ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

П Р Е Д В А Р И Т Е Л Ь Н Ы Й
Н А Ц И О Н А Л Ь Н Ы Й
С Т А Н Д А Р Т
Р О С С И Й С К О Й
Ф Е Д Е Р А Ц И И

ПНСТ
(проект)

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ.
Техническая структура федеративной системы машинного обучения

(ITU-T F.748.13 (2021), NEQ)

Издание официальное

Настоящий проект стандарта не подлежит применению до его утверждения

Москва
Российский институт стандартизации
2023

ПНСТ
(проект)

Предисловие

1 ПОДГОТОВЛЕН Научно-образовательным центром компетенций в области цифровой экономики Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова» (МГУ имени М.В.Ломоносова) и Обществом с ограниченной ответственностью «Институт развития информационного общества» (ИРИО) на основе собственного перевода на русский язык англоязычной версии документа, указанного в разделе 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 202 г. № -ст

4 Настоящий стандарт разработан на основе международного документа МСЭ-Т F.748.13 (2021) «Техническая структура федеративной системы машинного обучения» (Recommendation ITU-T F.748.13 (2021), Technical framework for a shared machine learning system)

5 ВВЕДЕН ВПЕРВЫЕ

Правила применения настоящего стандарта и проведения его мониторинга установлены в ГОСТ Р 1.16—2011 (разделы 5 и 6).

Национальный орган Российской Федерации по стандартизации собирает сведения о практическом применении настоящего стандарта. Данные сведения, а также замечания и предложения по содержанию стандарта можно направить не позднее, чем за девять месяцев до истечения срока его действия, разработчику настоящего стандарта по адресу: 119991, Российская Федерация, Москва, Ленинские горы, д. 1 и в национальный орган Российской Федерации по стандартизации по адресу: 123112, город Москва, Пресненская набережная, дом 10, строение 2.

В случае отмены настоящего стандарта соответствующее уведомление будет опубликовано в ежемесячно издаваемых информационном указателе

«Национальные стандарты» и журнале «Вестник технического регулирования». Уведомление будет размещено также на официальном сайте национального органа Российской Федерации по стандартизации в сети Интернет.

© ITU, 2021

© Оформление. ФГБУ «РСТ», 2023

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

Предисловие II
Введение VI
1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
3.1 Термины, определенные в других документах 2
3.2 Термины, определенные в настоящей Рекомендации 2
4 Сокращения и акронимы 3
5 Соглашения 3
6 Общее представление о федеративной системе машинного обучения 4
7 Роли в федеративной системе машинного обучения 5
7.1 Обзор 5
7.2 Провайдер данных 6
7.3 Вычислительная платформа 6
7.4 Получатель результата 7
7.5 Инициатор задачи 7
8 Технические требования к федеративной системе машинного обучения 7
8.1 Основные функциональные требования 7
8.2 Требования к масштабируемости 9
8.3 Требования к надёжности 10
8.4 Требования к совместимости 10
8.5 Требования к производительности 11
8.6 Требования к удобству использования 11
9 Требования к безопасности для федеративной системы машинного обучения 12
9.1 Требования к аутентификации 12
9.2 Требования к управлению доступом 12
9.3 Требования к аудиту безопасности 13
9.4 Требования к безопасности данных 13
9.5 Требования к защите персональных данных 14
10 Техническая архитектура, функциональные компоненты и процедура обработки федеративной системы машинного обучения в централизованном режиме 15
10.1 Техническая архитектура федеративной системы машинного обучения в централизованном режиме 15

10.2 Функциональные компоненты федеративной системы машинного обучения в централизованном режиме 17

10.3 Процедуры обработки федеративной системы машинного обучения в централизованном режиме 18

11 Техническая архитектура, функциональные компоненты и процедура обработки федеративной системы машинного обучения в децентрализованном режиме 20

11.1 Техническая архитектура федеративной системы машинного обучения в децентрализованном режиме 20

11.2 Функциональные компоненты федеративной системы машинного обучения в децентрализованном режиме 21

11.3 Процедуры обработки федеративной системы машинного обучения в децентрализованном режиме 22

Приложение I Варианты использования федеративной системы машинного обучения 24

I.1 Пример использования: Совершенствование модулей для выявления мошенничества в телекоммуникационных сетях с использованием данных из нескольких сетей 24

I.2 Вариант использования для интеллектуального управления кредитованием и рисками 25

I.3 Пример использования для интеллектуального маркетинга 26

Библиография 27

Введение

Данный стандарт определяет роли, технические требования и требования по безопасности для федеративной системы машинного обучения, а также описывает технические архитектуры, функциональные компоненты и процедуры обработки федеративной системы машинного обучения при централизованном и децентрализованном режимах работы.

ПНСТ
(проект)

П Р Е Д В А Р И Т Е Л Ь Н Ы Й Н А Ц И О Н А Л Ь Н Ы Й С Т А Н Д А Р Т
Р О С С И Й С К О Й Ф Е Д Е Р А Ц И И

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ.

Техническая структура федеративной системы машинного обучения

Artificial intelligence. Technical framework for a shared machine learning system

Дата введения – с __-____-____
до __-____-____

1 Область применения

В настоящей Рекомендация определено понятие «федеративной системы машинного обучения», а также определены роли, технические требования и требования по безопасности, технические архитектуры, функциональные компоненты и процедуры обработки федеративной системы машинного обучения при централизованном и децентрализованном режимах работы. В Приложении также описаны варианты использования федеративных систем машинного обучения.

2 Нормативные ссылки

Указанные ниже Рекомендации МСЭ-Т и другие публикации содержат положения, которые, посредством ссылок на них в данном тексте, становятся положениями настоящей Рекомендации. На момент публикации указанные редакции документов были действующими. Рекомендации и другие публикации могут быть пересмотрены; поэтому пользователям данной Рекомендации предлагается изучить возможность применения самых последних редакций перечисленных ниже Рекомендаций и других публикаций. Список действующих в настоящее время Рекомендаций МСЭ-Т регулярно публикуется. Ссылка на документ в тексте настоящей Рекомендации не придаёт ему как отдельному документу статус Рекомендации.

Издание официальное

Нормативные ссылки отсутствуют.

3 Термины и определения

3.1 Термины, определенные в других документах

В настоящей Рекомендации используются следующие термины, определённые в других документах:

3.1.1 машинное обучение (machine learning) [b-ITU-T Y.3172]: Процессы, позволяющие вычислительным системам, которые могут и не быть явным образом на это запрограммированными, понимать данные и извлекать из них знания.

