



**приоритет** 

SmartMLOps – система размещения и управления сервисами искусственного интеллекта

## Третий открытый онлайн семинар с командой SmartMLOps

Салех Хади Мухаммед, Департамент программной инженерии  
Факультета компьютерных наук НИУ ВШЭ г. Москва  
hsalekh@hse.ru

<https://mlops.hse.ru/>

Москва 2025



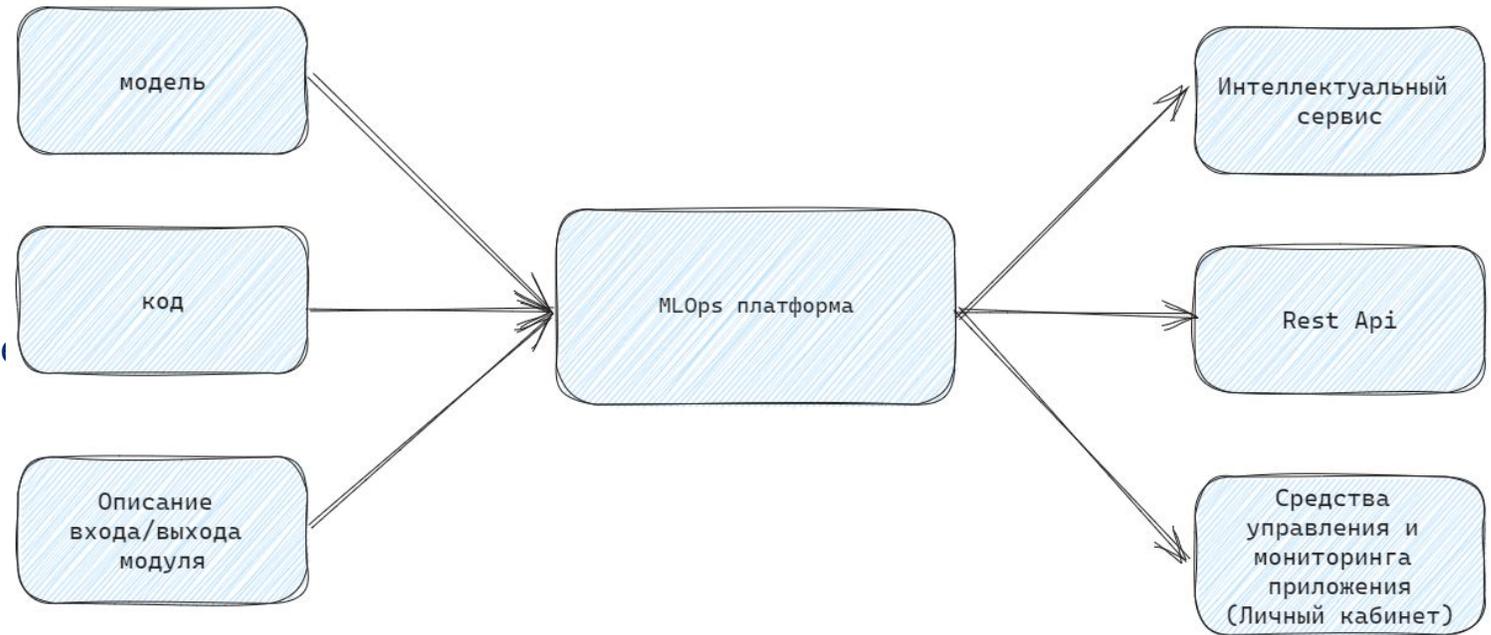
## План

1. Развертывание приложения через личный кабинет. Галынчик Максим.
2. Модуль управления файлами в личном кабинете. Галынчик Максим.
3. Дашборды и подсистема логирования. (Дашборды + Логи) Башун Владимир.
4. Организация и оказание технической поддержки. Салех Хади.
5. РИД, план работы и развитие системы SmartMLOps, Салех Хади.
6. Концепция озера данных. Прототип семантического слоя для озера данных. Силаев Юрий.
7. Обсуждение + вопросы

## MLOps платформа

Что такое MLOps платформа:

