Chapter 8
Securing the Future of Artificial Intelligence:
A Comprehensive Overview of AI Security Measures

Rinat Galiautdinov
https://orcid.org/0000-0001-9557-5250
Independent Researcher, Italy

ABSTRACT

The criticality of protecting artificial intelligence (AI) systems from malicious attacks has become increasingly paramount as they become ubiquitous in our daily lives. This chapter delves comprehensively into various dimensions of AI security, including detecting and preventing attacks, identifying vulnerabilities, ensuring privacy, establishing robust machine learning models, mitigating fraud, assessing risk, monitoring security, managing access, safeguarding against peripheral device attacks, and providing comprehensive security training. The chapter strongly emphasizes the urgent need to impose rigorous security measures to ensure the reliable and secure functioning of AI systems.

INTRODUCTION

The field of artificial intelligence (AI) is rapidly expanding, and its potential to transform many aspects of our lives is enormous. AI has already made significant strides in industries such as finance, healthcare, and e-commerce, where it is used to identify and prevent fraudulent activities. However, with advancements in AI

DOI: 10.4018/978-1-6684-9324-3.ch008

technology, there is an increased risk of cybercriminals exploiting vulnerabilities in these systems, which can compromise their integrity and lead to data breaches. This is why it is crucial to implement robust security measures to detect and mitigate potential risks.

One significant concern is fraud protection, which is a critical area of AI security. Organizations use AI to detect and prevent fraudulent activities, such as credit card fraud, identity theft, and money laundering. However, these systems are not immune to attacks, and cybercriminals are constantly finding new ways to bypass security measures. Therefore, it is necessary to develop advanced algorithms and machine learning models that can analyze vast amounts of data to identify patterns and anomalies that could indicate fraudulent activity.

In addition to fraud prevention, other areas of AI security include risk assessment, security monitoring, and access management. By employing these measures, organizations can safeguard their AI systems against potential threats and ensure that sensitive data remains secure. For instance, security monitoring involves continuously monitoring AI-based applications to identify any suspicious activity that may indicate a security breach. Similarly, access management involves controlling access to AI systems, ensuring that only authorized personnel can access sensitive data.

Another critical aspect of AI security is developing secure machine learning models. This entails implementing data encryption, secure data storage, and secure communication protocols to prevent attacks that could compromise the accuracy and integrity of machine learning models. Secure machine learning models are crucial for maintaining the integrity and trustworthiness of AI systems. Machine learning models that are not secure can be tampered with, leading to erroneous results and decisions.

The latest directions in Application Security based on Artificial Intelligence include developing algorithms to detect and prevent attacks on AI systems such as the injection of malicious data or modification of training data. For instance, Galiautdinov (2020) proposes a method to protect machine learning models from adversarial attacks by detecting and removing malicious data (Galiautdinov R., 2020). Hu et al. (2020) also developed tools that can automatically detect vulnerabilities in AI-based applications.

Another critical area of research is developing methods to protect personal user data used in AI systems. As AI systems become more ubiquitous, the need to protect personal data from unauthorized access becomes increasingly important. Secure data encryption, secure data storage, and secure communication protocols are essential for ensuring that personal data remains private.

Training programs that educate users and developers about security principles in AI-based applications and improve security levels are also vital. Korkala et al.