Описание функциональных характеристик

Наименование: Корпоративная платформа работы с языковыми моделями (КИИ), версия 1.0.

1. Назначение

Программа предназначена для организаций, которым нужен единый защищённый доступ сотрудников к языковым моделям, общая база документов и учёт расходов. Программа не содержит собственных языковых моделей: она подключается к моделям российских провайдеров (Яндекс, GigaChat) и к моделям, развёрнутым на серверах самой организации.

Программа работает как веб-приложение. Пользователи открывают её в браузере, администратор устанавливает на собственном сервере в виде контейнеров Docker.

2. Состав программы

Компонент	Назначение
Веб-приложение	Интерфейс пользователей и администраторов, программный интерфейс
Обработчик очереди	Фоновые задачи: анализ договоров, транскрибация
База данных PostgreSQL	Пользователи, чаты, документы, настройки, учёт расходов
Redis	Сессии и счётчики лимитов

3. Функции

3.1. Чаты с языковыми моделями

- Создание диалогов, выбор модели из разрешённых пользователю.
- Потоковый вывод ответа, отдельный блок рассуждений, если модель его передаёт.
- Вложения: документы из базы знаний, файлы с устройства, изображения, аудио.
- Закрепление, переименование, папки и поиск по названиям и тексту сообщений.
- Режим аналитика: ответ строится только по вопросу и прикреплённым документам.
- Автоматическое сжатие длинной истории (скользящее окно и суммаризация).
- Экспорт ответа в Word, таблиц из ответа в Excel.
- Общий доступ к чату по ссылке-приглашению.

3.2. Проекты

- Личные и корпоративные проекты, объединяющие чаты, документы и инструкции.
- Постоянные инструкции для ответов, подключаемые ко всем чатам проекта.

- Участники проекта, права доступа по ролям.
- Закрепление модели за проектом.

3.3. База знаний

- Иерархия папок и документов, личный и корпоративный (в пределах подразделения) доступ.
- Извлечение текста из PDF, DOC, DOCX и текстовых форматов, редактирование текста.
- Поиск по названию.

3.4. Проверка договоров

- Анализ договора из базы знаний по выбранному чек-листу.
- Перспективы: сбалансированная, заказчика, исполнителя.
- Показ хода анализа, сохранение результата, экспорт отчёта в DOCX.

3.5. Творческая студия

- Создание изображений и видео по текстовому описанию.
- Параметры генерации: формат, разрешение, длительность.
- Избранные модели, лента последних результатов, галерея.

3.6. Транскрибация

- Преобразование аудиофайлов в текст отдельным разделом и из чата.
- Распознавание на локальной модели организации (по умолчанию); внешний российский сервис подключается по выбору администратора.
- Сохранение результата в документы пользователя.

3.7. Нейроны

- База «вопрос — ответ» подразделения, которая автоматически добавляется в контекст ответа модели.

3.8. Пользователи и подразделения

- Одна организация, внутри неё подразделения; база знаний, нейроны и корпоративные проекты общие в пределах подразделения.
- Суперадминистратор создаёт подразделения и видит всех пользователей; администратор подразделения видит и настраивает только своё подразделение.
- Роли: пользователь, администратор подразделения, суперадминистратор.
- Лимиты символов и расходов на пользователя, ограничения списка моделей.
- Сброс пароля по коду на электронную почту.

3.9. Статистика и счета

- Учёт токенов и стоимости в рублях по пользователям, моделям и периодам.

- Счета по подразделениям.

3.10. Администрирование

- Подключение провайдеров и моделей, цены, окна контекста, активность.
- Настройка логотипа, названия продукта и движка транскрибации.

4. Технические характеристики

Параметр	Значение
Языки и платформа	PHP 8.2, Laravel 12, JavaScript
СУБД	PostgreSQL 16
Кэш и сессии	Redis 7
Веб-сервер	Apache 2.4
Поставка	Образы Docker, запуск через Docker Compose
Интерфейс	Веб-браузер, русский язык
Шифрование соединений с провайдерами	TLS, с проверкой сертификатов, включая корневой сертификат УЦ Минцифры
Хранение секретов	Ключи провайдеров хранятся в базе в зашифрованном виде
Шифрование сообщений	Необязательное, AES-256-GCM (параметр `MESSAGE_ENCRYPTION_ENABLED`)

5. Внешние обращения программы

Назначение	Адрес	Когда
Яндекс GPT, картинки, видео	`llm.api.cloud.yandex.net`	Выбрана модель Яндекса
GigaChat	`ngw.devices.sberbank.ru`, `gigachat.devices.sberbank.ru`, `api.giga.chat`	Выбрана модель GigaChat
Локальная модель и локальная транскрибация	Адрес из карточки провайдера, только внутри сети организации	Выбрана локальная модель

Назначение	Адрес	Когда
Внешняя транскрибация	`speech2text.ru`	Только если администратор выбрал этот движок

Других обращений программа не выполняет. Сторонние скрипты, шрифты и стили загружаются только с самого сервера.

6. Требования к окружению

См. раздел 2 документа «Инструкция по установке и эксплуатации».