**Direct Independent Study Summer 2025 MILESTONE PLAN**

* Name: Onitiju Taiwo
* Topic: Al-Powered Phishing Detection Using NLP & Explainable Al.
* Duration of Study: Summer 2025
* Professor: Dr. Roy Swapnoneel

**DIS: Summer 2025 Milestone Plan**

|  |  |  |
| --- | --- | --- |
| **Week 1 - 2**  **(May 12th – May 23rd)** | Background Research & Setup | * ~~Literature Review:~~   ~~NLP for phishing detection~~  ~~Transformer models in cybersecurity~~  ~~Explainable AI methods (SHAP, LIME)~~   * ~~Technical Setup:~~   ~~Set up Python environment with key libraries (e.g., transformers, scikit-learn, shap, lime, pandas, etc.)~~  ~~Download and explore datasets (PhishTank, Ember, or others)~~ |
| **Week 3- 4**  **(May 26th – June 6th)** | Baseline & Dataset Preparation | * ~~Data Preprocessing:~~   ~~Clean and preprocess email/text data~~  ~~Tokenization for transformer models (e.g., using BERT tokenizer)~~   * ~~Baseline Model:~~   ~~Implement a basic traditional model (e.g., logistic regression, SVM)~~  ~~Record baseline performance (accuracy, F1, precision/recall).~~ |
| **Week 5 – 6**  **(June 9th – June 20th)** | Transformer Model Development | * ~~BERT-based Model:~~   ~~Fine-tune BERT or another transformer on phishing vs. legitimate email classification~~  ~~Evaluate performance using standard metrics~~  ~~Compare with baseline~~ |
| **Week 7 – 8**  **(June 23rd – July 4th)** | Explainable AI Integration | * ~~Apply SHAP/LIME:~~   ~~Use SHAP and/or LIME to interpret model decisions~~  ~~Visualize which tokens/phrases influenced classification~~   * ~~Qualitative Analysis:~~   ~~Analyze explanations for both correct and incorrect classifications~~ |
| **Week 9 – 10**  **(July 7th – July 18th)** | Robustness Testing | * Adversarial Manipulation:   Introduce adversarial examples (e.g., misspellings, word substitutions)  Measure model degradation and robustness   * Mitigation:   Explore techniques like adversarial training or data augmentation |
| **Week 11**  **(July 21st – July 25th)** | Final Evaluation | * Performance Summary:   Evaluate model accuracy, robustness, and explanation quality  Generate tables/figures for results |
| **Week 12**  **(July 28th – August 1st)** | Final Report /Presentation | * Write Final Report:   Include methodology, results, analysis, and visualizations   * Prepare Presentation:   Summarize project goals, implementation, findings, and future work |