3.1.2 доверенная среда исполнения (trusted execution environment, TEE) [b-ITU-T J.1201]: Защищенная область основного процессора в кабельной телевизионной приставке и телевизоре, поддерживающих функции интегрированной вещательной широкополосной системы (IBB), которая обеспечивает хранение, обработку и защиту чувствительных данных в изолированной и надёжной среде. Такая среда поддерживает изолированное защищённое выполнение авторизованного программного обеспечения системы безопасности, которое позволяет обеспечить сквозную безопасность посредством исполнения в защищённом режиме аутентифицированного кода и обеспечения конфиденциальности, аутентичности, защиты персональных данных, целостности системы и прав доступа к дан.

3.2 Термины, определенные в настоящей Рекомендации

В настоящей Рекомендации определены следующие термины:

3.2.1 федеративное (совместное) машинное обучение (shared machine learning, SML): Парадигма машинного обучения, позволяющая агрегировать принадлежащие ряду сторон данные и обеспечивать многостороннюю защиту персональных данных[1] в ситуациях, когда различные провайдеры данных и вычислительная платформа не доверяют друг другу.

3.2.2 многосторонние вычисления (multi-party computation): Подраздел криптографии, занимающийся созданием методов, позволяющих сторонам совместно вычислять значение функции на основе индивидуально хранимых сторонами частей

[1] Понятие «data privacy protection» переводится здесь и далее по тексту как «защита персональных данных», однако в контексте настоящего документа его в ряде случаев можно интерпретировать шире, включая в него защиту конфиденциальности не относящейся к персональным данным деловой информации – *прим. переводчик*

входных данных, сохраняя при этом конфиденциальность этих частей входные данных[2].

3.2.3 удаленная аттестация (remote attestation): Метод, с помощью которого вычислительный узел (клиент) осуществляет аутентификацию конфигурации своего оборудования и программного обеспечения для удалённого вычислительного узла (сервера). Цель удаленной аттестации – дать возможность одной удаленной системе (запрашивающей доказательства) определить уровень доверия к целостности платформы другой системы (заявляющей о такой целостности).

4 Сокращения и акронимы

В настоящей Рекомендации используются следующие сокращения и акронимы:

API	Application Programming Interface	Интерфейс прикладного программирования
QPS	Queries Per Second	Количество запросов в секунду
SML	Shared Machine Learning	Федеративное машинное обучение
SMS	Short Messaging Service	Служба коротких сообщений (СМС)
TEE	Trusted Execution Environment	Доверенная среда исполнения
WOE	Weight of Evidence	Весомость доказательст

5 Соглашения

В настоящей Рекомендации используются следующие соглашения:

- **Формулировка «требуется, чтобы»** (is required to) означает требование, которое должно строго соблюдаться, и от которого не допускается никаких отклонений, если предполагается заявить о соответствии настоящему документу;
- **Формулировка «недопустимо, чтобы»** (is prohibited from) означает требование, которое должно строго соблюдаться, и от которого не допускается никаких отклонений, если предполагается заявить о соответствии настоящему документу

[2] Для сравнения: «Подмножество методов безопасных вычислений, известное как безопасные многосторонние вычисления, представляет собой метод, при котором стороны могут совместно вычислять некоторое значение на основе индивидуально хранимых частей информации, не раскрывая в процессе обработки данные части друг другу» (ГОСТ Р 59407-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Базовая архитектура защиты персональных данных», п.7.2.4.8, на основе ISO/IEC 29101:2018 (п.8.2.4.9), см. https://protect.gost.ru/document1.aspx?control=31&baseC=6&id=240677) – прим. переводчика

- **Формулировка «рекомендуется» (is recommended)** означает требование, соблюдать которое рекомендуется, но которое не является абсолютно обязательным. Таким образом, для того, чтобы заявить о соответствии настоящей Рекомендации, выполнение такого требования не требуется;
- **Формулировка «не рекомендуется»** (is not recommended) означает требование, соблюдать которое не рекомендуется, но которое явно не запрещено. Таким образом, о соответствии настоящей Рекомендации всё ещё можно заявлять даже в том случае, если такое требование присутствует;
- Термин «**функции**» (functions) определяется как совокупность функциональных возможностей.

6 Общее представление о федеративной системе машинного обучения

В федеративных системах машинного обучения (SML-системах) несколько участников совместно используют зашифрованные данные и/или обмениваются параметрами моделей с целью обеспечить безопасность данных и защиту персональных данных. Чтобы обеспечить наиболее эффективное использование данных, зашифрованные данные каждой стороны и/или предоставленные ею параметры моделей собираются и используются для обучения модели федеративного машинного обучения. Модели федеративного машинного обучения продолжают обучаться в интересах само-оптимизации, а участники или иные лица, авторизованные на доступ к модели, могут вводить информацию для получения результатов или прогнозов на основе совместно используемых значений. Федеративные системы машинного обучения могут, например, применяться (не ограничиваясь ими) в мультимедийных и игровых приложениях (media applications).

Федеративное машинное обучение может использоваться в централизованном и децентрализованном режимах. Централизованный режим - это решение для многостороннего совместного использования зашифрованных данных и обучения на результатах слияния данных в доверенной среде исполнения (TEE). Децентрализованный режим - это решение для совместного использования и обучения несколькими участниками, основанное на безопасных многосторонних вычислениях, при котором осуществляется обмен неоригинальными данными, не раскрывающими персональные данные.

При использовании централизованного режима сбор данных и обучение на них

осуществляются доверенной третьей стороной. Данный режим можно применять в ситуациях, где участники готовы обмениваться данными, не содержащими персональные данные, а стоимость доступа к услугам низкая. Его также можно применять в случаях, когда требуется проведение сложных вычислений. В централизованном режиме могут также поддерживаться любые алгоритмы, развертывание кластера и централизованные вычисления.

При использовании децентрализованного режима применяются методы многосторонних вычислений для обмена данными или параметрами моделей с целью выполнения задач обучения. Данный режим можно применять в ситуациях, в которых участникам требуется строгая защита персональных данных, или же в случаях, когда у участников имеются большие объёмы локальных данных - поскольку в рамках обучения модели машинного обучения стороны обмениваются параметрами моделей, а не исходными данными. Многосторонние вычисления подходят в случае простых вычислений, поскольку они способны поддерживать меньше алгоритмов по сравнению с решением на основе доверенной среды исполнения.

7 Роли в федеративной системе машинного обучения

7.1 Обзор

Роли в федеративной системе машинного обучения и взаимодействие между ними показаны на рисунке 1.

