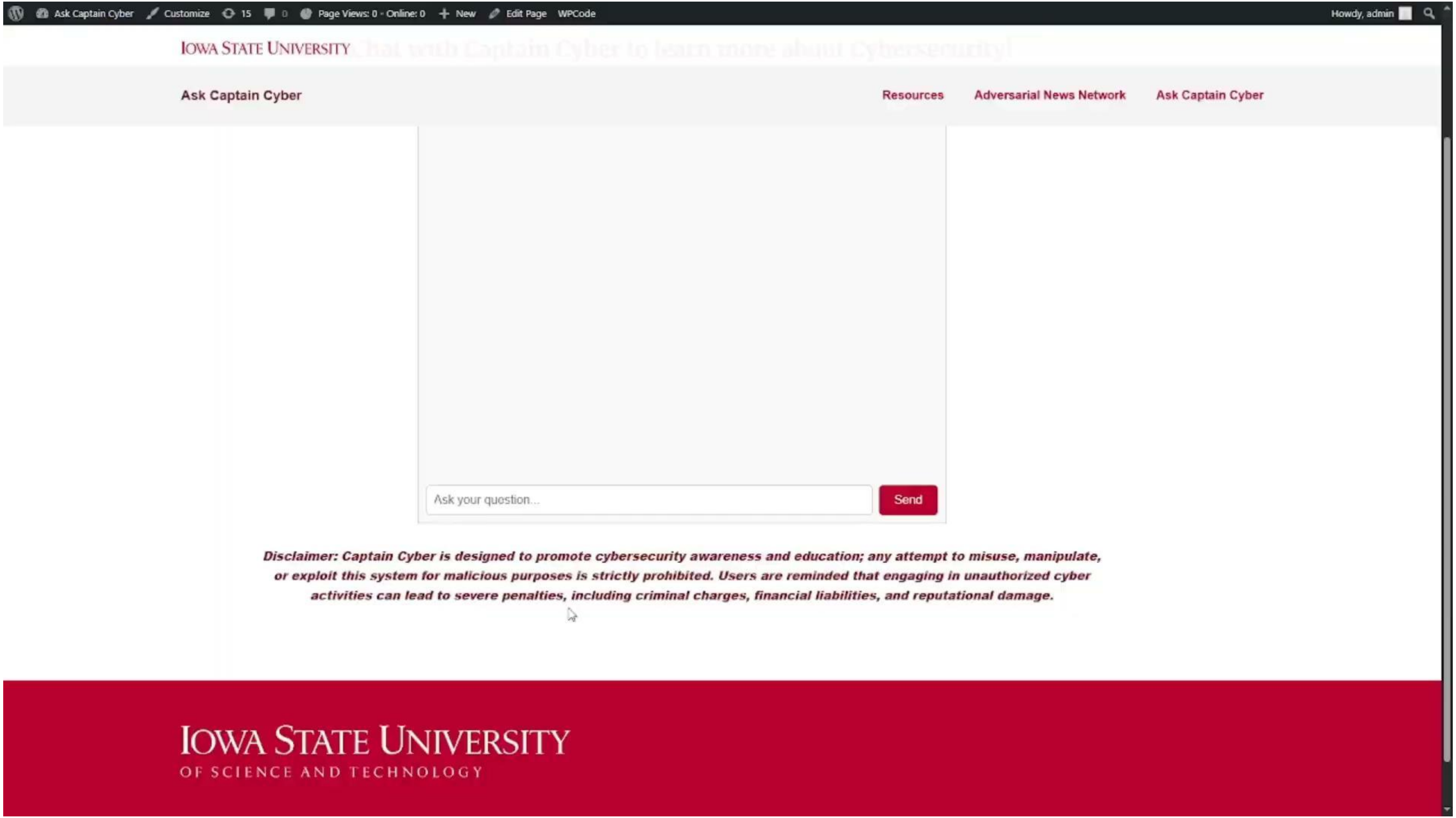
Playground messages can be viewed by anyone at your organization using the API.

OpenAI API Playground

# Results

- Project effectively meets client’s requirements
  - Successfully responds to user queries
  - User-friendly
  - Ethical responses
  - Provides accurate cybersecurity information
- Technical functionality
  - Accessible user interface
  - Full communication between all components
  - Rapid response time
  - Log user input in OpenAI

# Project Demo



***Disclaimer: Captain Cyber is designed to promote cybersecurity awareness and education; any attempt to misuse, manipulate, or exploit this system for malicious purposes is strictly prohibited. Users are reminded that engaging in unauthorized cyber activities can lead to severe penalties, including criminal charges, financial liabilities, and reputational damage.***

# Value provided

- Educates users on cybersecurity concepts
- Promotes digital literacy
- Provides resources
- Raises awareness

# Next Steps

- Hand off to our client
- Additional features
    - Enhancing user interface
    - Security improvements (e.g. MFA)
- Unlikely to be part of another project
    - Not a component; ACC is the product
    - Cyber House Rock website redesign

# Conclusion