"Human usability" in cybersecurity means making security systems and processes simple and clear for users. This helps ensure that people can use these systems effectively without making mistakes or finding ways around them due to complicated designs. The goal is to create a balance between having strong security and making it easy for users to interact with.

Human usability in Explainable Artificial Intelligence (XAI) systems focuses on how easily and effectively users can interact with these systems. Usability concerns the ease and pleasure with which users can utilize features of XAI systems. The key goals in enhancing human usability in XAI include transparency, ensuring that the AI's decision-making process is understandable; trustworthiness, building confidence in the system's reliability and fairness; actionability, enabling users to make informed decisions based on AI explanations; and efficiency, aiming to reduce cognitive load while making explanations concise and clear.

2. Research Questions

The study addresses three main research questions:

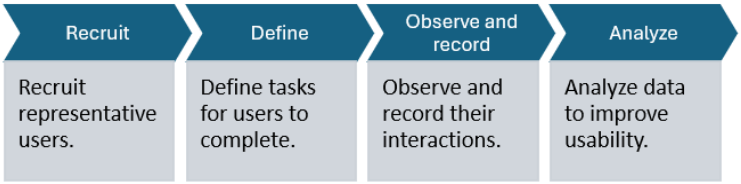
1. How do XAI models falter in managing temporal datasets for security applications?
2. What impact do limitations in feature correlations have on security decisions, such as malware detection?
3. What are the human-centric usability issues with current XAI models in security contexts?

Objective & Hypothesis:

The goal of studying how people use Explainable Artificial Intelligence (XAI) tools is to see if these tools are helpful in real-life situations, especially for people working in cybersecurity. These studies look at whether XAI tools can make it easier for cybersecurity experts to make better decisions, such as finding and understanding unusual activities in their systems.

The main idea being tested is whether XAI tools help people understand the decisions made by AI models better than if the models didn't explain their decisions. This means that with XAI tools, cybersecurity professionals can get clear explanations about why the AI is suggesting something is a threat, which helps them feel more confident and make informed decisions.

4. Usability Study Design (Methodology)



The study's methodology comprises qualitative and quantitative assessments involving cybersecurity researchers and graduate students. Participants compare decision-making processes using traditional black-box models and XAI-enabled models, focusing on anomaly detection tasks.

5. Experimental Setup

In this study, cybersecurity researchers and graduate students were recruited to evaluate the performance of XAI (Explainable Artificial Intelligence) models compared to traditional non-explainable (black box) models in tasks like anomaly detection. XAI models such as SHAP and LIME provide clear explanations for their decisions, which enhances user understanding and trust in the AI's judgments. The study's interview process is structured into two parts: qualitative and quantitative.

Qualitative Interviews (Semi-structured): This section involves open-ended responses, where participants can give short or long answers. Participants discuss their experiences using both model types, offer suggestions for improvements, and comment on how the explanations provided by XAI models help avoid errors and speed up decision-making processes.

Quantitative Interviews: This part employs multiple choice questions, including Likert scales (ranging from 1 to 5) and ranking methods. Participants are asked to rate their understanding of the system's explanations, how well these explanations align with their prior knowledge, and the actionability of the information provided. This quantitative data helps quantify the perceived effectiveness and utility of XAI models in practical cybersecurity applications.

Experiment Design:

1. ***Qualitative (Semi - Structure) Interview Questions:***

- 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? E. g. Time to detect and time to response, Alert closure rate per day

2. **Quantitative Interview Questions:**
   In this segment of the usability study, cybersecurity researchers and graduate students evaluate the Explainable Artificial Intelligence (XAI) models on several dimensions:

   Expertise Level: Participants are asked "how much domain-specific knowledge or training they believe is necessary to effectively understand and utilize the explanations provided by the XAI system? They rate the complexity of the explanations on a scale from 1 to 5, where 1 indicates that the explanations are very easy to understand and 5 signifies that they are highly complex.

   Explanation Type: Participants categorize the type of explanation provided by the XAI models. This could include methods like feature importance ranking for text or tabular data or visual heatmaps for image data, detailing how each explanation type relates to the data processed.

   Coherence: This measure assesses whether the explanations align with the participants' prior knowledge and understanding of the cybersecurity task at hand. Participants also identify any discrepancies between what the explanations show and their expectations. They rate the coherence of the explanation on a scale from 1 to 5, with 1 being completely incoherent and 5 being highly consistent and aligned.