В федеративной системе машинного обучения её участники могут выполнять ряд ролей, включая роли инициатора задачи, провайдера данных, вычислительной платформы и получателя результатов. Следует иметь в виду, что, хотя на рисунке 1 показаны только два провайдера данных, однако при реальном развёртывании решения таких провайдеров данных может быть много. Одна и та же сторона может выступать в качестве исполнителя нескольких ролей - например, один из провайдеров данных может быть инициатором задачи и/или получателем результата.

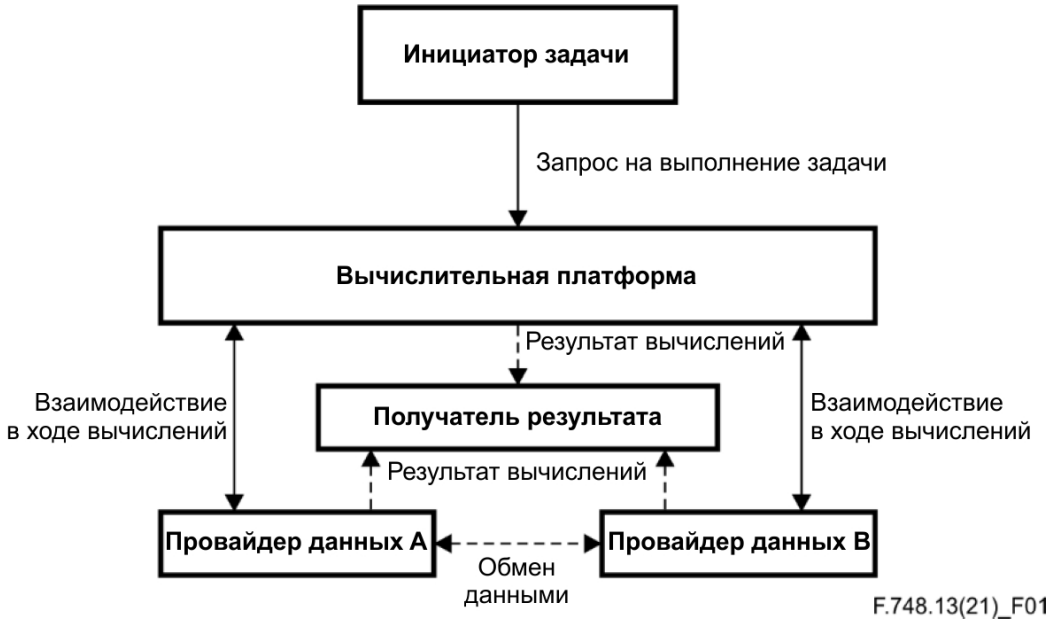

Рисунок 1 - Роли в федеративной системе машинного обучения.

7.2 Провайдер данных[3]

Провайдер данных является владельцем данных и предоставляет эти данные в качестве входных данных вычислительной платформы или другому провайдеру данных. Следует обратить внимание на то, что провайдеры данных также располагают вычислительными ресурсами.

В централизованном режиме данные шифруются и передаются от провайдера данных вычислительной платформе.

В децентрализованном режиме проводится обработка секретов и/или параметров модели, обмен которыми осуществляется между провайдерами данных при координации со стороны вычислительной платформы.

И в том, и в другом режиме обмениваемые данные представляют собой полученные в результате расчётов коэффициенты, а не сами исходные данные. Ни у платформы вычислений, ни у поставщиков данных не должно иметься возможности ввести какую-либо информацию об исходных данных из этих коэффициентов.

7.3 Вычислительная платформа

Вычислительная платформа получает от инициатора задачи запрос на выполнение задачи и отправляет зашифрованный результат получателю результата.

[3] Термин «data provider» переводится здесь как «провайдер данных» с целью отличить его от суб-роли провайдера данных «поставщик данных» (data supplier), упоминаемой в других Рекомендациях МСЭ-Т – *прим. переводчика*

Она взаимодействует с провайдерами данных с целью выполнения вычислительных задач.

В централизованном режиме вычислительная платформа получает зашифрованные данные от провайдеров данных, расшифровывает их и выполняет вычисления в доверенной среде исполнения.

В децентрализованном режиме вычислительная платформа распределяет выполнение вычислительной задачи между участниками и координирует деятельность поставщиков данных по обмену параметрами или секретами между собой и по индивидуальному выполнению вычислительных задач.

Вычислительная платформа объединяет вычислительные алгоритмы

7.4 Получатель результата

В централизованном режиме по завершении вычислительной задачи вычислительная платформа посылает зашифрованный результат получателю результата, который может его расшифровать и получить окончательный результат.

В децентрализованном режиме провайдеры данных посылают результаты вычислений получателю результатов, который объединяет полученные результаты вычислений и получает окончательный результат.

Роль получателя результатов может выполняться и провайдерами данных, и инициатором задачи, и вычислительной платформой.

7.5 Инициатор задачи

Инициатор задачи инициирует вычислительную задачу на вычислительной платформе.

Роль инициатора задачи может выполняться вычислительной платформой или поставщиками данных.

8 Технические требования к федеративной системе машинного обучения

8.1 Основные функциональные требования

8.1.1 Функциональные требования к управлению данными

Функциональные требования к управлению данными для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна поддерживать управление метаданными и их отображение;

- Федеративная система машинного обучения должна поддерживать функцию авторизации использования данных;
- Федеративная система машинного обучения должна поддерживать функцию синхронизации данных (data alignment);
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала аудит поведения при использовании данных;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала хранение данных в зашифрованном виде.

8.1.2 Функциональные требования к управлению алгоритмами

Функциональные требования по управлению алгоритмами для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна поддерживать по крайней мере один широко распространённый алгоритм машинного обучения и прогнозирования, способный обеспечить защиту персональных данных – например, линейную регрессию, логистическую регрессию, древовидную модель, глубокую нейронную сеть, графовую нейронную сеть;
- Федеративная система машинного обучения должна поддерживать функции настройки параметров для получения более качественных результатов обучения;
- Федеративная система машинного обучения должна поддерживать по крайней мере один метод секционирования данных (горизонтальное секционирование или вертикальное секционирование); и рекомендуется, чтобы она поддерживала оба метода секционирования данных одновременно;
- Федеративная система машинного обучения должна поддерживать модели сопротивления атакам получестного противника;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала учитывающую потребности защиты персональных данных функцию, обеспечивающую сведение данных при повторном объединении вертикально секционированных данных;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала безопасные функции анализа данных, такие как

статистический анализ максимальных и минимальных значений, средней дисперсии;

- Рекомендуется, чтобы федеративная система машинного обучения поддерживала функции предварительной обработки данных и оценки эффективности алгоритмов машинного обучения, включая разделение набора данных, заполнение пропущенных значений, метод совокупности доказательств (WoE), оценку корреляции признаков и другие инструменты для повышения показателей производительности моделей машинного обучения.

