УДК 004.056

КОНЦЕПЦИЯ АДАПТИВНОЙ СИСТЕМЫ ДЛЯ СОВМЕСТНЫХ КОНФИДЕНЦИАЛЬНЫХ ВЫЧИСЛЕНИЙ

Щетинин Д.С. (Университет ИТМО)
Научный руководитель – кандидат технических наук, Менщиков А.А.
(Университет ИТМО**)**

Введение. Новые результаты в области глубинного обучения способствуют повышению интереса исследователей и крупных компаний из разных сфер. Уже сейчас существует множество примеров успешного применения моделей машинного обучения (МО) и нейронных сетей для извлечения прибыли, что безусловно привлекает злоумышленников. Последние успехи свидетельствуют о прямой зависимости качества конечных моделей от размера обучающей выборки. Этот факт обостряет и без того серьезную проблему постоянного недостатка качественных данных для обучения. Для преодоления этой проблемы, современные крупные компании, ведущие бизнес в одной или смежных областях, могут объединить свои данные, для получения большой, релевантной выборки. Однако, в этой ситуации на первый план встает вопрос безопасности этих данных. Распространенные подходы обеспечения безопасности, такие как анонимизация, зашумление и другие, имеют существенные недостатки [1-3].

Основная часть. Целью данной работы является построение концепции адаптивной системы для совместных конфиденциальных вычислений, обучения моделей МО и нейронных сетей. Общность темы обоснована ее громоздкостью и сложностью. Гипотеза заключается в том, что можно объединить три и более равных стороны, владеющих схожими по признакам данными, для выполнения совместных конфиденциальных вычислений, обучения моделей МО и нейронных сетей. Адаптивность концепции заключается в том, что конечный набор этапов для обучения модели и обеспечения безопасности этого процесса формируется в зависимости от оценки, которая основывается на имеющихся признаках наборов данных и статистических характеристиках, предварительно вычисляемых по индивидуальным и общему наборам. Все вышесказанное должно способствовать оптимальному соотношению безопасности обучения и качества конечной модели.

Выводы. В работе предложена концепция адаптивной системы для совместного конфиденциального обучения моделей машинного обучения и нейронных сетей.

Список использованных источников:

1. Ahuja, Mohit & Belaid, Mohamed-Bachir & Bernabé, Pierre & Collet, Mathieu & Gotlieb, Arnaud & Lal, Chhagan & Marijan, Dusica & Sen, Sagar & Sharif, Aizaz & Spieker, // Opening the Software Engineering Toolbox for the Assessment of Trustworthy AI // CEUR Workshop Proceedings. – 2020. – Vol. 2659. – С. 67 - 70.
2. Prasser, Fabian & Eicher, Johanna & Spengler, Helmut & Bild, Raffael & Kuhn, Klaus. // Flexible data anonymization using ARX-Current status and challenges ahead // Software: Practice and Experience. – 2020. – Vol. 50.
3. Park, Saerom & Kim, Seongmin & Lim, Yeon-sup. // Fairness Audit of Machine Learning Models with Confidential Computing // WWW '22: Proceedings of the ACM Web Conference. – 2022. – С. 3488–3499.

Щетинин Д.С. (автор) _____

Менщиков А.А. (научный руководитель) _____