



Платформа доступа к моделям ИИ в локальной среде

Использование моделей генеративного ИИ в защищенном локальном режиме для автоматизации рутинных задач организации

Платформа разворачивается в инфраструктуре организации и представляет собой комплексный набор инструментов и сервисов, включающий: базовый или специально подобранный под требования заказчика набор больших генеративных моделей, интерфейсы для взаимодействия с пользователями, интеграцию платформы с существующими корпоративными данными организации.

ЦЕННОСТНОЕ ПРЕДЛОЖЕНИЕ

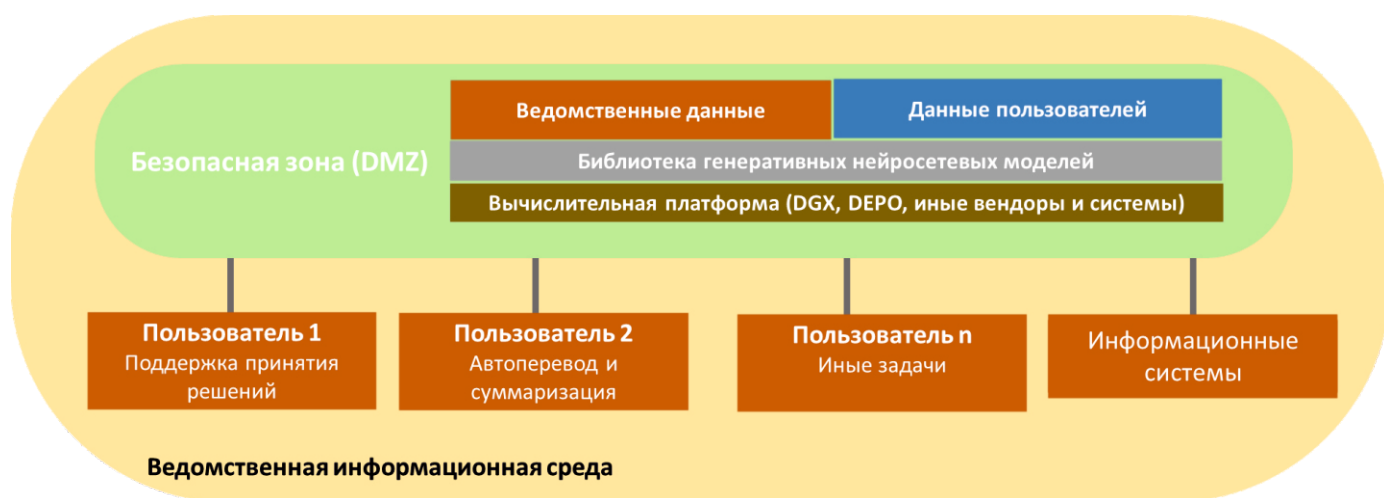
Локализованная в инфраструктуре организации платформа доступа к моделям ИИ для автоматизации выполнения рутинных задач и обеспечения безопасной работы с корпоративной информацией, исключающей возможность несанкционированного доступа к данным со стороны третьих лиц

ОБЛАСТИ ПРИМЕНЕНИЯ

- Органы власти (министерства и ведомства)
- Оборонно-промышленный комплекс
- Интеграторы ИТ-решений

КОНКУРЕНТНЫЕ ПРЕИМУЩЕСТВА

- Конфиденциальность корпоративной информации: исключение риска утечки корпоративных данных
- Гибкость и масштабируемость: индивидуальная адаптация платформы под оборудование и задачи заказчика, включая выбор оптимальных моделей ИИ и настройку интерфейса
- Минимизация влияния отказов оборудования на процесс вычислений
- Отсутствие скрытых платежей и оплаты за токены



Архитектура платформы

СТАДИЯ РАЗРАБОТКИ

- Разработана архитектура платформы, включая набор моделей ИИ, баз данных, фреймворков
- Проведены работы по выбору и интеграции генеративных моделей ИИ в комплексное программное решение
- Разработан пользовательский интерфейс доступа к моделям ИИ
- Проведено тестирование возможности интеграции платформы с корпоративными информационными системами

ОСНОВНЫЕ ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

- Технические характеристики платформы:
 - операционная система: Linux
 - фреймворк управления ресурсами и моделями ИИ: vLLM
 - мультимодальность: работа с любыми моделями ИИ
 - модуль интеграции с корпоративными данными: векторные и реляционные БД, модуль графовой БД
 - пользовательские веб-сервисы для доступа к ресурсам моделей
- Базовые требования к техническим характеристикам ИТ-инфраструктуры:
 - объем видеопамати: от 24 Гб
 - объем оперативной памяти сервера: от 128 Гб
 - объем дискового массива: от 1 Тб

ПРАВОВАЯ ОХРАНА

- Свидетельство о государственной регистрации программы для ЭВМ № 2025693868 «Программный комплекс для организации мультипользовательского доступа к большим языковым моделям через веб-интерфейс»
- Свидетельство о государственной регистрации программы для ЭВМ № 2025697521 «Система контроля доступа к сессии высокопроизводительных вычислений с интеллектуальной диспетчеризацией GPU ресурсов»
- Свидетельство о государственной регистрации программы для ЭВМ № 2024685757 «Программный модуль для управления агентами вычислительного кластера в составе платформы данных и нейросетевых моделей»
- Свидетельство о государственной регистрации программы для ЭВМ № 2024686458 «Система оркестрации распределенными вычислительными ресурсами GPU для формирования среды управления экспериментами в области ML и ИИ»