8.1.3 Функциональные требования к управлению вычислениями

Функциональные требования к управлению вычислениями для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна предоставлять базовые возможности для управления задачами, такие как создание и отмена задач;
- Рекомендуется, чтобы федеративная система машинного обучения предоставляла развитые возможности для управления задачами, такие как мониторинг хода выполнения задач, постановка задач в очередь, анализ исторической информации о задачах;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала возможности для управления распределёнными ресурсами и планирования задач.

8.2 Требования к масштабируемости

Требования к масштабируемости для федеративной системы машинного обучения включают следующее:

- Рекомендуется, чтобы федеративная система машинного обучения обладала хорошей масштабируемостью и допускала добавление новых функциональных компонентов в соответствии с деловыми потребностями;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала возможность доступа пользователей к системе через интерфейс прикладного программирования (API).

8.3 Требования к надёжности

Требования к надежности для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна гарантировать доступность системы и предотвращать отключения из-за некорректных входных данных;
- Федеративная система машинного обучения должна быть способна выполнять автоматическое аварийное восстановление после сбоев (таких, как сбой сервера, сбой жёсткого диска, сбой сети, выключение, перезапуск), включая резервное копирование и восстановление данных и т.д.;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала кластеризацию обучения и прогнозирования, а также поддерживала проведение аварийного восстановления из нескольких компьютерных залов, с целью обеспечения катастрофоустойчивости и возможности переключения сервиса на другой ресурс в случае отказа, тем самым повышая доступность системы;
- Рекомендуется, чтобы федеративная система машинного обучения в интересах повышения надёжности системы поддерживала такие эксплуатационные возможности и возможности для технического обслуживания, как предварительное тестирование и развёртывание по модели «grayscale release»[4], мониторинг и оповещение, а также отслеживание ссылок.

8.4 Требования к совместимости

Требования к совместимости для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна быть совместима с другими алгоритмами машинного обучения;

[4] Для справки: «Модель тестирования «greyscale» (также известная как модель развёртывания «grayscale release» или «canary release») предусматривает выбор конкретной группы людей для проведения испытаний продукта до того, как тот будет официально выпущен; и постепенное расширение количества пользователей-тестеров с тем, чтобы вовремя обнаружить и исправить проблемы. По сути, это предварительное тестирование с целью сбора отзывов пользователей», см. https://www.zentao.pm/blog/On-the-Introduction-of-Grayscale-Testing-1448.html - *прим. переводчика*

- Рекомендуется, чтобы федеративная система машинного обучения поддерживала обратную совместимость в процессе обновления, включая совместимость между системными модулями и совместимость между системами и файлами;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала требования к развертыванию в различных средах, таких как облачная среда, виртуальная машина, физическая машина;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала нормальную работу решения на основе доверенной среды исполнения в различных широко распространённых доверенных средах исполнения.

8.5 Требования к производительности

Требования к производительности для федеративной системы машинного обучения включают следующее:

- Рекомендуется, чтобы федеративная система машинного обучения поддерживала распределённое обучение с тем, чтобы увеличить объем данных, используемых в процессе обучения, и сократить время обучения;
- Рекомендуется, чтобы федеративная система машинного обучения была способна повысить производительность, измеряемую в запросах в секунду (QPS), за счёт горизонтального масштабирования.

8.6 Требования к удобству использования

Требования к удобству использования для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна предоставлять полные инструкции по развёртыванию и эксплуатации, чтобы облегчить пользователям понимание системы, доступ к ней и её использование;
- Рекомендуется, чтобы федеративная система машинного обучения предоставляла платформенный интерфейс для снижения стоимости обучения пользователей и использования системы;
- Рекомендуется, чтобы федеративная система машинного обучения предоставляла платформу разработки алгоритмов, с тем, чтобы на её

основе пользователи могли разрабатывать собственные алгоритмы, отвечающие ограничениям по безопасности;

- Рекомендуется, чтобы федеративная система машинного обучения была спроектирована таким образом, чтобы сократить продолжительность перерывов в работе пользователей во время выполнения обновлений системы.

9 Требования к безопасности для федеративной системы машинного обучения

9.1 Требования к аутентификации

Требования к аутентификации для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна поддерживать функции аутентификации личности пользователей, которые получают доступ к федеративной системе машинного обучения. К пользователям относятся провайдеры данных, системные пользователи и получатели результатов;
- Федеративная система машинного обучения должна поддерживать аутентификацию сертификатов для поставщиков данных и получателей результатов;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала две или более комбинации технологий (таких, как проверка пароля, проверка с помощью электронной почты, проверка с помощью СМС и т.д.) при аутентификации личности пользователей системы.

9.2 Требования к управлению доступом

Требования к управлению доступом для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна поддерживать механизм авторизации для пользователей, чтобы гарантировать авторизованность использования их данных в федеративном машинном обучении;
- Рекомендуется, чтобы федеративная система машинного обучения в полной мере поддерживала роли и системы авторизации. Федеративная

система машинного обучения должна явным образом назначать роли и устанавливать права доступа после входа пользователя в систему;

- Рекомендуется, чтобы федеративная система машинного обучения требовала от пользователя, в соответствии с политикой машинного обучения, повторной аутентификации или повторной активации сеанса работы после продолжительного бездействия.

9.3 Требования к аудиту безопасности

Требования к аудиту безопасности для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна поддерживать функции протоколирования и анализа журналов аудита для операций с данными, выполняемых при обращении пользователей к федеративной системе машинного обучения;
- Рекомендуется, чтобы федеративная система машинного обучения поддерживала функцию отслеживания и аудита истории операций для основных операций в системе.

9.4 Требования к безопасности данных

Требования к безопасности данных для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна поддерживать функцию передачи и хранения важных данных в зашифрованном виде;
- Обрабатываемые данные не должны содержать чувствительные (в т.ч. секретные и конфиденциальные) данные, обработка которых не допускается в соответствии с применимым законодательством;
- Федеративная система машинного обучения должна поддерживать возможность ограничения передачи данных;
- Федеративная система машинного обучения должна поддерживать защищенные протоколы передачи и/или защищенные каналы передачи, чтобы обеспечить безопасность и надёжность каналов передачи данных и предотвратить возможные атаки;
- Федеративная система машинного обучения должна обеспечивать свойства конфиденциальности, целостности и доступности данных и

выборок данных, а также предотвращать незаконное получение данных неавторизованными пользователями;

- Федеративная система машинного обучения должна уничтожать доверенную среду исполнения после завершения выполнения вычислительных задач в централизованном режиме.

