The Research Monograph Series in Computing, Electrical & Communication Networks
Vol. 1 (01), June 2023, pp. 1–4

CHAPTER 1

INTRODUCTION

Explainable Artificial Intelligence (XAI) is an emerging field that aims to develop AI systems that can explain their decision-making process in a human-understandable way. XAI is particularly relevant in the context of Intrusion Detection Systems (IDS), which are used to detect and respond to cyber attacks.

An IDS uses machine learning algorithms to analyze network traffic and identify anomalies that could indicate a security breach. However, these algorithms can be complex, and their decision-making process may not be transparent or easy to understand for humans. This lack of transparency can be a problem in situations where a human operator needs to make a decision based on the IDS output, such as whether to take action against a potential threat.

XAI techniques can help to address this problem by providing explanations of the IDS output. For example, one approach is to use visualizations to show which features of the network traffic were most important in the decision-making process. This can help a human operator to understand why the IDS made a particular decision and to evaluate whether it is trustworthy.

Another approach is to use natural language explanations to describe the reasoning behind the decision. For example, the IDS could generate a report that explains which rules or heuristics were used to classify the network traffic and how they contributed to the decision.

Artificial intelligence's cognitive capabilities are developed in this sector. Another area where it can demonstrate explainability is in the user interface, as explainable artificial intelligence demands extremely intricate user interactions. And finally, a method of approaching artificial intelligence that can be explained relies heavily on deep learning models (Figure 1).

Intrusion detection for smart grid communication systems is an important aspect of ensuring the security and reliability of the smart grid. Smart grids are modern power grids that rely on communication systems to transmit data and control signals between power generation, distribution, and consumption points. These communication systems

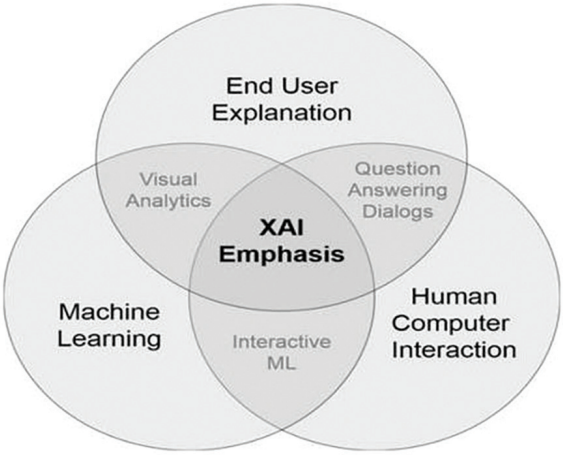

Figure 1. Explainable artificial intelligence (xAI) [1].

are vulnerable to cyber attacks, which can compromise the integrity and reliability of the smart grid.

Intrusion detection for smart grid communication systems involves monitoring the communication network for suspicious activities or behavior that may indicate a cyber attack. This can be done using a variety of techniques, such as signature-based detection, anomaly-based detection, and behavior-based detection.

Signature-based detection involves comparing network traffic against a known set of attack signatures or patterns. This technique is effective at detecting known attacks, but may not be able to detect new or unknown attacks.

Anomaly-based detection involves monitoring the network for behavior that deviates from normal or expected behavior. This technique is effective at detecting previously unknown attacks, but may also generate false positives.

Behavior-based detection involves monitoring the network for behavior that is consistent with known attack patterns or strategies. This technique is effective at detecting advanced or targeted attacks, but may require a high level of expertise to configure and maintain.

Intrusion detection for smart grid communication systems should be combined with other security measures, such as access control, encryption, and network segmentation, to provide a comprehensive security solution. Additionally, regular security audits and updates should be performed to ensure that the intrusion detection system remains effective and up-to-date.

AI-based security controls have the potential to significantly enhance cybersecurity by enabling faster and more accurate threat detection and response. However, it's important to note that AI-based security controls are not a silver bullet and should be used as part of a comprehensive cybersecurity strategy that also includes other security controls, such as firewalls, endpoint protection, and security awareness training for employees.

