MODULE PRACTICE

Unexpected and undesired AI behaviors

Not providing a large amount of trained data can lead to unexpected and undesired AI behaviors.

Inference attack

Inference attacks can be introduced when a specific connection in an AI system can be unraveled and detected.

Data poisoning

Data poisoning is when trained data has been tainted.

Vulnerabilities in AI Systems

AI systems vulnerabilities are often input handling and logic flaws.

Extraction attack

An extraction attack is when a malicious actor feeds data into a model and tracks how the model manipulates the data.

Evasion attack

Evasion attacks are designed to interrupt AI's ability to perform classification and identification.

Underlying Technology Behind AI System

The underlying technology behind AI systems is software and algorithms.