9.5 Требования к защите персональных данных

Требования к защите персональных данных для федеративной системы машинного обучения включают следующее:

- Федеративная система машинного обучения должна обеспечивать, чтобы имеющиеся у пользователей персональные данные не попадали в результате утечек к другим провайдерам данных, координаторам или пользователям;
- Федеративная система машинного обучения должна обеспечить наличие механизма, предотвращающего возможность каким-либо образом осуществить точную или приблизительную идентификацию физического лица на основе информации, которой обмениваются провайдеры данных;
- Федеративная система машинного обучения должна предоставлять провайдерам данных возможность использования адекватных методов шифрования, обеспечивающих, что никакие другие стороны, помимо получателя данных, не смогут извлечь персональные данные пользователя из пересылаемых зашифрованных данных - чтобы предотвратить такие возможные атаки, как взлом методом грубой силы, логические атаки и т.д.;
- В рамках защищённого многостороннего вычислительного решения, федеративная система машинного обучения должна обеспечить, что имеющиеся у пользователя персональные данные не покидают локальное хранилище, а различные провайдеры данных обмениваются только случайными числами и/или зашифрованными параметрами.

10 Техническая архитектура, функциональные компоненты и процедура обработки федеративной системы машинного обучения в централизованном режиме

10.1 Техническая архитектура федеративной системы машинного обучения в централизованном режиме

В централизованном режиме работы федеративной системы машинного обучения локальные данные ряда участников обрабатываются, шифруются и передаются в доверенную среду исполнения для обучения федеративной модели. Поскольку данные каждой стороны зашифрованы, их не может увидеть ни платформа, ни другие участники. Когда участники или иные лица имеют авторизацию на доступ к федеративной модели, они могут получить к ней доступ посредством интерфейсов прикладного программирования (API) для ввода исходных данных и получения выходных данных или прогнозов на основе совместно используемых значений.

Перед предоставлением данных участник (провайдер данных) проверяет корректность и правильность работающего в доверенной среде исполнения программного обеспечения посредством удалённой аттестации. После этого участник и доверенная среда исполнения согласовывают механизм шифрования. Затем участник шифрует данные, используя согласованный механизм, и отправляет зашифрованные данные в доверенную среду исполнения. Для шифрования могут применяться механизмы как симметричного, так и асимметричного шифрования.

Чтобы исключить утечку вычисленной модели, обученная модель должна быть зашифрована после завершения всех вычислений.

Техническая архитектура представлена на рисунке 2.

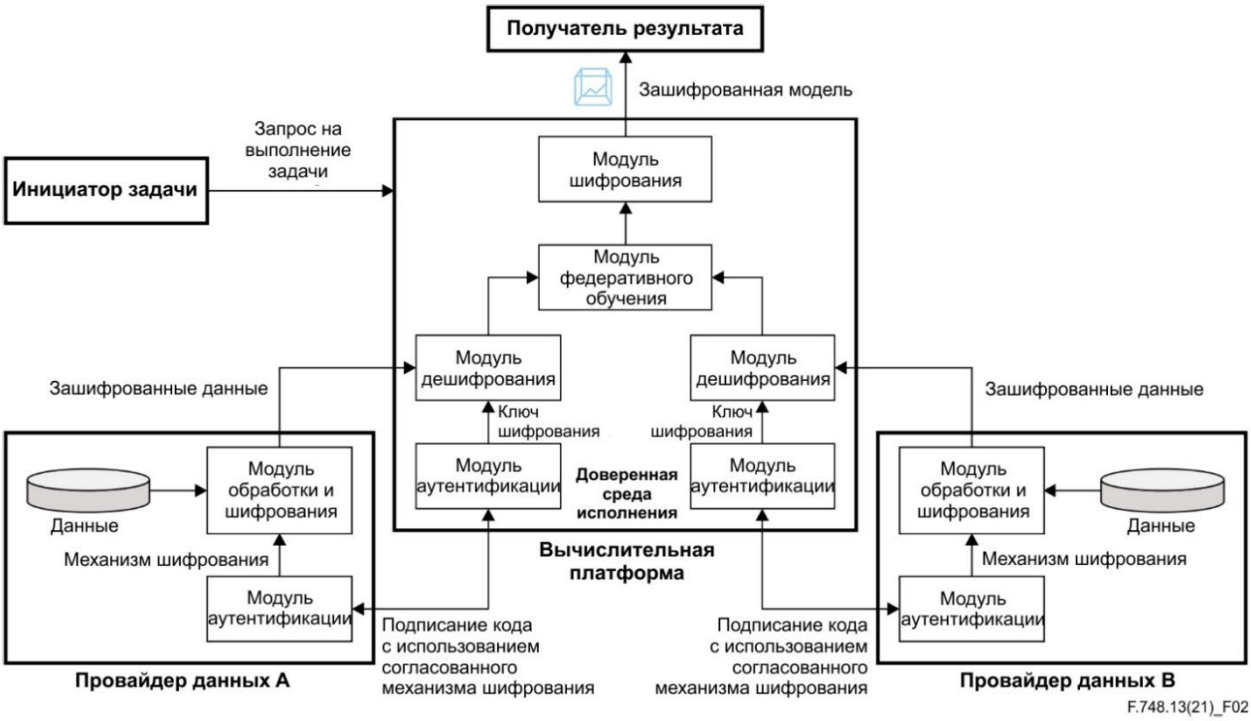

Рисунок 2 – Техническая архитектура федеративной системы машинного обучения в централизованном режиме

Техническая архитектура федеративной системы машинного обучения в централизованном режиме в основном состоит из вычислительной платформы и нескольких провайдеров данных. Вычислительная платформа включает модуль аутентификации, модули шифрования и дешифрования, а также модуль федеративного обучения. У каждого провайдера данных имеются модуль обработки и шифрования, модуль аутентификации и данные. В технической архитектуре модули являются функциональными компонентами.

Данные каждого провайдера данных обрабатываются и шифруются с помощью ключа шифрования, полученного от модуля аутентификации, а затем зашифрованные данные загружаются в доверенную среду исполнения вычислительной платформы. Модуль аутентификации вычислительной платформы расшифровывает зашифрованные данные с помощью ключа дешифрования, а затем отправляет расшифрованные данные в модуль федеративного обучения. Модуль федеративного обучения выполняет операцию федеративного машинного обучения с использованием расшифрованных данных, полученных от нескольких провайдеров данных.

