2. Large Language Models for Cyber

Description:
This project explores the application of large language models (LLMs) in cybersecurity to enhance real-time threat detection, analysis, and automated response. LLMs can process and understand complex textual data, such as system logs, threat intelligence reports, and user behavior descriptions, enabling advanced capabilities for identifying potential security breaches and mitigating them efficiently.

Objectives:

1. **Threat Intelligence Analysis:** Use LLMs to analyze threat intelligence feeds and extract actionable insights.
2. **Anomaly Detection in Logs:** Train an LLM on network/system logs to detect anomalies indicative of cyber threats.
3. **Automated Incident Response:** Develop a conversational AI-based assistant to guide or automate response actions in real-time.
4. **Phishing Email Detection:** Fine-tune an LLM to identify phishing attempts in email communication.
5. **Dynamic Threat Hunting:** Use LLMs to assist cybersecurity analysts in creating and executing search queries across large datasets.

Implementation Steps:

1. **Dataset Preparation:** Collect and preprocess cybersecurity datasets, including:
    - Threat intelligence feeds (e.g., STIX, TAXII).
    - Phishing email datasets.
    - Network/system logs.
2. **Model Training:** Fine-tune pre-trained LLMs (e.g., GPT, BERT) for specific cybersecurity tasks.
3. **Prototype Development:** Build and evaluate an end-to-end system that integrates LLM capabilities with a security operations center (SOC) workflow.
4. **Evaluation:** Assess the performance using metrics like precision, recall, and real-world testing scenarios.

Expected Outcomes:

- A prototype system demonstrating real-time cyber threat detection and automated response.
- Enhanced threat intelligence processing and actionable insight generation.
- A report on the feasibility, strengths, and limitations of using LLMs in cybersecurity applications.

Potential Challenges:

- Ensuring the model's security against adversarial inputs.

- Managing computational requirements for real-time inference.
- Reducing false positives in threat detection.

Recent References:

1. Li, X., et al. (2023). "Applying Transformers in Cybersecurity: A Survey." *IEEE Access*.
2. Liu, W., et al. (2022). "Fine-Tuning BERT for Intrusion Detection in System Logs." *ACM Transactions on Security and Privacy*.
3. Wired Article: "Researchers Have Ranked AI Models Based on Risk—and Found a Wild Range".
4. Brundage, M., et al. (2022). "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation." *ArXiv Preprint*.