Chapter 9
Enabling Explainable AI in Cybersecurity Solutions

Imdad Ali Shah
Taylor's University, Malaysia

Noor Zaman Jhanjhi
https://orcid.org/0000-0001-8116-4733
Taylor's University, Malaysia

Sayan Kumar Ray
Taylor's University, Malaysia

ABSTRACT

The public needs to be able to understand and accept AI's decision-making if it is to acquire their trust. A compelling justification can outline the reasoning behind a choice in terms that the person hearing it will find "comfortable." A suitable level of complexity is present in the explanation's combination of facts. As AI becomes increasingly complex, humans find it challenging to comprehend and track the algorithm's actions. These "black box" models are built purely from this information. It might be required to meet regulatory standards, or it might be crucial to provide people impacted by a decision the opportunity to contest. With explainable AI, a company may increase model performance and solve issues while assisting stakeholders in comprehending the actions of AI models. Evaluation of the model is sped up by displaying both positive and negative values in the model's behaviour and using data to generate an explanation.

1. INTRODUCTION

While cyber defensive mechanisms appear at the application, network, host, and data levels, cyber security protects computer systems, networks, and data from unwanted access or misuse. The number of connected systems increases as Internet use becomes ubiquitous. Recent technological developments in networking, server capacity, and mobile device accessibility have dramatically increased Internet use.

DOI: 10.4018/978-1-6684-6361-1.ch009

However, the Internet's popularity also encourages hackers to find ways to launch increasingly complex and damaging attacks. Even though there were 0.3 billion more internet users in 2021 than in the year before the number of global cyber attacks climbed by 29% in 2021, as reported by the 2021 Cyber Trends Report A. B. Arrieta 2020. Thousands of people in several U.S. states lost their unemployment benefits and job-search assistance after a cyberattack on a software company in June 2022 B. Goodman and S.2017. This will cause serious societal instability during the COVID-19 pandemic. It is envisaged that the new social and economic norms generated by the COVID-19 outbreak would place an even greater premium on a secure and reliable internet, as stated in a paper by the European Union Agency for Network and Information Security (ENISA) J. Dastin.2017. These statistics and numbers show that cybercriminals and cyberattacks against the Internet and its associated networks and devices have increased significantly in recent years.

A reliable and secure cybersecurity computer system must be set up to protect the confidentiality, availability, and integrity of data during transmission over the Internet. However, traditional cyber defensive mechanisms like signature-based and rule-based approaches need help to keep up with the proliferation of data on the Internet K. Hao 2019. But cybercriminals are persistent in their efforts to stay one step ahead of the law by developing and employing ever-more-advanced methods of attack, such as using artificial intelligence (AI) and other cutting-edge technologies.

However, cyber security researchers and engineers occasionally prioritise accuracy over explainability, resulting in more complex and less intuitive models for users H. Ledford 2019. The General Data Protection Regulation of the European Union has revealed this lack of explainability, which is the ability to understand the reasoning behind an AI algorithmic decision that has a detrimental effect on a person P. Voigt and A. 2017. Therefore, AI must be open and interpretable if we are to have faith in the decisions made by cybersecurity systems. Several approaches have been proposed to improve the human understandability of AI decision-making to meet these needs. These explainable methods, sometimes abbreviated as "XAI," are already being used in various fields of application, including medicine, NLP, and finance M. MacCarthy. 2018. Furthermore, this study aims to examine the many cyber-security applications of XAI.

To verify real-world AI applications, "explainable AI" (XAI) is a collection of advanced methods for explaining AI outcomes to humans. It contrasts with the "black box" concept in machine learning (ML), in which not even the programmers or designers can explain how the AI model arrived at a particular result S. Mohseni 2018. As such, XAI might be viewed as the technological equivalent of the human right to explanation. When there are no regulations, XAI techniques can still be used successfully. Through XAI, end-users trust in AI's ability to make the best decision can be cultivated, leading to a better user experience for the service or product in question. Therefore, the primary objective of XAI is to highlight the explanations or facts on which the actions are based, as well as to explain what tasks have been performed, what activities will be performed, and what future jobs may be. These operations can be used to (1) back up previously held beliefs, (2) cast doubt on previously established data, and (3) come up with novel hypotheses. We may classify ML algorithms into "black box" and "white box." In machine learning, white-box algorithms produce straightforward results even for trained experts to decipher F. K. Dosilovi 2018. On the other hand, black-box systems are so complicated that not even specialists can fully understand them. These three qualities—explainability, transparency, and interpretability—form the backbone of any XAI system. Although explaining ability is crucial, it may not have a universally accepted meaning. In machine learning, it is suggested that explainability can be thought of as a collec-