Actionability: This dimension evaluates how actionable the explanations are, meaning how well they help participants make decisions. Questions focus on whether the explanations provide clear next steps or guidelines. Actionability is rated on a scale from 1 to 5, where 1 means the explanations are not actionable and 5 indicates that they are immediately actionable.

User Studies and Results: Example1: Anomaly Detection

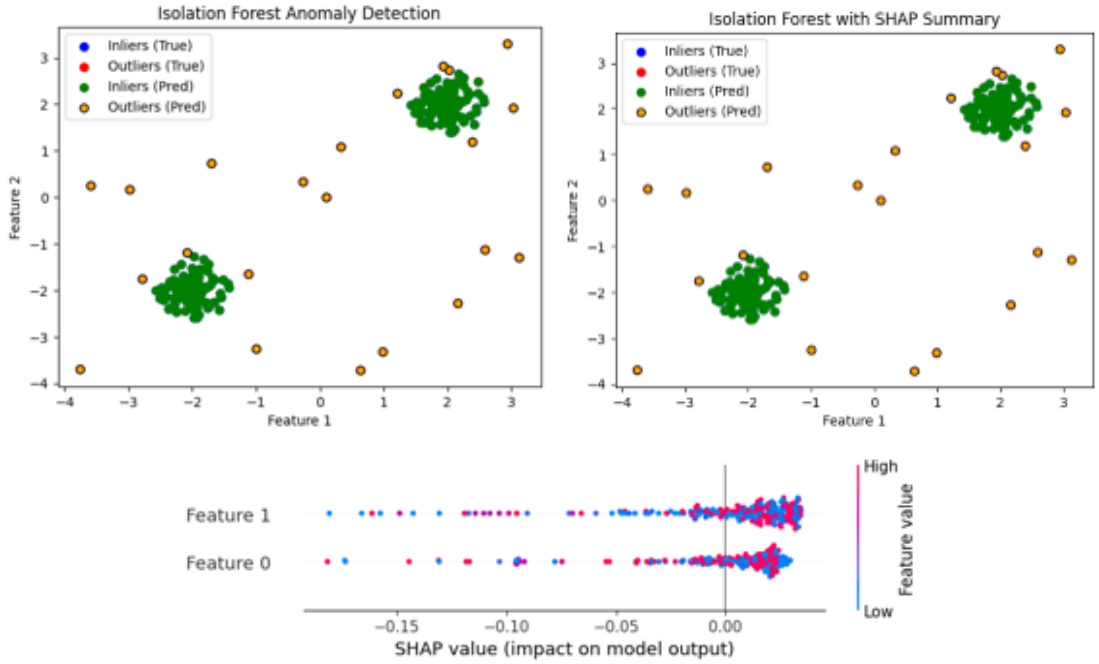


Figure1: Anomaly detection - Example 1: Comparative Visualization of Anomaly Detection Using Isolation Forest and SHAP Value Analysis

The Figure1 illustrates how the Isolation Forest algorithm identifies normal and anomalous data points. In the Blackbox Summary, true normal points are shown in blue, true anomalies in red, predicted normal in green, and predicted anomalies in orange. Two clusters of green points indicate areas where the model accurately predicts normal data, while scattered orange points signal detected anomalies.

The SHAP Summary Plot shows the impact of each feature on the model's decisions. On the x-axis, SHAP values close to zero have low impact, while negative values indicate a strong influence towards labeling anomalies. The y-axis lists features, with color coding indicating the magnitude of feature values, helping to highlight their importance in anomaly detection.

Result: Participants Response(From Example #1): *Qualitative Interview*

After Showing the output of example 1 we captured participants' responses. Following table would be results and comparison of blackbox and XAI model

Qualitative Aspect	Traditional Model (Black Box)	XAI-Enhanced Model (SHAP)
Experience with the system	Lacks transparency; users do not understand why predictions are made.	✔ Provides interpretable explanations for predictions ✔ increasing trust and usability.
Suggested Improvements for Real-World Use	Need to Better handling of false positives/negatives.	✔ Add real-time alerts. ✔ Enable interactive explanations. ✔ Support dynamic thresholds. ✔ Optimize for large/streaming datasets.
Avoiding mistakes or wrong decisions	✔ Prone to errors due to lack of interpretability. ✔ False positives and negatives require manual validation.	Reduces errors by providing clear explanations of feature importance, aligning predictions
Decision-making time	Longer time due to need for manual investigation of predictions.	Faster decision-making as SHAP explanations immediately highlight contributing features.
Time to Detect and Respond	Detection is fast, but response is slower due to lack of actionable insights.	Both detection and response times are improved with interpretable SHAP explanations.
Alert Closure Rate per Day	Lower closure rates due to unresolved alerts and higher false positives requiring manual checks.	Higher closure rates as SHAP explanations clarify alerts, enabling quicker resolution.