In Figure 2, we show Machine learning, deep learning, and explainable artificial intelligence in relation to each other.

Explainable Artificial Intelligence (XAI) refers to the ability of an artificial intelligence system to provide understandable explanations of its decision-making process. This is particularly important in the context of intrusion detection for Smart Grid, where it is crucial to be able to trust the system's outputs and understand how they were generated.

Intrusion detection systems for Smart Grid use various machine learning algorithms to detect anomalous behavior that may indicate a cyber attack. However, these algorithms can be opaque and difficult to interpret, which can lead to a lack of trust in the system's outputs. This is where XAI comes in.

By using XAI techniques, the system can provide explanations for its decisions, such as highlighting the specific features of the input data that led to a particular classification or flagging certain patterns as suspicious. These explanations can help users understand how the system is detecting intrusions and make it easier to identify false positives or false negatives.

In addition to enhancing trust management, XAI can also aid in system improvement by identifying areas where the system may be prone to errors or bias. This feedback can be used to refine the system's algorithms and improve its accuracy and performance over time.

The classification of XAI approaches is shown in Figure 3.

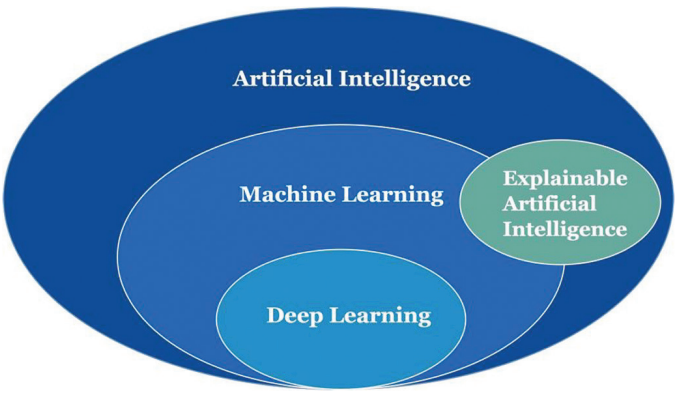

Figure 2. The relationship between AI, Machine learning, Deep learning, and XAI.

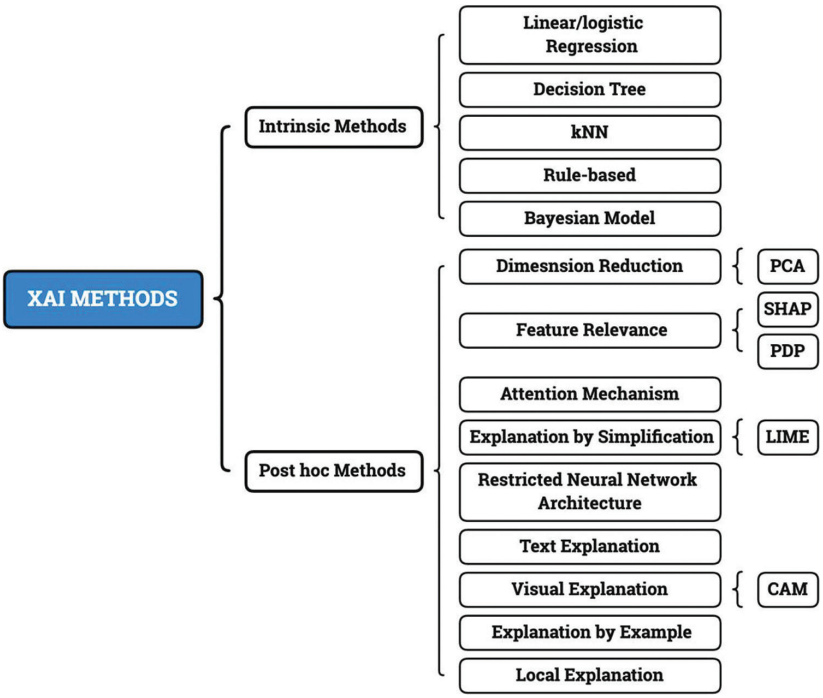

Figure 3. The classification of XAI approaches.