Experiment Design for Human Usability Studies

Objective: Evaluate the effectiveness of XAI tools to real-world applicability and decision-making improvements of cybersecurity professionals.

- **Hypothesis**: XAI tools improve the ability of cybersecurity analysts (Cyber security researchers) to understand model outputs and make informed decisions compared to non-explainable models.

Participants

- Cybersecurity researcher with varied levels of experience (e.g., junior, mid-level, senior) to test if XAI tools impact decision-making across expertise levels.

Survey Questions

Qualitative (Semi-structure) Interview:

- How did your experience with the XAI system compare to using a traditional black-box model without explanations?

- What features or improvements would you suggest making the XAI system more applicable to real-world cybersecurity applications?
- Were the explanations helpful in avoiding mistakes or wrong decisions
- Did the explanations reduce the time it took for you to make a decision compared to previous non-XAI-supported methods?

Quantitative Interview

- **Expertise Level:** How much domain-specific knowledge or training do you feel is needed to effectively understand and utilize the explanations provided by this system?
  (Rate the complexity of the explanations on a scale from 1 to 5)
  (e.g. 1 = very easy to understand, 5 = highly complex)

- **Explanation Type:** How would you categorize the type of explanation provided (e.g., feature importance ranking for text/tabular data or visual heatmaps for image data)?

- **Coherence**: Do the explanations align with your prior knowledge and understanding of the security task? Can you identify any discrepancies between the explanations and your expectations? (Rate the coherence of the explanation on a scale from 1 to 5) (e.g. 1 = completely incoherent, 5 = highly consistent and aligned)

- **Actionability**: How actionable do you find the explanations provided by the system? Do they offer clear next steps or guidelines for making decisions? (Rate the actionability of the explanations on a scale from 1 to 5) (e.g. 1 = not actionable, 5 = immediately actionable)

References:

1. Nadeem, Azqa, et al. "Sok: Explainable machine learning for computer security applications." *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2023.
2. Bhusal, Dipkamal, et al. "Sok: Modeling explainability in security analytics for interpretability, trustworthiness, and usability." *Proceedings of the 18th International Conference on Availability, Reliability and Security*. 2023.