Result: Participant Response (From Example #1): Quantitative Interview:

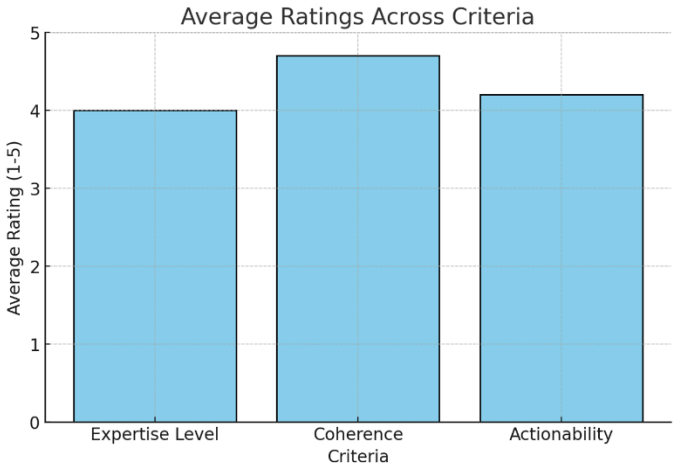
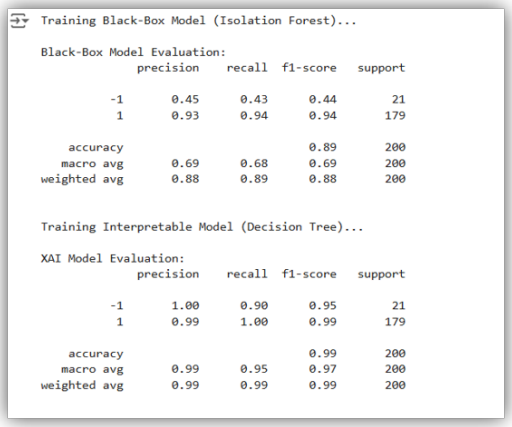


Figure 2: Participants Average Rating on Quantitative Interview (Example 1)

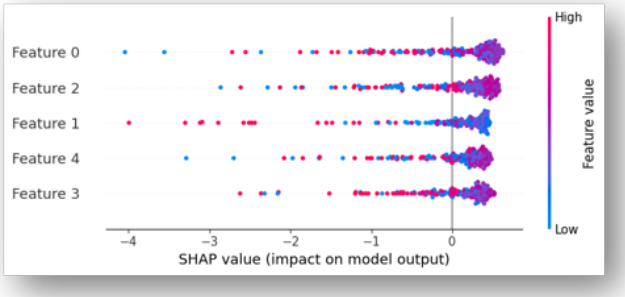
Explanation Type:

Visuals like scatter plots and SHAP summary graphs were categorized as feature importance explanations. Visuals are the most helpful.

Example2: Anomaly Detection



```
Training Black-Box Model (Isolation Forest)...

Black-Box Model Evaluation:
              precision    recall  f1-score   support

          -1       0.45      0.43      0.44        21
           1       0.93      0.94      0.94       179

    accuracy                           0.89       200
   macro avg       0.69      0.68      0.69       200
weighted avg       0.88      0.89      0.88       200


Training Interpretable Model (Decision Tree)...

XAI Model Evaluation:
              precision    recall  f1-score   support

          -1       1.00      0.90      0.95        21
           1       0.99      1.00      0.99       179

    accuracy                           0.99       200
   macro avg       0.99      0.95      0.97       200
weighted avg       0.99      0.99      0.99       200
```



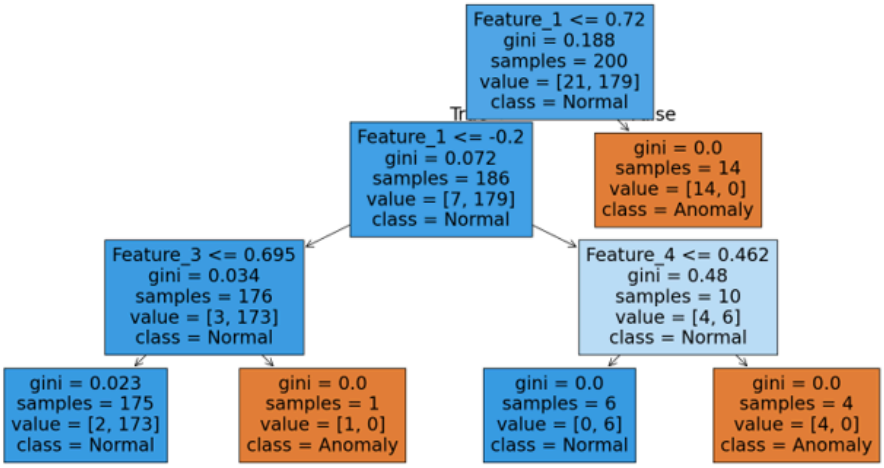


Figure 3 : Comparative Analysis and Decision Tree Visualization for Anomaly Detection Using XAI Models

