Research Questions

Hierarchical Explanations for Complex Data

1. Paper: Hierarchical Explanations for Video Action Recognition

Focus: This paper presents a hierarchical prototype-based approach for explaining video action recognition at multiple levels. It emphasizes multi-level explanations that offer high-level summaries and detailed insights.
Relevance: Inspired the need for hierarchical explanations in cybersecurity, especially for SOC analysts who benefit from layered insights that can reveal both general patterns and detailed feature interactions.

2. Feature Interaction Interpretability

Paper: How Does This Interaction Affect Me? Interpretable Attribution for Feature Interactions
Focus: This paper examines how to provide interpretable explanations that account for feature interactions in complex datasets.
Relevance: This study raised awareness of the challenges in handling feature correlations, which is crucial in cybersecurity contexts where multiple variables interact over time. It led to questions about adapting these methods for temporal and spatial dependencies in cybersecurity data.

3. Model-Agnostic Multilevel Explanations

Paper: Model Agnostic Multilevel Explanations
Focus: This research provides a framework for model-agnostic explanations that operate at multiple levels, offering local (instance-level) and global (model-wide) insights regardless of the underlying model.
Relevance: It highlighted the value of multi-level explanations that work across different models and data types, motivating the idea of model-agnostic hierarchical explanations tailored for the cybersecurity domain.

4. Faithfulness in Explanations (NLP Context)

Paper: Towards Faithful Model Explanation in NLP: A Survey
Focus: This survey discusses faithfulness in model explanations—ensuring that explanations genuinely reflect model behavior. It focuses on ensuring that the explanation truly represents what drives the model's decisions, particularly in NLP.
Relevance: This paper underscored the importance of faithfulness in explanations, leading us to consider trustworthy and accurate insights for cybersecurity analysts. It inspired questions around ensuring that explanations are not only interpretable but also actionable and aligned with real-world outcomes.

5. Local-to-Global Explanations for Decision Trees

Paper: From Local Explanations to Global Understanding with Explainable AI for Trees
Focus: This research explores how local explanations for decision trees can be scaled up to provide a global understanding of model behavior, blending instance-specific and model-wide perspectives.
Relevance: This paper highlighted the benefits of combining local and global explanations, particularly for complex datasets like those in cybersecurity, where both instance-level and broad trends are crucial for understanding and mitigating threats.

- How can hierarchical explainability methods improve interpretability and decision-making in Network Intrusion Detection Systems (NIDS) by enabling SOC analysts to access both high-level insights and low-level feature-specific explanations?
    - Goal: This question addresses how to design explanations that are both broad and specific, ensuring that the hierarchy helps analysts understand general patterns and drill down into detailed feature interactions when needed.
- What are the limitations of existing explainability methods (e.g., SHAP, LIME, TCAV) in handling temporal dependencies and feature correlations within cybersecurity datasets, and how can these be addressed to provide more accurate and actionable insights?
    - Goal: This targets specific challenges around temporal data and feature correlation, which are often present in cybersecurity logs. Answering this could improve the real-time usefulness of XAI models in detecting evolving threats.
- How can explainable AI models be adapted or developed to generate insights that are actionable for SOC analysts, rather than merely visualizing model behavior, particularly for anomaly detection and intrusion prevention?
    - Goal: This question emphasizes creating explanations that translate directly into actions, helping analysts take immediate steps rather than just interpreting model outputs.
- What usability challenges arise when SOC analysts use hierarchical explanations in real-world cybersecurity contexts, and how can we measure and optimize the cognitive load, response time, and confidence levels in their decision-making?
    - Goal: This focuses on human-centered evaluation, exploring how hierarchical explanations impact usability and confidence for SOC analysts, especially when handling complex cybersecurity events.
- How robust are hierarchical explanation models in replicating consistent explanations across similar network security events, and what evaluation metrics can be established to measure robustness and reliability for cybersecurity applications?
    - Goal: Robustness is crucial in cybersecurity, where models need to be consistently accurate. This question addresses the evaluation framework, focusing on establishing metrics that measure consistency and reliability in explanations.
- How can hierarchical explanations outperform traditional single-level explainability methods (e.g., SHAP, LIME) in both interpretability and usability for SOC analysts when applied to complex cybersecurity datasets?
    - Goal: A comparative question that encourages testing the effectiveness of hierarchical explanations against more traditional methods, focusing on the unique requirements of cybersecurity.

- In what ways do feature interaction limitations in current XAI models affect critical cybersecurity tasks like malware detection, and how can hierarchical models mitigate these limitations to enhance SOC analysts' response accuracy and speed?
    - Goal: This question aims to explore how feature interactions play a role in cybersecurity and how hierarchical models might better capture these to support accurate and timely responses.