# Personal Challenge Proposal

Student Name: Bruno Carvalho  
Date: 01-03-2025  
Course: AI for Society Minor

## 1. Introduction

Cybersecurity awareness remains a major challenge due to the complexity of technical concepts and the lack of accessible educational resources. My project aims to bridge this gap by developing an AI-powered cybersecurity assistant that explains security concepts in a way that is clear, actionable, and tailored to individual users. This tool will not only educate users but also provide personalized recommendations to help them improve their cybersecurity practices.

## 2. Project Objectives & Added Benefit

### Technical Objectives (LO3, LO4, LO5)

#### Data Preparation (LO3):

* Identify and preprocess cybersecurity datasets (e.g., MITRE ATT&CK, OWASP, NIST guidelines).
* Incorporate insights from my own Forms survey (<https://forms.office.com/e/ritvzRPr7e>) to analyze human behavior in cybersecurity.
* Explore additional open-source cybersecurity datasets (Google Dataset Search).

#### Machine Teaching (LO4):

* Since I am new to this, I will research different machine teaching approaches and explore the best way to train an AI model to explain cybersecurity effectively.
* Investigate fine-tuning pre-trained models (GPT, Llama, or BERT) with security-related datasets.

#### Data Visualization (LO5):

* Research and define how to visually present cybersecurity insights in a way that engages users and makes security advice more actionable.
* Consider interactive formats such as decision trees, dynamic charts, or simplified security checklists.

## Contextual Objectives (LO1, LO2, LO6)

#### Societal Impact (LO1):

* Many people do not actively seek to improve their security practices. This project will provide a proactive tool that helps them become more aware of cybersecurity threats and encourages them to take concrete steps to improve their online safety.

#### Investigative Problem-Solving (LO2):

* Research why people ignore cybersecurity best practices and how AI-driven education could engage them better.
* Identify barriers to cybersecurity learning and explore how AI can adapt to user knowledge levels to improve engagement.

#### Reporting (LO6):

* Document my research, AI implementation process, and findings in my Personal Development Report (PDR).
* Reflect on feedback received and track my learning progress throughout the semester.

### Added Benefit

This AI-powered assistant will allow users to understand cybersecurity risks and act, reducing vulnerability to cyber threats like phishing, weak passwords, and social engineering. Unlike existing AI-driven security tools, this assistant is educational-first, ensuring that users not only receive security tips but understand the reasoning behind them.

## 3. Feasibility & Approach

### Data & Resources

* MITRE ATT&CK, OWASP Top 10, NIST Framework, phishing datasets.
* Pre-trained LLMs (GPT-3.5, Llama, or BERT) to fine-tune for security explanations.
* AI development tools (Hugging Face, OpenAI API, TensorFlow/PyTorch).
* Form Survey.
* Open-source datasets from (<https://datasetsearch.research.google.com/>) .

### Model Training & Implementation

* Begin with retrieval-augmented generation (RAG) to ensure factual accuracy.
* Fine-tune an LLM with cybersecurity-specific datasets to improve explanation quality.
* Implement a feedback mechanism where users can clarify doubts or ask for simpler explanations.

### Challenges & Considerations

* Data Privacy: Ensuring that no sensitive data is stored or misused.
* Ethical AI Use: Preventing misinformation and ensuring security advice remains reliable.
* User Accessibility: Making explanations adaptable to different knowledge levels.

## 4. Personal Goal (LO8)

Since my long-term career goal is in cybersecurity, my **Personal Challenge** is to develop AI expertise within this field by:

1. Gaining hands-on experience with AI model fine-tuning and data preparation.
2. Improving my ability to bridge technical cybersecurity knowledge with real-world application.
3. Developing a fully functional AI prototype that demonstrates clear, actionable cybersecurity recommendations.

## 5. Expected Outcomes

By the end of this project, I aim to:  
✔ Develop a prototype AI assistant that explains cybersecurity in a user-friendly way.  
✔ Research how to make AI-driven cybersecurity education engaging.  
✔ Receive feedback from my Semester Coach and consultants to improve my work.  
✔ Strengthen my technical AI skills while addressing a real cybersecurity challenge.

## 6. Next Steps

* Gather feedback from my Semester Coach and Technical/Contextual Consultants.
* Continue research on AI model training and cybersecurity datasets.
* Begin documenting progress in my Personal Development Report (PDR).
* Develop an initial prototype to test feasibility and collect feedback.