10.2 Функциональные компоненты федеративной системы машинного обучения в централизованном режиме

Основу вычислительной платформы составляют модуль аутентификации, модули шифрования и дешифрования, а также модуль федеративного обучения – все они функционируют в доверенной среде исполнения (TEE).

- Модуль аутентификации: Данный модуль отвечает за согласование механизма шифрования с каждым провайдером данных и за предоставление ключа дешифрования модулю дешифрования. Ключ дешифрования используется для расшифровки зашифрованных данных, загруженных на платформу провайдером данных. Модуль аутентификации вычислительной платформы отвечает за подписание программного кода, работающего в доверенной среде исполнения, и поддерживает проведение проверки программного кода провайдером данных;
- Модуль дешифрования: Данный модуль отвечает за расшифровку загруженных провайдером данных зашифрованных данных с использованием ключа дешифрования. Ключ дешифрования может быть как асимметричным, так и симметричным;
- Модуль федеративного обучения: Данный модуль отвечает за обучение модели на основе расшифрованных данных, полученных от нескольких провайдеров данных, с целью вычисления федеративной модели.
- Модуль шифрования: Данный модуль отвечает за шифрование вычисленной модели после завершения всех вычислений.

Основу архитектуры провайдера данных составляют данные, модуль обработки и шифрования, а также модуль аутентификации. Имеется как минимум два провайдера данных.

- Данные: Данные, предоставленные провайдером данных, могут быть обработаны, зашифрованы и загружены на вычислительную платформу. Зашифрованы могут быть первичные данные, рассчитанные коэффициенты и любыми иные данные, которые необходимо защитить при передаче между участниками и вычислительной платформой;
- Модуль обработки и шифрования: Данный модуль используется для обработки и шифрования данных на основе механизма шифрования,

согласованного с вычислительной платформой, после чего зашифрованные данные загружаются на вычислительную платформу;

- Модуль аутентификации: Модуль аутентификации провайдера данных используется для выполнения удаленной аутентификации провайдера данных и вычислительной платформы, включая согласование с вычислительной платформой механизма шифрования и проверку подписи кода.

Инициатор задачи отправляет на вычислительную платформу запрос на выполнение вычислительной задачи, и инициирует вычислительную задачу.

Получатель результата получает зашифрованный результат от вычислительной платформы и, расшифровав его, получает окончательный результат.

10.3 Процедуры обработки федеративной системы машинного обучения в централизованном режиме

В централизованном режиме как любые провайдеры данных, так и вычислительная платформа могут инициализировать вычислительные задачи, после чего вычислительная платформа создает доверенную среду исполнения. Локальные данные, предоставляемые каждым провайдером данных, могут быть обработаны, зашифрованы и загружены на вычислительную платформу. Платформа расшифровывает полученные от провайдеров данных зашифрованные данные в доверенной среде исполнения, и выполняет обучение модели на основе расшифрованных данных для получения федеративной модели. Обработка данных, шифрование, дешифрование и шаги обучения могут повторяться несколько раз. По завершении вычислительной задачи доверенная среда исполнения уничтожается для обеспечения безопасности и защиты персональных данных.

Рекомендуемые процедуры обработки представлены на рисунке 3:

Рисунок 3 – Процедуры обработки федеративной системы машинного обучения в централизованном режиме

Технический процесс обработки в функционирующей в централизованном режиме федеративной системе машинного обучения включает следующие шаги:

- Шаг 1: Провайдеры данных скачивают программные инструменты с вычислительной платформы и развертывают их;
- Шаг 2: Провайдеры данных выполняют подготовку данных, включая их шифрование и авторизацию;
- Шаг 3: Провайдеры данных загружают зашифрованные данные на вычислительную платформу;
- Шаг 4: Инициатор задачи инициирует вычислительные задачи на вычислительной платформе, включая обучаемую модель и алгоритмы;
- Шаг 5: Вычислительная платформа создает доверенную среду исполнения;
- Шаг 6: Вычислительная платформа расшифровывает зашифрованные данные в доверенной среде исполнения;
- Шаг 7: Вычислительная платформа в доверенной среде исполнения выполняет вычисления с использованием расшифрованных данных для получения результата вычислений;

- Шаг 8: Вычислительная платформа доставляет результат вычисления получателю результата;
- Шаг 9: Вычислительная платформа уничтожает доверенную среду исполнения и данные в ней.

Следует иметь в виду, что описанные выше процедуры обработки являются типовыми. В условиях каждого конкретного развёртывании решения порядок и детали каждой процедуры могут отличаться. Например, инициатор задачи может инициировать вычисление на первом шаге. Провайдеры данных могут многократно загружать зашифрованные данные на вычислительную платформу. Относящиеся к обучению модели данные и обученная модель также могут быть «запечатаны» и сохранены вне вычислительной платформы для последующего повторного использования.

11 Техническая архитектура, функциональные компоненты и процедура обработки федеративной системы машинного обучения в децентрализованном режиме

11.1 Техническая архитектура федеративной системы машинного обучения в децентрализованном режиме

В децентрализованном режиме работы федеративной системы машинного обучения каждой стороне необходимо локально развернуть модуль обучения и передать ему данные. Модули обучения различных провайдеров данных обмениваются параметрами с использованием различных методов шифрования, обеспечивая совместное использование данных в отсутствие обмена первичными данными, что делается для защиты персональных данных. Вычислительная платформа помогает инициировать обновление модуля обучения для каждого участника и координирует взаимодействие между различными сторонами.

Техническая архитектура представлена на рисунке 4.

