# **Personal Development Report (PDR)**

**Student Name:** [My Name]  
**Date:** 16-03-2025  
**Course:** AI for Society Minor

## **1. Introduction**

* **Abstract**: This project explores the use of artificial intelligence to improve cybersecurity awareness. The goal is to build a digital assistant that helps everyday users understand and respond to online threats. The assistant answers cybersecurity-related questions using pre-trained machine learning models and a large language model. It only responds to cybersecurity topics and filters out unrelated content. The assistant is being developed using data collected from a large survey, open cybersecurity datasets, and browser automation tools. While still in progress, it already includes several trained models and a working chatbot interface. The final aim is to create a tool that supports users in safer online behavior.

**Research Question**: How can AI be leveraged to enhance cybersecurity awareness and improve individual security behaviors?

**Sub-Research Questions**:

* What are the primary barriers preventing individuals from engaging in proactive cybersecurity measures?
* How can AI models be trained to deliver tailored security advice based on user knowledge levels?
* What visualization techniques improve user engagement with cybersecurity recommendations?

**Technologies and Methods**: The project will utilize large language models (LLMs), survey-based insights, and interactive visualizations to provide cybersecurity guidance.

## **2. Learning Outcome Evaluations**

Each section follows this structure:  
**(a) Explanation of the Learning Outcome**  
**(b) Self-Assessment & Current Progress**  
**(c) Learning Process & Evidence** (Feedback, research, datasets, initial models, survey results, etc.)  
**(d) Reflection & Next Steps**

### **2.1 LO1 - Societal Impact**

* **Explanation**: Enhancing cybersecurity awareness and promoting proactive security behaviors among users.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**: Survey results from 231 participants provided insights into online behavior, showing that a large portion of respondents feel underprepared for common cybersecurity threats. This data shaped the assistant's content priorities.
* **Reflection & Next Steps**: Continue integrating public behavior data into model responses. Future work will include interactive modules informed by survey feedback.

### **2.2 LO2 - Investigative Problem Solving**

* **Explanation**: Addressing the challenges in cybersecurity awareness and proposing AI-driven solutions.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**: Identified key limitations in existing solutions. Researched alternatives such as LLM customization, real-time agent assistance, and threat-specific model training. Integrated the browser-use framework to prototype task-based cybersecurity helpers.
* **Reflection & Next Steps**: Investigate hybrid AI-agent workflows to automate responses to real-world cybersecurity queries. Evaluate performance across categories.

### **2.3 LO3 - Data Preparation**

* **Explanation**: Collecting and refining data sources to enhance model training and cybersecurity insights.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**:
  + Curated labeled cybersecurity datasets (phishing, malware, CVEs, awareness).
  + Collected 231 survey responses analyzing behavior and awareness levels.
  + Trained models using scikit-learn pipelines and evaluated performance (e.g., phishing model: ~0.87 accuracy, CVE model: ~0.82 accuracy)
* **Reflection & Next Steps**: Finalize cleaning scripts, integrate more balanced label distributions, and document preprocessing steps for reproducibility.

### **2.4 LO4 - Machine Teaching**

* **Explanation**: Training an AI model to generate cybersecurity insights in an accessible manner.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**:
  + Developed 10 domain-specific models, each corresponding to key awareness areas such as phishing, CVEs, and malware.
  + Each model was trained using TF-IDF vectorization and standard classifiers like logistic regression and gradient boosting, with preprocessing handled via scikit-learn pipelines.
  + Accuracy ranged from 80–90% depending on dataset complexity and label quality.
  + Models were serialized and stored in the models/trained\_pipelines folder as .pkl files for consistent and efficient runtime loading.
  + These models were integrated into the assistant backend to support topic-specific predictions, although some still require input reshaping and validation logic.
  + A local instance of DeepSeek-R1 14B was integrated using *Ollama*. The assistant uses this LLM only when needed, and its responses are filtered to stay within the cybersecurity domain.
  + Hardcoded fallbacks and category checks were developed to ensure no irrelevant responses are provided.
* **Reflection & Next Steps**: Improve data compatibility across models, refine the prediction interface, and finalize fallback logic for invalid inputs.

### **2.5 LO5 - Data Visualization**

* **Explanation**: Enhancing cybersecurity learning through interactive and visual representations.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**:
  + Began plotting behavior trends from the survey using matplotlib.
  + Identified patterns in security habits (e.g., password reuse, response to phishing).
* **Reflection & Next Steps**: Expand interactive components of the assistant with visual guides, alerts, and confidence ratings per response.

### **2.6 LO6 - Reporting**

* **Explanation**: Documenting research findings, methodology, and results.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**:
  + Structured Personal Challenge Proposal and PDR.
  + Collected evidence (screenshot logs, chatbot output examples, survey graphs).
* **Reflection & Next Steps**: Structured Personal Challenge Proposal and PDR.
* Collected evidence (screenshot logs, chatbot output examples, survey graphs).

### **2.7 LO7 - Personal Leadership**

* **Explanation**: Developing initiative and leadership within AI and cybersecurity.
* **Self-Assessment**: Beginning.
* **Learning Process & Evidence**:
  + Management of project milestones and coordination of research efforts.
* **Reflection & Next Steps**: Prioritize robustness and testing. Seek external feedback before final release.

### **2.8 LO8 - Personal Goal**

* **Explanation**: Gaining AI expertise in the context of cybersecurity.
* **Self-Assessment**: Beginning/Proficient.
* **Learning Process & Evidence**:
  + Built an AI assistant tailored for cybersecurity queries.
  + Customized multiple components (classification, fallback logic, agent tasks).
* **Reflection & Next Steps**: Built an AI assistant tailored for cybersecurity queries.
* Customized multiple components (classification, fallback logic, agent tasks).

## **3. Retrospect (Final Submission Only)**

* **Course Experience**: Analysis of AI for Society minor’s impact on skill development.
* **Challenges & Improvements**: Review of project difficulties and areas for enhancement.
* **Future Applications**: Exploration of long-term applications of AI in cybersecurity.

## **4. Conclusion (Final Submission Only)**

* **Success Assessment**: Justification of learning outcomes achieved and project impact.

## **5. Appendices**

* **Relevant references, datasets, survey results, consultant feedback screenshots, etc.**