SIT799 Human Aligned Artificial Intelligence

Pass Task 5.1: Adverse use of AI Quiz

Overview

During week 5, you have been introduced to: The use of AI for malicious purpose; Several real-world examples of the use of AI for malicious purposes; Some solutions to fight against the use of AI for malicious purposes. This quiz gives you a chance to demonstrate your understanding of what you have learned.

Quiz

For each of the following scenarios, indicate what type of adverse use of AI it is. **Please justify your answers**.

1. Developing an AI system that generates someone's video face to fool a face recognition system of a phone.

A: This scenario is an example of using AI for unauthorised access. A face recognition system of a phone is generally used to lock the phone and prevent private data. AI solutions like Deepfakes can be used to create a video face to bluff the phone's face recognition system into granting unauthorised access.

2. Developing an AI system to talk/chat as a person to collect information from his colleagues on a company cyber protection system.

A: This scenario is an example of misusing AI for information gathering. An AI system can be trained to fool a human into giving up sensitive information that might be harmful to both organisation and its employees. By using this adverse AI solution, a person can collect information which might be used to perform potential breaches and attacks on the organisation.

3. Developing an AI system to identify users susceptible to click on a malicious link.

A: This scenario is an example for using AI for gathering information for potential misuse. Once the AI system has identified a group of users that may click on susceptible link, the attacker can use the information to launch multiple attacks such as phishing, installing remote malware and keyloggers etc. This attack is an example of probing where the attackers look for potential targets before launching the actual attack.

4. Developing an AI system to identify users susceptible to share banking information – for instance credit card number.

A: This scenario is also an example for using AI for gathering information for potential misuse. Once the attacker has information of the targets after the AI system has successfully finished its probe, the attacker can launch the attacks to gain unauthorized access into the user's account or use the information for other malicious purposes such as misuse of funds.