УДК 004.8-049.65:339.138

АТАКИ НА ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ИЛИ КАК ОТРАВИТЬ ДАННЫЕ И ОБМАНУТЬ АЛГОРИТМ ДЛЯ ПОДРЫВА ЭКОНОМИКИ КОМПАНИИ

В. С. Кухто

студент, Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Беларусь, kuhtoveronika04@gmail.com

Научный руководитель **Е. Н. Макеева**

старший преподаватель, Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Беларусь, e.makeeva@bsuir.by

В статье рассматривается применение искусственного интеллекта в бизнесе и связанные с этим риски безопасности. Особое внимание уделяется уязвимостям искусственного интеллекта к различным видам атак, таким как отравление данных и их возможным последствиям для компаний. Также описываются современные методы защиты ИИ-систем, направленные на повышение их устойчивости и предотвращение экономических и репутационных потерь.

Ключевые слова: искусственный интеллект; модель обучения; атаки; искажение данных; защита данных.

ATTACKS ON ARTIFICIAL INTELLIGENCE OR HOW TO POISON DATA AND DECEIVE AN ALGORITHM TO UNDERMINE A COMPANY'S ECONOMY

V. S. Kukhta

student, Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus, kuhtoveronika04@gmail.com

Supervisor **E. N. Makeeva**

senior lecturer, Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus, e.makeeva@bsuir.by

The article examines the application of artificial intelligence in business and the associated security risks. Special attention is paid to the vulnerabilities of AI to various types of attacks, such as data poisoning, and their potential consequences for companies. It also describes modern methods of protecting AI systems, aimed at increasing their resilience and preventing economic and reputational losses.

***Keywords*:** artificial intelligence; training model; attacks; data distortion; data protection.

Цифровая революция радикально изменила способ ведения современного бизнеса. Искусственный интеллект (ИИ) стал катализатором трансформации традиционных бизнес-процессов. Сегодня ИИ помогает нам за секунды решать задачи, на которые раньше тратилось несколько часов, дней, месяцев или даже лет. С развитием технологий, корпорации начали активно использовать возможности искусственного интеллекта.

Компании используют искусственный интеллект для решения таких задач, как выявление потребностей клиента, разработка и реализация маркетинговых стратегий, подбор кандидатов, сбор и хранение данных о клиентах, сокращение расходов предприятия, выстраивание пользовательских путей.

Основываясь на экспертные оценки и частоту упоминания различных направлений искусственного интеллекта в профессиональной литературе, можно составить диаграмму (рисунок) распределения по направлениям использования.

Диаграмма распределения по направлениям использования ИИ

Однако важно осознавать ограничения ИИ. Несмотря на впечатляющие достижения, искусственный интеллект не способен полностью воспроизвести сложность человеческого мышления и интуиции.

Современные системы ИИ основываются на машинном обучении и обучаются на большом объеме данных [1]. Эффективность работы таких систем зависит от качества, полноты и достоверности данных.

Атаки на ИИ становятся сегодня все более изощренными и потенциально разрушительными для предприятий. Один из самых распространенных видов атак на искусственный интеллект – метод отравления данных (Data poisoning). Злоумышленник намеренно вносит вредоносные или неточные данные в обучающий набор. Целью этого является обучение модели на искаженной информации, что приведет к некорректным действиям системы. Существуют и другие виды атак: модификация существующих данных (Data Changing), удаление важной информации и ее подмена (Information Substitution).

Совершив атаку на систему ИИ предприятия, злоумышленники могут манипулировать данными кредитной истории клиентов, искажать отчеты о работе предприятия, подменять персональные данные. Успешные атаки на ИИ могут нанести серьезный экономический ущерб компаниям.

Каков же принцип действия для злоумышленника? Он изучает архитектуру и параметры модели ИИ данной компании, чтобы создать входные данные, которые кажутся идентичными исходным, однако приводят к ошибочной классификации или неверному дальнейшему прогнозу и оценке данных [2]. Неверные решения или сбои в работе ИИ могут подорвать доверие клиентов и нанести им ущерб. Компании испытывают финансовые трудности, например, из-за сбоев торговых алгоритмов.

Из-за существующего риска атак на ИИ появились различные способы защиты системы предприятия. Внедряются механизмы для выявления и фильтрации подозрительных и неточных данных [3]. Производится регулярная оценка точности модели искусственного интеллекта для повышения устойчивости модели. Используются методы обучения ИИ, которые делают модели менее восприимчивыми к отравлению данных. Разрабатываются инструменты для автоматического выявления атак в реальном времени.

В заключение необходимо сказать, атаки на ИИ представляют собой серьезную и постоянно растущую угрозу для бизнеса. Для защиты от этих угроз компаниям необходимо инвестировать в надежные системы безопасности. Эффективная защита ИИ требует не только технических решений, но и осведомленности, ответственности и активного сотрудничества между разработчиками, специалистами по безопасности и руководством компании. Защита данных в ИИ-безопасности требует ежедневного внимания, и успех в борьбе с атаками на искусственный интеллект будет зависеть от способности предвидеть, предотвращать и реагировать на новые и возникающие угрозы.

Библиографические ссылки

1. *Архипова Л. И.* Большие данные и искусственный интеллект в бизнесе: развитие и регулирование. Минск : Изд. Белпринт, 2020.

2. *Дубровский Д.* Сознание, мозг, искусственный интеллект. М. : Изд. Стратегия-Центр, 2007.

3. *Бостром Н.* Искусственный интеллект. Возможные пути, стратегии, опасности. М. : Изд. Манн, Иванов и Фербер, 2015.