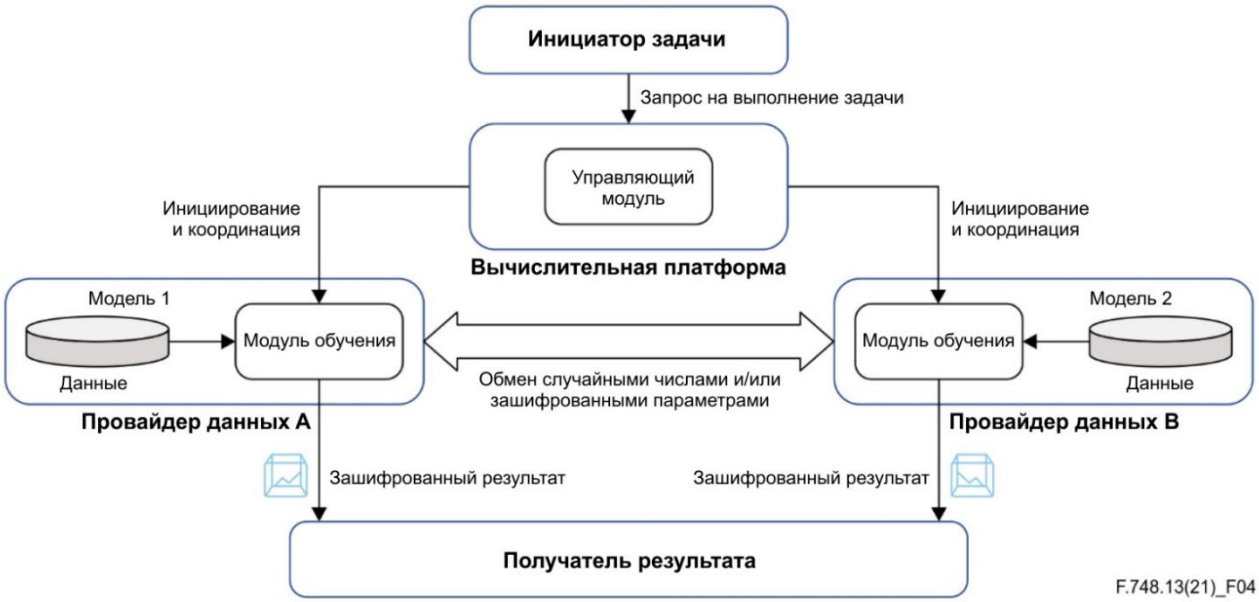

Рисунок 4 – Техническая архитектура федеративной системы машинного обучения в децентрализованном режиме

Техническая архитектура федеративной системы машинного обучения в децентрализованном режиме в основном формируется из вычислительной платформы, получателя результатов, инициатора задачи и нескольких поставщиков данных.

11.2 Функциональные компоненты федеративной системы машинного обучения в децентрализованном режиме

Основу вычислительной платформы составляют управляющий модуль, который распределяет вычислительные задачи между различными провайдерами данных и осуществляет координацию обучающих модулей провайдеров данных. Затем модули обучения провайдеров данных обмениваются между собой случайными числами и/или зашифрованными параметрами с целью выполнения операции федеративного машинного обучения.

Основу архитектуры провайдера данных составляют данные и модуль обучения. Имеется как минимум два провайдера данных.

- Данные: Данные провайдера данных предоставляются локальному обучающему модулю для целей машинного обучения. Провайдер данных обеспечивает отсутствие утечек имеющихся у пользователей персональных данных;

– Модуль обучения: Данный модуль используется для получения задач машинного обучения, выдаваемых вычислительной платформой, и для выполнения операцию федеративного машинного обучения на основе локальных данных, а также случайных чисел и/или зашифрованных параметров, которыми обмениваются между собой провайдеры данных.

Инициатор задачи отправляет на вычислительную платформу запрос на выполнение задачи.

Получатель результата получает зашифрованный результат от провайдеров данных и, расшифровав его, получает окончательный результат.

11.3 Процедуры обработки федеративной системы машинного обучения в децентрализованном режиме

В децентрализованном режиме провайдеры данных выполняют операции федеративного обучения, обмениваясь на основе установленного протокола многосторонних вычислений информацией, которая не приведёт к утечке имеющихся у пользователей персональных данных.

Рекомендуемые процедуры обработки для децентрализованного решения представлены на рисунке 5.

Рисунок 5 – Процедуры обработки федеративной системы машинного обучения в децентрализованном режиме

Процедуры технической обработки в функционирующей в децентрализованном режиме федеративной системе машинного обучения включает следующие шаги:

- Шаг 1: Каждый провайдер данных скачивает программные инструменты с вычислительной платформы;
- Шаг 2: Каждый провайдер данных развертывает скачанные инструменты и подготавливает локальные данные;
- Шаг 3: Инициатор задачи инициирует вычислительные задачи через вычислительную платформу;
- Шаг 4: Вычислительная платформа координирует вычислительные задачи, передавая их на исполнение провайдерам данных;
- Шаг 5: Провайдеры данных используют локальные данные для вычислений;
- Шаг 6: Провайдеры данных обмениваются случайными числами и/или зашифрованными параметрами посредством множественных взаимодействий на основе протокола многосторонних вычислений;
- Шаг 7: Провайдеры данных выполняют расчёт плотных слоёв нейронной сети и получают соответствующий результат;
- Шаг 8: Провайдеры данных передают результаты расчёта плотных слоёв нейронной сети получателю результата;
- Шаг 9: Получатель результатов объединяет результаты расчётов плотных слоёв и получает окончательный результат.

Приложение I

Варианты использования федеративной системы машинного обучения

(Данное Приложение не является неотъемлемой частью настоящей Рекомендации.)

Федеративное машинное обучение может использоваться в сервисах различных типов, которым необходимо совместное использование данных для обучения федеративных моделей, и которым также требуется обеспечить безопасность данных.

В настоящем Приложении описываются варианты использования федеративной системы машинного обучения с целью проиллюстрировать их концепцию и техническую архитектуру.

I.1 Пример использования: Совершенствование модулей для выявления мошенничества в телекоммуникационных сетях с использованием данных из нескольких сетей

Своевременное и автоматическое выявление и предотвращение мошенничества в сфере электросвязи является полезной и важной задачей. Во время общения друг с другом мошенники и их потенциальные жертвы могут находиться как в одной и той же, так и в разных сетях связи. Таким образом, для обучения модулей распознаванию мошенничества в телекоммуникационных сетях требуется тесное сотрудничество между несколькими сетевыми операторами, - однако в соответствии с действующим законодательно-нормативным регулированием данные сетей связи являются защищёнными и отделенными друг от друга, и раскрывать их и обмениваться ими нельзя.

Режим обучения федеративного машинного обучения может помочь преодолеть ограничения на доступ к данным нескольких сетей. В этом режиме обучения сети связи хранят у себя свои первичные данные и лишь выгружают необходимую обезличенную информацию (например, метки данных и индекс, а не сами данные) на платформу управления моделью (см. рисунок I.1). Без ограничения общности платформа управления моделью может использовать обезличенную информацию для организации выполнения задач обучения модуля на локальных обучающих платформах соответствующих сетей связи. Задачи обучения модуля

выполняются исполнителями обучения модуля на локальных обучающих платформах, с использованием локальных первичных данных. По ходу обучения локальные обучающие платформы может обмениваться и синхронизировать параметры обучаемых модулей с параметрами на платформе управления моделью.