- вычислительные ресурсы
- хранилища данных
- система безопасности
- платформа для пользователей
- доступ к суперкомпьютеру CHARISMA
- мониторинг ресурсов и работы модели
- управление доступом и лицензиями к модулям
- личный кабинет
- консультации и сопровождение от команды техподдержки



Все для вашего интеллектуального приложения

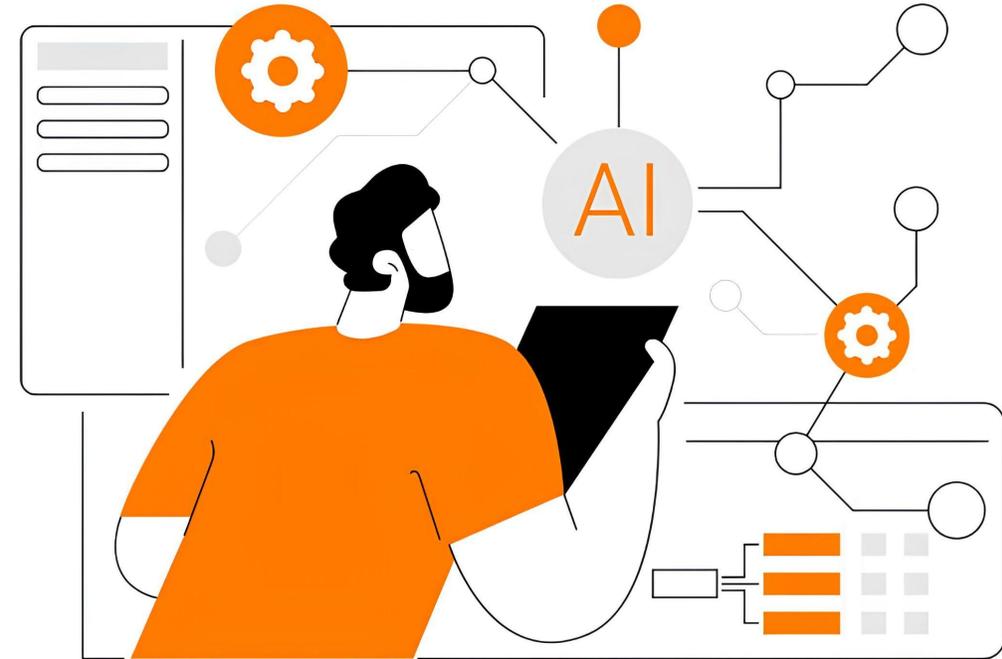


## Цели проекта

- создание и развитие центров компетенций по развитию и применению технологий ИИ НИУ ВШЭ;
- разработка и интеграция ИИ-помощников в образовательный процесс университета
- разработка и апробация технологий ИИ в административных процессах, в сферах управления и медицины совместно с партнерами
- снижение трудоемкости развертывания ИИ-решений как сервисов на основе данных корпоративных информационных систем НИУ ВШЭ
- предоставление необходимой инфраструктуры разработчикам моделей