In Figure 3 summarizes, the black-box model has moderate accuracy in detecting anomalies, with only 45% of its anomaly predictions being correct and correctly identifying 43% of actual anomalies. It performs better with normal instances, achieving 93% precision and 94% recall, with an overall accuracy of 89%. On the other hand, the XAI model using a decision tree is highly effective, correctly predicting every anomaly and missing only 10% of them. It identifies nearly all normal instances accurately, with an overall accuracy of 99%. The SHAP analysis for the XAI model shows how each feature influences predictions, with more important features and their values affecting whether an instance is classified as normal or an anomaly. This model not only performs better but also provides clear reasons for its decisions.

Result: Participants Response(From Example #2): *Qualitative Interview*

After Showing the output of example 2 we captured participants' responses. Following table would be results and comparison of blackbox and XAI model

Qualitative Aspect	Black-Box Model (Isolation Forest)	XAI Model (Decision Tree)
Experience with the system	Hard to understand; users have to trust it blindly.	✔ Easy to understand; users can see how decisions are made.
Suggested Improvements for Real-World Use	✔ Add explanations for decisions. ✔ Show which features are most important.	✔ Add real-time, ✔ easy-to-read explanations. ✔ Handle bigger datasets better. ✔ Use visual tools like SHAP to make it even clearer.
Avoiding mistakes or wrong decisions	✔ More chances of mistakes because decisions aren't explained.	Fewer mistakes because the system shows clear reasons for decisions.
Decision-making time	It takes more time because users have to double-check results.	Quicker because explanations make decisions easy to trust.
Time to Detect and Respond	Slower because users need more time to understand alerts.	Faster because decisions are clear and easy to act on.
Alert Closure Rate per Day	Fewer alerts closed because users spend more time investigating.	More alerts closed because explanations help users act faster.

R

Result: Participant Response (From Example #2): Quantitative Interview:

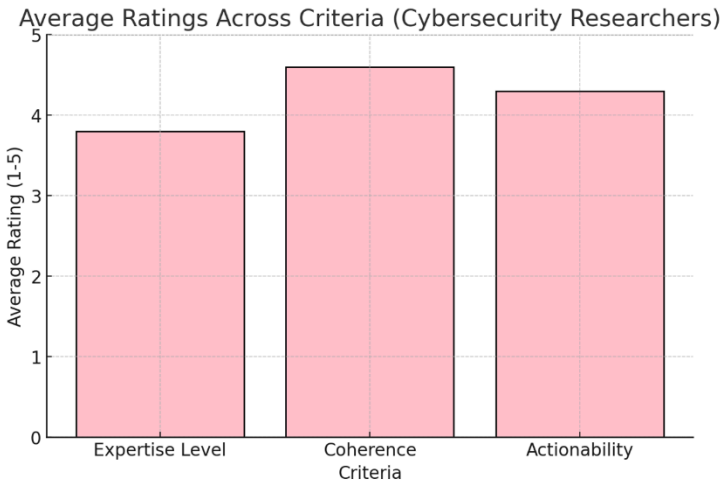


Figure 4: Participants Average Rating on Quantitative Interview (example 2)

Limitation and Future Work

The study on explainable AI in anomaly detection faces several user challenges and limitations. Some users find the explanations provided hard to understand, potentially leading to confusion or feeling overwhelmed by the volume of information. Moreover, the study's findings are based on a small number of participants, suggesting that results might vary with a larger, more diverse user base. Additionally, the user group was limited to cybersecurity student researchers, not including professionals across various experience levels such as junior, mid-level, or senior, which could affect the generalizability and applicability of the study's conclusions to a broader professional context.

Future work should focus on creating more user-friendly explanations, expanding studies to include a larger and more diverse group of participants, and incorporating professionals at various experience levels. These steps will enhance the understandability and generalizability of XAI systems, making them more effective across different user groups and real-world applications.