УДК 004.89

ВНЕДРЕНИЕ ДОВЕРЕННЫХ ВЫЧИСЛЕНИЙ В БИБЛИОТЕКИ ГЛУБОКОГО ОБУЧЕНИЯ

Власова И.С., студент гр. 24.М02-мкн, I курс
Санкт-Петербургский государственный университет, г. Санкт-Петербург

С развитием технологий искусственного интеллекта и ростом объёмов обрабатываемых данных проблема обеспечения безопасности и приватности вычислительных процессов становится всё более актуальной. Особенно это касается библиотек глубокого обучения, которые используются для создания и обучения моделей, работающих с чувствительной информацией, такой как медицинские данные, финансовая аналитика или персональные профили пользователей. Внедрение доверенных вычислений в эти библиотеки представляет собой важное направление, позволяющее гарантировать целостность, конфиденциальность и надёжность обработки данных в условиях потенциальных угроз и атак. В данной статье рассматриваются основные подходы и технологии, способствующие интеграции доверенных вычислений в популярные фреймворки глубокого обучения [1].

Доверенные вычисления подразумевают использование аппаратных и программных средств, которые обеспечивают выполнение вычислительных задач в изолированной и защищённой среде, гарантируя, что данные и код не будут изменены или раскрыты третьим лицам. В контексте библиотек глубокого обучения это достигается через интеграцию с технологиями Trusted Execution Environment (TEE), такими как Intel SGX, ARM TrustZone или AMD SEV, которые создают аппаратные изоляционные зоны, защищённые от внешних вмешательств. Это позволяет безопасно выполнять критичные операции – от обработки исходных данных до обучения и инференса модели – без риска утечки или модификации.

Кроме аппаратных решений, важную роль играют программные методы, направленные на обеспечение доверия. К ним относятся криптографические протоколы, например, мультипартийные вычисления (MPC) и гомоморфное шифрование, позволяющие выполнять операции над зашифрованными данными. Интеграция таких методов в библиотеки глубокого обучения позволяет организовать распределённое обучение и инференс, сохраняя при этом конфиденциальность данных каждого участника. Это особенно актуально для сценариев с федеративным обучением, где данные остаются локальными,

а обмен происходит лишь посредством защищённых агрегированных параметров [2].

При внедрении доверенных вычислений в библиотеки глубокого обучения также необходимо учитывать вопросы производительности и масштабируемости. Аппаратные изоляционные среды и криптографические протоколы, как правило, вносят дополнительные вычислительные затраты и задержки. Поэтому архитектура и реализация таких механизмов должны быть оптимизированы для минимизации влияния на скорость обучения и точность моделей, сохраняя при этом высокий уровень безопасности. Важным элементом является поддержка гибкой настройки, позволяющей выбирать компромисс между производительностью и уровнем защиты в зависимости от требований конкретного приложения [3].

Для успешной интеграции доверенных вычислений в библиотеки глубокого обучения необходима тесная координация с экосистемой инструментов разработки и развертывания, включая системы контейнеризации, оркестрации и мониторинга. Это обеспечивает удобство использования и совместимость с существующими рабочими процессами, а также позволяет реализовать автоматическое управление жизненным циклом моделей с учётом требований безопасности [4].

Немаловажным аспектом является обеспечение прозрачности и аудита процессов, выполняемых в доверенных вычислительных средах. Механизмы логирования и отчётности позволяют отслеживать действия и изменения, гарантируя возможность проверки корректности и соответствия установленным политикам безопасности. Это особенно важно в условиях регулируемых отраслей, где требования к аудиту и контролю постоянно ужесточаются.

Таким образом, внедрение доверенных вычислений в библиотеки глубокого обучения представляет собой комплексную задачу, требующую интеграции аппаратных и программных технологий, оптимизации производительности и обеспечения совместимости с существующей инфраструктурой. Развитие данного направления способствует повышению уровня безопасности и приватности ИИ-систем, расширяет возможности применения глубокого обучения в чувствительных областях и формирует основу для доверенной экосистемы искусственного интеллекта будущего [5].

Список литературы:

1. Sabt, M. Trusted Execution Environment: What It Is, and What It Is Not / M. Sabt, M. Achemlal, A. Bouabdallah // IEEE Trustcom/BigDataSE/ISPA. – 2015. – P. 57–64. – DOI 10.1109/Trustcom.2015.357.

2. Hunt, T. Ryoan: A Distributed Sandbox for Untrusted Computation on Secret Data / T. Hunt, Z. Zhu, Y. Xu [et al.] // USENIX Symposium on Operating Systems Design and Implementation. – 2016. – P. 533–549.

3. Trāmèr, F. Slalom: Fast, Verifiable and Private Execution of Neural Networks in Trusted Hardware / F. Trāmèr, D. Boneh // International Conference on Learning Representations. – 2019. – P. 1–18.

4. Ohrimenko, O. Oblivious Multi-Party Machine Learning on Trusted Processors / O. Ohrimenko, F. Schuster, C. Fournet [et al.] // USENIX Security Symposium. – 2016. – P. 619–636.

5. Costan, V. Intel SGX Explained / V. Costan, S. Devadas // IACR Cryptology ePrint Archive. – 2016. – Report 2016/086. – 118 p.