# Introductions

AI chatbots are starting to take over the globe rapidly in terms of customer service and in other applications as it is an effective way to aid customers without having to train people. However, there are vulnerabilities that these AI chatbots can have which humans would not have. This report will highlight one vulnerability which is bot repurposing and listing all the way that a chatbot can be vulnerable to this attack. The report will also touch on the ways to prevent these vulnerabilities.

# Ways AI chatbots can be vulnerable to bot repurposing

There are several ways that AI chatbots can be vulnerable to bot repurposing:

Insecure API: If the chatbot's API is not properly secured, it can be accessed and modified by unauthorized users.

Weak authentication: If the authentication mechanism used to access the chatbot's API is weak, it can be easily bypassed by attackers.

Poorly designed intent recognition: If the chatbot's intent recognition mechanism is not robust, attackers can use cleverly crafted messages to trick the chatbot into performing unintended actions.

Lack of input validation: If the chatbot does not properly validate user inputs, it can be vulnerable to attacks that exploit input validation vulnerabilities.

Insufficient testing: If the chatbot is not tested thoroughly, it may have vulnerabilities that can be exploited by attackers.

# Measures to prevent bot repurposing

To prevent bot repurposing, AI chatbot developers should implement security measures such as secure API design, strong authentication, intent recognition that is resistant to attacks, input validation, and thorough testing. Additionally, regular security audits and vulnerability scans should be performed to ensure that the chatbot is secure and up-to-date with the latest security standards.

# Conclusion

As great as AI chatbots are, they are still susceptible to potential attacks such as bot repurposing. These vulnerabilities can be mitigated by applying proper security principles to ensure that AI chatbot has the least number of vulnerabilities that attackers can exploit. These principles include implementing a secure API design, strong authentication, robust intent recognition, input validation, and thorough testing. By doing this, the users of the AI are less likely to experience any weird anomalies that weren’t intended by its creator, which could potentially be hazardous.

Sources:

Comiter, M., (2019). *Attacking artificial intelligence*. *Belfer Center Paper*, *8*.,