

# Техническое задание: Интеграция ии ассистента в чат тех поддержки

## 1. Обзор проекта

**Цель:** Внедрить ии сервис для поиска ответов из базы знаний в существующий чат тех поддержки для автоматизации обработки типовых вопросов.

**Принцип работы:** Пользователь задает вопрос → система находит готовый ответ из FAQ → возвращает результат без генерации новых текстов.

## 2. Архитектура решения

- **FAQ-сервис** — микросервис с API для поиска ответов
- **База знаний** — 145 вопросов-ответов в векторном индексе
- **Модель эмбедингов** — BAAI/bge-m3 для семантического поиска
- **FAISS индекс** — быстрый поиск похожих векторов

## 3 Интеграция в существующий чат

1. **Перехват сообщений** — до передачи оператору
2. **Вызов FAQ API** — асинхронный запрос к микросервису
3. **Обработка ответа** — по алгоритму выше
4. **Логирование** — сохранение статистики использования

## 4. Технические требования

- **Порт:** 8000
- **Протокол:** HTTP/HTTPS
- **Формат:** REST API
- **Время ответа:** < 3 секунд

### 4.1 Интеграция в чат

- **Асинхронные запросы** — не блокировать UI
- **Таймаут:** 5 секунд
- **Fallback** — при недоступности сервиса передать оператору
- **Кэширование** — для частых вопросов

### 4.2 Мониторинг и логирование

- **Метрики:** количество запросов, confidence, время ответа
- **Логи:** все запросы к FAQ API
- **Алерты:** при недоступности сервиса

## 5. Пользовательский опыт

1. **Автоответ** — пользователь получает готовый ответ мгновенно
2. **Уточнение** — система предлагает похожие вопросы
3. **Эскалация** — сложные вопросы передаются оператору

## 6. UI/UX требования

- **Индикатор загрузки** — при поиске ответа
- **История поиска** — для повторных вопросов
- **Обратная связь** — кнопки "👍/👎" для оценки ответов

## 7. Развертывание

- **Docker контейнер** — для FAQ-сервиса
- **Environment variables** — для конфигурации
- **Health checks** — для проверки состояния

## 8. Мониторинг

- **Метрики:** количество запросов, confidence, время ответа
- **Логи:** структурированное логирование
- **Алерты:** при критических ошибках