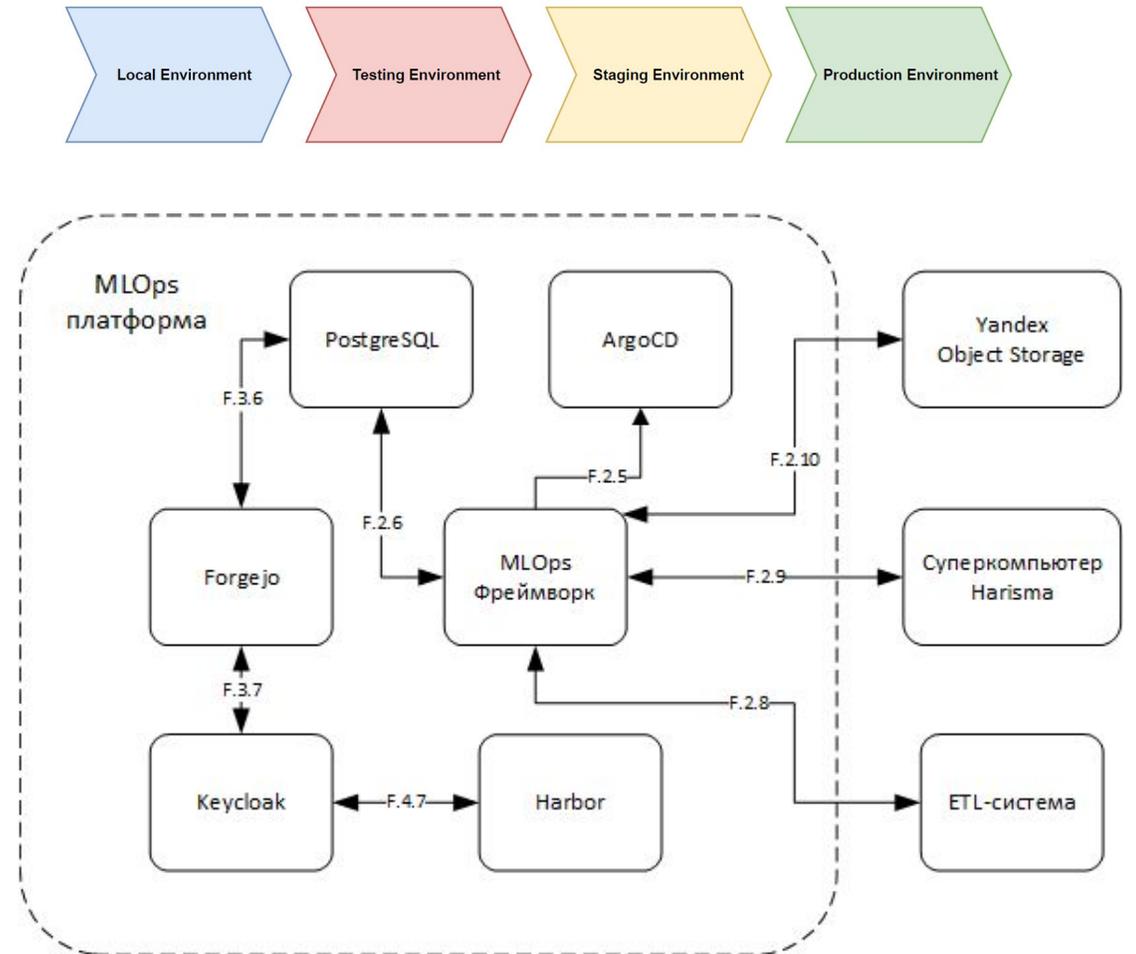
### Варианты использования:

- Размещение интеллектуальных сервисов, использование в рабочих процессах НИУ ВШЭ и для внешних пользователей
- Размещение учебных моделей, проведение опытной эксплуатации и внедрение результатов НИОКР лабораторий НИУ ВШЭ

