УДК 004.056

Использование технологии Blockchain для обеспечения целостности и безопасности данных обучающих выборок

Хатенов И.А. (ИТМО)
Научный руководитель –
ассистент факультета безопасности информационных технологий Федоров И.Р.
(ИТМО)

Введение. Развитие машинного обучения и искусственного интеллекта привело к впечатляющим достижениям во многих областях. Однако, одной из главных проблем, с которой сталкиваются исследователи и практики, является недостоверность и ненадежность данных, используемых для обучения моделей. Безопасность и целостность обучающих выборок являются критически важными факторами, которые определяют точность и надежность модели[1]. В данном докладе рассматривается использование технологии Blockchain в контексте обеспечения безопасности и целостности данных.

Основная часть. Возможности применения технологии Blockchain в области обучения моделей:

1. Децентрализация и распределенное хранение данных[2]: технология Blockchain позволяет хранить данные обучающих выборок в децентрализованной сети, что гарантирует их надежность и доступность. Распределенное хранение данных обеспечивает отсутствие единой слабой точки в системе, что повышает безопасность и защищает от хакерских атак и потери данных.
2. Неизменность блокчейн-транзакций и невозможность подделки[3]: блокчейн - это цепочка блоков, которая является неизменной, неподдельной и прозрачной. Каждая добавленная запись становится неотъемлемой частью цепочки, и изменение прошлых записей требует согласия большинства участников сети. В контексте данных обучающих выборок, это означает, что информация о каждой выборке сохраняется и невозможно подделать или модифицировать уже существующие данные. Это обеспечивает надежность и целостность данных.
3. Обеспечение прозрачности и доверия между участниками сети: в блокчейне каждый участник имеет доступ к полной истории транзакций и может проверить подлинность данных. Это создает доверие между участниками и устраняет необходимость доверять централизованным структурам. Использование технологии Blockchain для обучающих выборок позволяет подтверждать источник данных, проверять их достоверность и обеспечивать прозрачность для всех участников.

Примеры применения технологии Blockchain в машинном обучении.

1. Обеспечение целостности и безопасности медицинских данных для обучения моделей[4]: В медицинских исследованиях и разработке новых лекарственных препаратов данные критически важны. Blockchain позволяет обеспечить их целостность и безопасность, учитывая конфиденциальность пациентов и защищая от несанкционированного доступа;
2. Защита финансовых данных при обучении моделей для прогнозирования рынков[5]: финансовые данные являются ценными и конфиденциальными. Blockchain при использовании для обучения моделей позволяет сохранить целостность и безопасность финансовых данных, минимизирует риски и обеспечивает доверие между участниками;
3. Повышение доверия клиентов в сфере маркетинга и анализа данных: в контексте сбора и анализа данных клиентов Blockchain может создать доверие, предоставляя полную прозрачность и подтверждение подлинности данных. Клиенты могут быть уверены в безопасности и конфиденциальности своих личных данных.

Ограничения и вызовы применения технологии Blockchain в обучении моделей:

1. Сложности масштабирования и производительности с использованием Blockchain[6]:

процессы синхронизации и проверки данных в блокчейн-сети могут быть медленными и требовать больших вычислительных ресурсов;

2. Проблемы конфиденциальности и защиты персональных данных: хранение и обработка персональных данных в блокчейне вызывает вопросы о конфиденциальности. Необходимо разработать методы, обеспечивающие защиту личной информации.

Выводы. Проведен анализ использования технологии Blockchain для обеспечения данных, рассмотрены сильные и слабые стороны данного подхода при работе с обучающими выборками.

Список использованных источников:

1. Integrity, Accuracy, Consistency: 3 Keys to Maintaining Data Quality in Machine Learning // «Label studio» [Электронный ресурс] URL: https://labelstud.io/blog/integrity-accuracy-consistency-3-keys-to-maintaining-data-quality-in-machine-learning/ (01.01.2024)
2. Сатоши Накамото, Bitcoin: A Peer-to-Peer Electronic Cash System // bitcoin [Электронный ресурс] URL: https://bitcoin.org/bitcoin.pdf/ (02.01.2024)
3. Zhenpeng Liu et al. Data Integrity Audit Scheme Based on Quad Merkle Tree and Blockchain//2022 IEEE symposium on security and privacy (SP). – IEEE, 2023.
4. Хлопотов Р. С. Технологии защиты конфиденциальных медицинских данных в информационной системе врача-иммунолога // Журнал «Известия Тульского государственного университета. Технические науки». – 2023. – С. 31–40.
5. Апатова Н. В., Королев О. Л. Финансовая безопасность и технологии блокчейн // Журнал «Научный вестник». – 2017 – С. 35–40.
6. Абдулжалилов А.З. Методы и стратегии масштабируемости блокчейн-технологий: анализ, сравнение и перспективы // Журнал «Вестник науки» . – 2023 – С. 625–633.