Обученные модули для обнаружения мошенничества в телекоммуникационных сетях могут быть развернуты и выполняться в нескольких соответствующих сетях связи.

В данном случае сетевые операторы могут использовать механизмы федеративного машинного обучения и соответствующие системы для обучения и выполнения модулей для выявления мошенничества в телекоммуникационных сетях без прямого обмена первичными данными.

Общая схема выявления мошенничества в телекоммуникационных сетях с использованием данных из нескольких сетей представлена на рисунке I.1.

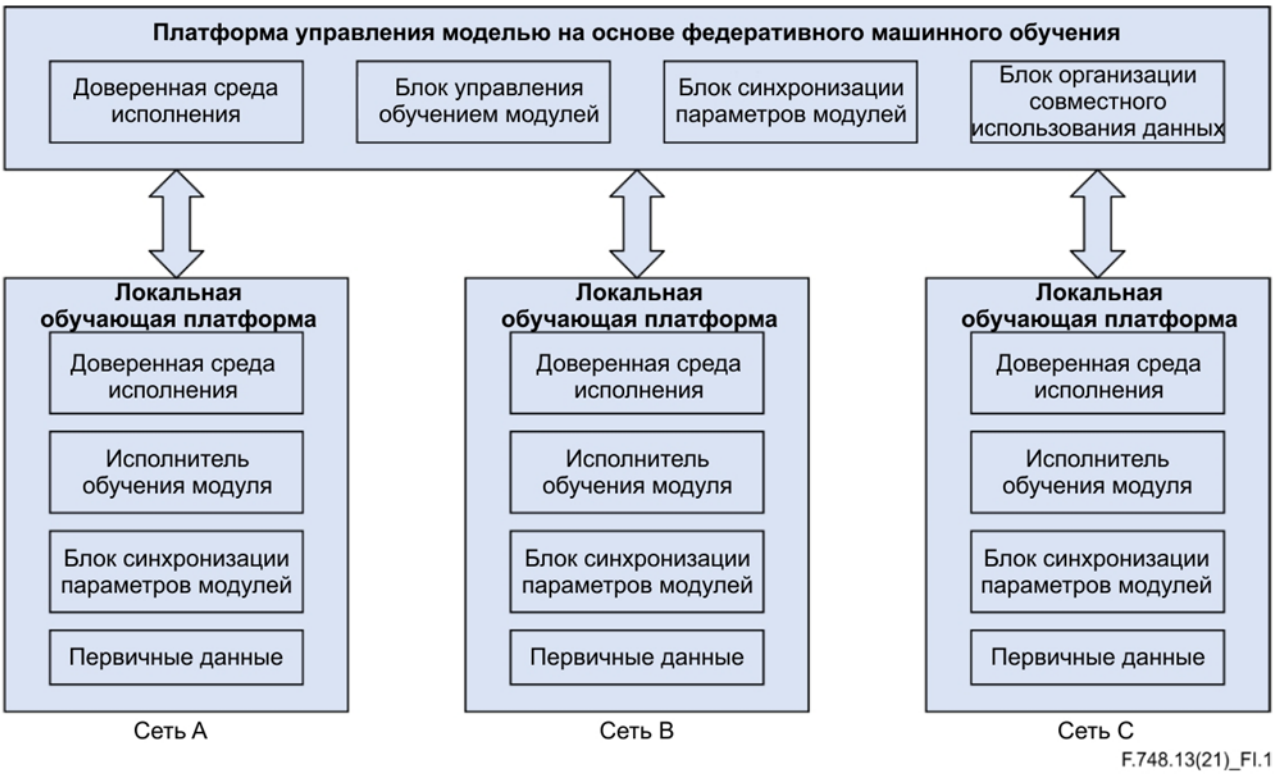

Рисунок I.1 – Общая схема выявления мошенничества в телекоммуникационных сетях с использованием данных из нескольких сетей

I.2 Вариант использования для интеллектуального управления кредитованием и рисками

Федеративное машинное обучение можно использовать в сценариях интеллектуального управления кредитованием в финансовой сфере.

Зашифрованные данные кредитной истории пользователей собираются у различных сторон с целью обучения федеративной модели управления рисками, которое осуществляет доверенная третья сторона. Таким образом, провайдеры данных, такие как банки, финансовые учреждения и другие взаимосвязанные стороны, могут совместно использовать федеративную модель для проверки кредитоспособности физических лиц при определении суммы кредита. Технология федеративного обучения способствует интеллектуальному проведению процесса проверки и помогает снизить затраты на онлайн-проверку кредитоспособности сотрудниками-людьми. Кроме того, федеративное машинное обучение в таком сценарии может поддерживать авторизацию на стороне клиента и защиту персональных данных с целью лучшего управления рисками.

I.3 Пример использования для интеллектуального маркетинга

Федеративное обучение может обеспечить точную стратегию в отношении активов клиентов и защищённую среду обмена информацией, которая способствует выявлению рисков. Например, в сценарии автострахования федеративное машинное обучение может значительно улучшить дифференцированный подход в автостраховании. Это может помочь страховым компаниям разработать лучшие стратегии действий до того, как клиенты приобрели страховку, при наличии согласия клиентов. Благодаря использованию информации о владельцах автомобилей, клиентов можно разделить на группы с разным уровнем риска. Кроме того, может быть проведен точный анализ репутации и рисков владельцев автомобилей, чтобы реализовать точную стратегию в отношении активов клиентов.

Библиография

[b-ITU-T J.1201] Рекомендация МСЭ-Т J.1201 (2022) «Операционная система «умного» телевидения – Функциональные требования» (Functional requirements of a smart TV operating system), см. https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-J.1201-202201-I!!PDF-E&type=items [5]

[b-ITU-T Y.3172] Рекомендация МСЭ-Т Y.3172 (2019) «Основа архитектуры машинного обучения в будущих сетях, включая IMT-2020» (Architectural framework for machine learning in future networks including IMT-2020), см. https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.3172-201906-I!!PDF-E&type=items

[5] Ссылка на устаревшую редакцию 2019 года заменена ссылкой на действующую редакцию 2022 года – *прим. переводчика*

ПНСТ
(проект)

УДК 004.01:006.354 ОКС 35.020

Ключевые слова: информационные технологии; искусственный интеллект; зашифрованные данные; многосторонние вычисления; федеративное машинное обучение; доверенная среда исполнения.

Руководитель разработки
Председатель совета директоров
ООО «Институт развития
информационного общества» Ю. Е. Хохлов

Исполнитель А. А. Храмцовская