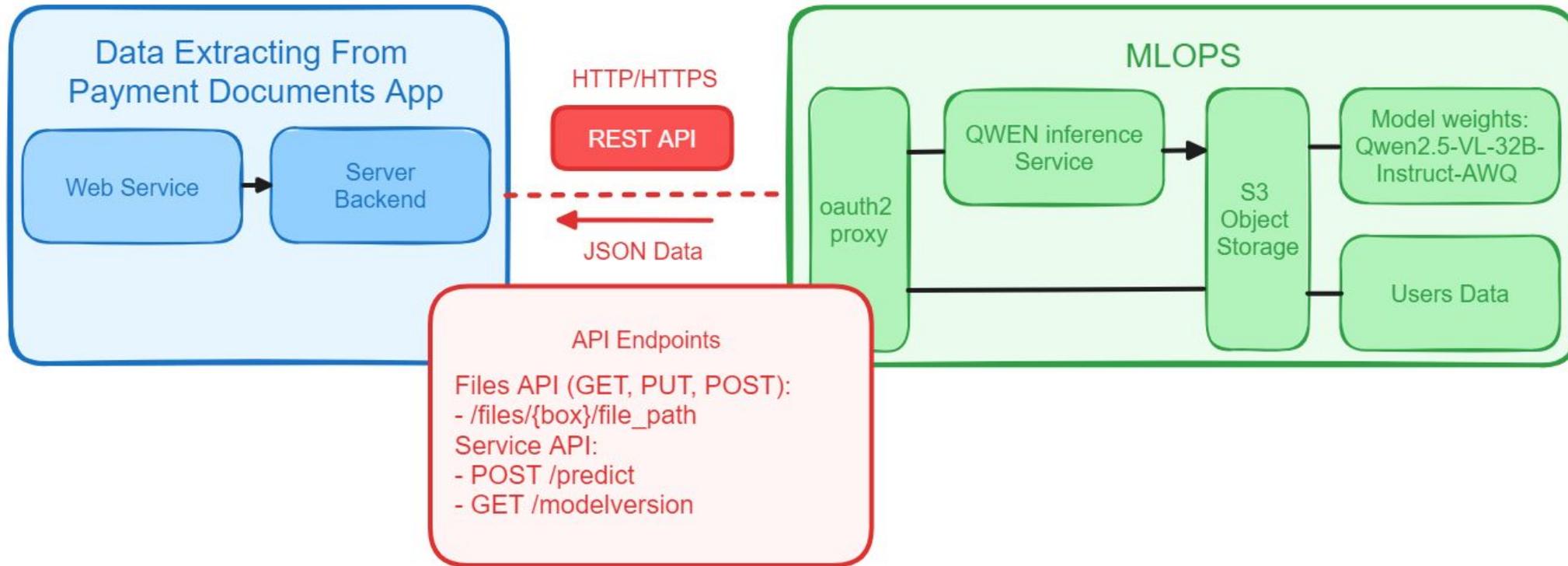




# Инфраструктура платформы



## Архитектура клиент-серверного сервиса ИИ, размещенная на MLOPS



### Data Flow:

1. User sends Payment Document via Web Service
2. Web Service, forwards data to Backend Server
3. Backend Server performs data preprocessing and calls QWEN service hosted by MLOPS via REST API
4. QWEN service recognize data from Payment Document and returns the result



## Организация и оказание технической поддержки

<p><b>Каналы технической поддержки</b></p> <p>Тех поддержка 16 Электронная почта 14 Телеграмм 40 Репозиторий документации 16</p>	<p><b>Темы обращений:</b></p> <p>Регистрация на платформе Вопросы по работе с сервисами платформы Развертывание моделей Локализация ошибок при запуске и работе моделей</p>
<p><b>В репозитории задач:</b></p> <ul style="list-style-type: none"><li>- всего 16:</li><li>- 8 открыто, 8 закрыто.</li><li>- 2 вопроса, 1 предложение, 13 багов.</li></ul> <p>Отслеживают 7 чел. (5 внутр., 2 внеш. разработ.) 3 автора(ов) отправили 54 коммитов в main и 54 коммитов во все ветви. На main, 46 файлов изменены, и там было 3485 добавлений и 380 удалений</p>	<p><a href="https://platform-forgejo.stratpro.hse.ru/mlops_platform/documentation/issues">https://platform-forgejo.stratpro.hse.ru/mlops_platform/documentation/issues</a></p>

