Week 4 Direct Independent Study Progress

* **Topic Covering**: Al-Powered Phishing Detection Using NLP & Explainable Al.
  + **Covered Paper: Comparative Investigation of Traditional Machine Learning Models and Transformer Models for Phishing Email Detection.**
    - **Link**: <https://www.mdpi.com/2079-9292/13/24/4877>
    - **Citation:**  *M. Ammar, M. A. Khan, M. A. Khan, and H. R. Qureshi, "Comparative Investigation of Traditional Machine Learning Models and Transformer Models for Phishing Email Detection," Sensors 23, no. 20 (2023): 8594.* <https://www.mdpi.com/2079-9292/13/24/4877>.
    - **Notes**: The authors look at how traditional machine learning (ML) models (Logistic Regression, Random Forest, etc.) compare with transformer models (DistilBERT, BERT, among others) in detecting phishing emails using a composite dataset.
      * They make use of a Comprehensive dataset combining various sources (Enron, Ling‑Spam, SpamAssassin, Nazario, Nigerian, CEAS), similar to my approach.
      * Their technique of preprocessing the dataset involved
        + Standard text cleaning and tokenization.
        + Balanced sampling to mitigate class imbalance.
        + Input sequences formatted suitably for both classic ML and transformer models.
  + **Covered Paper: A Systematic Review of Deep Learning Techniques for Phishing Email Detection.**
    - **Link**: <https://www.mdpi.com/2079-9292/13/19/3823>
    - **Citation:**  *D. F. Kyaw, J. Gutierrez, and A. Ghobakhlou, "A Systematic Review of Deep Learning Techniques for Phishing Email Detection," Electronics 13, no. 4 (2024): 765.* <https://www.mdpi.com/2079-9292/13/19/3823>.
    - **Notes**: This paper does a literature review investigating 33 studies focused on deep learning for phishing email detection. Its objective is to classify techniques, measure their performance, and locate gaps in present research.
      * Issues identified by the authors:
        + Models often lack adaptability to evolving phishing tactics, leading to performance degradations.
        + Few studies integrate Explainable AI methods like SHAP or LIME, limiting transparency in decision-making.
        + Deep architectures, especially those with embeddings, may be resource-intensive, challenging real-time deployment

**Relevance of Paper to my Study**

* From the first paper, I could implement logistic regression or SVM to mirror their approach. I could also implement transformer fine-tuning by starting with DistilBERT or RoBERTa based on their performance insights.
* The first paper’s overall approach was similar to my current research but with my research extending their work by adding explainability (SHAP/LIME) as well as evaluating adversarial robustness.
* The second paper reinforces my present research’s idea for using explainability in the form of SHAP or LIME to enable interpretability. They also proposed the idea of documenting computational resource considerations if I plan to scale up the project.

# **Data Preprocessing**

* ***Still in progress of data cleaning of Phishing Email Dataset in preparation for Transformer Model Development phase.***