# АКТУАЛЬНЫЕ УГРОЗЫ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МАШИННОГО ОБУЧЕНИЯ

И.И. Фролов

*Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», Минск, Беларусь*

С развитием искусственного интеллекта (ИИ) и машинного обучения (МО) появились новые угрозы, связанные с этой технологией. В этом обзоре рассмотрим некоторые из основных угроз и проблем, связанных с развитием ИИ и МО [1].

Атаки на модели МО. Злоумышленники могут использовать различные методы для атаки на модели МО. Например, они могут изменять или подменять входные данные, чтобы система приняла неправильные решения. Это может быть особенно опасно, если МО используется в критических областях, таких как медицина или автономные транспортные системы.

Вредоносные атаки с использованием ИИ. С развитием ИИ возрастает вероятность использования его для разработки и распространения вредоносного программного обеспечения. Злоумышленники могут создавать интеллектуальные атаки, которые способны обманывать системы обнаружения и проникать в защищенные сети.

Уязвимости в алгоритмах и моделях МО. Недостаточно защищенные алгоритмы и модели МО могут стать уязвимыми для атак. Например, злоумышленники могут настроить модель МО таким образом, чтобы она давала неправильные результаты или была подвержена взлому.

Неправильное использование данных [2]. Сбор и использование больших объемов данных для обучения моделей МО могут повлечь за собой проблемы конфиденциальности и нарушение приватности. Если некорректные или недостоверные данные используются при обучении модели, это может привести к искаженным результатам и неправильным выводам.

Этические и социальные вопросы. Развитие ИИ и МО влияет на широкий спектр этических и социальных вопросов. Например, проблема автономных систем, способных принимать решения о жизни и смерти, вызывает серьезные вопросы о нравственности и ответственности.

Генерация фальшивых данных. ИИ и МО могут быть использованы для генерации фальшивых данных, включая фальшивые изображения, тексты или видео. Это может привести к распространению дезинформации, манипуляции или созданию поддельных доказательств.

Атаки на инфраструктуру ИИ. Распределенные системы ИИ требуют значительных вычислительных ресурсов и специализированных инфраструктурных компонентов. Атаки на такую инфраструктуру могут привести к нарушению работы систем и серьезным последствиям для организаций, зависящих от ИИ.

Быть в курсе последних тенденций в области безопасности, внедрять надежные меры безопасности, проводить регулярные оценки рисков и повышать осведомленность пользователей – все это необходимо для эффективного снижения угроз информационной безопасности при использовании искусственного интеллекта и машинного обучения.

**Список литературы**

1. Режимы сбоя в машинном обучении [Электронный ресурс] – Режим доступа: https://learn.microsoft.com/ru-ru/security/engineering/failure-modes-in-machine-learning. – Дата доступа: 01.05.2023.

2. Атаки на искусственный интеллект [Электронный ресурс] – Режим доступа: https://media.kaspersky.com/ru/business-security/attacks-on-artificial-intelligence-whitepaper.pdf. – Дата доступа: 01.05.2023.