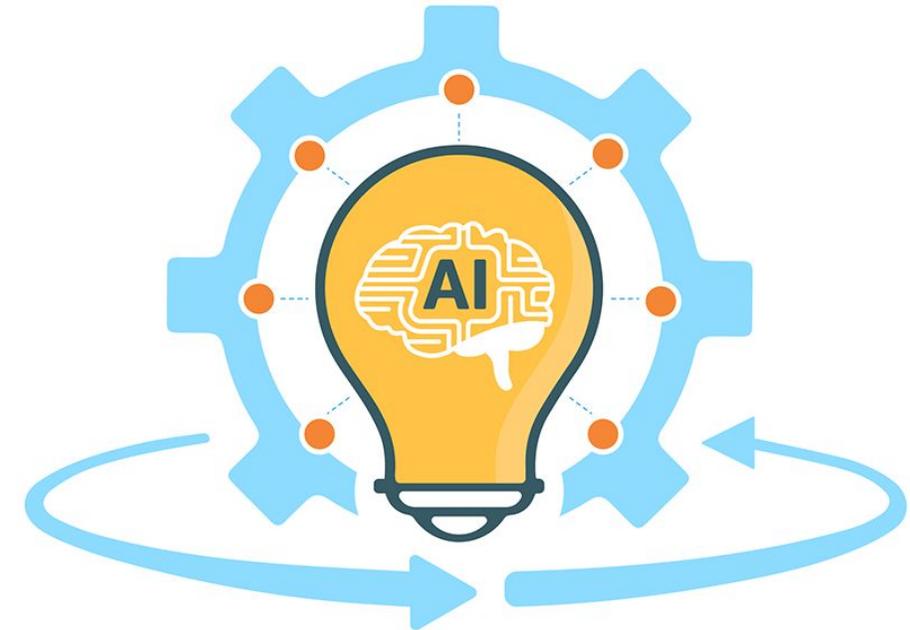


## Развернутые сервисы

Модель с RAG DeepSeek+ElasticSearch	Модель на основе нейросетевой модели DeepSeek, отвечающая на вопросы студентов НИУ ВШЭ на основе локально-нормативных актов НИУ ВШЭ, а также материалов с официального сайта НИУ ВШЭ.
Классификатор вопроса пользователя	Модель получает вопрос пользователя и возвращает категорию данного вопроса. Используется в чат-боте для ответов на вопросы студентов НИУ ВШЭ. Помогает улучшать поиск релевантных документов для ответов на вопросы.
Модель с RAG GigaChat + Chroma	Модель на основе GigaChat, отвечающая на вопросы студентов НИУ ВШЭ на основе локально-нормативных актов НИУ ВШЭ, а также материалов с официального сайта НИУ ВШЭ.
NevoScan – ИИ-ассистент для скрининга рака кожи	Программный модуль определяет риск злокачественности кожных новообразований (родинок) по фотографии. Сервис использует каскад из нескольких моделей глубокого обучения.
Сервис трудоустройства выпускников (Пифия)	Сервис предиктивной аналитики трудоустройства выпускников «Пифия» — это информационно-аналитическая платформа, позволяющая оценить потенциальный доход в зависимости от личных характеристик студента, университета, специальности и интересов.
ИИ агенты для образовательных и медицинских целей	Большая языковая модель из семейства Qwen3. Используется в качестве чат бота, который отвечает на вопросы пользователя касательно результатов поиска и помогает улучшить поиск.

## План работ и развития системы

- Разработка, настройка, доработка CI и CD процессов
- Разработка модуля логирования событий
- Предоставление фреймворком статусов развертывания компонентов приложения
- Разработка и развертывание дашбордов разработчика, дашборда администратора, шаблонов дашбордов;
- Разработка сервиса анонимизации данных
- Интеграция автоматического сбора метаданных и обновление онтологии для озера данных
- Создание механизма запроса и получения квот
- Ограничение использования сервисов приложений и фреймворка лимитами, устанавливаемыми разработчиком приложения
- Обновление веб Портала



### **РИДы** (Программа для ЭВМ):

1. Личный кабинет
2. Сервис анонимизации данных
3. Платформа SmartMLOps (программный комплекс)



## Концепция озера данных. Прототип семантического слоя

Корпоративное озеро данных – это единая платформа для централизованного хранения, обработки и выдачи данных из разрозненных систем организации

### Самообслуживание

Бизнес-пользователи получают доступ к данным через BI.

### Централизация данных

Централизованное хранилище для всех данных.



### Автоматизация

Автоматизированное управление пайплайном данных через Airflow.

### Зоны хранения

Зоны сырых, промежуточных и курируемых данных.

Спасибо за внимание!

Обратная связь:



Присоединяйтесь:



Салех Хади Мухаммед, Департамент программной инженерии  
Факультета компьютерных наук НИУ ВШЭ г. Москва  
hsalekh@hse.ru