



# Руководство администратора и технолога

Руководство администратора и технолога

В данном документе приведен порядок установки и настройки модуля «ИИ Агент». Веб-интерфейс модуля предоставляет доступ к функционалу после авторизации. Доступны разделы для управления промптами, справочниками, провайдерами и просмотра логов. Администратор может редактировать описания промптов, выбирать модели и тестировать запросы. Также реализованы функции управления паролем и настройки демонстрационного режима «Витрина ИИ». Система обеспечивает гибкую настройку интеграции с различными AI-провайдерами и моделями.

ИИ агент устанавливается и функционирует на устройствах с ОС: Windows, Linux.

Книга не является полной документацией к программному обеспечению. Для использования программы необходимо наличие других книг, включаемых в комплект.

Подсистема постоянно совершенствуется и в связи с этим возможны некоторые несоответствия, касающиеся описания пользовательского интерфейса.

Документация разработана совместно специалистами ГК «ЭОС».

Мобильные подсистемы являются торговой маркой ГК «ЭОС».

Право на тиражирование программных продуктов и документации принадлежит ГК «ЭОС».

Все упомянутые в данном издании товарные знаки и зарегистрированные товарные знаки принадлежат своим законным владельцам.

© ГК «Электронные Офисные Системы». Все права защищены

107113, Москва, ул. Шумкина, д. 20 стр. 1.

Тел/факс: (495) 221-24-31

e-mail: support@eos.ru

<http://www.eos.ru>

## Оглавление

|           |                                                                  |           |
|-----------|------------------------------------------------------------------|-----------|
| <b>1.</b> | <b>Условные обозначения .....</b>                                | <b>6</b>  |
| <b>2.</b> | <b>Архитектура .....</b>                                         | <b>7</b>  |
| <b>3.</b> | <b>Требования к программному и аппаратному обеспечению .....</b> | <b>8</b>  |
| 3.1.      | Требования к серверу (Сервис ИИ) .....                           | 8         |
| 3.2.      | Требования к программному обеспечению .....                      | 8         |
| 3.3.      | Требования к аппаратному обеспечению .....                       | 10        |
| 3.4.      | Требования к серверу LLM (vLLM + LLM).....                       | 11        |
| <b>4.</b> | <b>Установка vLLM и LLM.....</b>                                 | <b>15</b> |
| 4.1.      | Загрузка компонентов из интернета .....                          | 15        |
| 4.2.      | Перенос LLM.....                                                 | 16        |
| 4.3.      | Проверка .....                                                   | 20        |
| <b>5.</b> | <b>Установка агента на Windows сервер .....</b>                  | <b>21</b> |
| 5.1.      | Установка и настройка агента на Windows.....                     | 21        |
| 5.1.      | Обновление существующего сервера .....                           | 27        |
| 5.2.      | Управление несколькими серверами .....                           | 29        |
| 5.3.      | Удаление сервера .....                                           | 30        |
| <b>6.</b> | <b>Установка агента на Linux сервер .....</b>                    | <b>32</b> |
| 6.1.      | Установка и настройка агента на Linux .....                      | 32        |
| 6.2.      | Обновление существующего сервера .....                           | 40        |
| 6.3.      | Управление несколькими серверами .....                           | 42        |
| 6.4.      | Удаление сервера .....                                           | 43        |
| <b>7.</b> | <b>Настройка сайта агента .....</b>                              | <b>45</b> |

---

|                                            |    |
|--------------------------------------------|----|
| 7.1. Авторизация в модуле «ИИ Агент» ..... | 45 |
| 7.2. Витрина ИИ.....                       | 48 |
| 7.3. Структура сайта агента .....          | 48 |
| 7.3.1. Раздел «Промпты».....               | 49 |
| 7.3.2. Раздел «Справочники».....           | 56 |
| 7.3.3. Раздел «Провайдеры».....            | 59 |
| 7.3.3.1.Окно «Описание провайдера».....    | 60 |
| 7.3.4. Раздел «Протоколирование».....      | 63 |
| 7.3.5. Раздел «Настройки».....             | 65 |

## 1. Условные обозначения

- ИИ агент - название приложения
- ВКЛ - включено
- ВЫКЛ - выключено
- ОС - операционная система
- LLM - языковая модель
- LLM-сервер – сервер, содержащий LLM
- vLLM - программный продукт <https://vllm.ai/> для управления LLM моделями



### Примечание:

Данный документ соответствует версии:

- ИИ Агент 1.0.0.0
- СЭД «ДЕЛО» версии 26.0\_20260928.88 и выше



### Примечание:

ИИ агент предназначен для работы с русскоязычными текстами. Обработка запросов и документов на других языках не гарантируется и может приводить к некорректным результатам.

## 2. Архитектура

Подсистема состоит из следующих компонентов:

- ИИ агент, имеющий клиентскую часть для конфигурирования. Осуществляет связь между СЭД «ДЕЛО» и LLM сервером,
- LLM сервер используется для вычислений и генерации ответов,
- модуль интеграции «Дело» и ИИ агента.

## 3. Требования к программному и аппаратному обеспечению

### 3.1. Требования к серверу (Сервис ИИ)

#### Ограничения и прочие требования

Сервер может быть виртуальный или физический

### 3.2. Требования к программному обеспечению

ИИ-агент работает с СЭД «ДЕЛО» версии 26.0\_20260928.88 и выше

Для работы требуется:

|                   |                                                                                                                                                                                                                                                                |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ОС                | Windows:<br>Windows 10/11 или Windows Server 2019+ 64-бит)<br>Linux:<br>Linux-дистрибутив (Ubuntu, Debian, Fedora, CentOS/RHEL, Arch, openSUSE, ALT, Astra, РЕД ОС и т.д.)                                                                                     |
| Системные утилиты | Linux:<br>На целевой машине обязательно должны быть установлены: `jq`, `nginx`, `systemd` (`systemctl`), `ss` (пакет `iproute2`):<br>ALT Linux:<br><a href="#">apt-get update</a><br><a href="#">apt-get install -y zstd iproute2 nginx jq</a><br>Astra Linux: |

|            |                                                                                                                                                                                                                                                                              |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|            | <i>apt update &amp;&amp; apt install -y zstd iproute2 nginx jq</i><br>РЕД ОС:<br><i>dnf upgrade</i><br><i>dnf install -y zstd iproute nginx jq</i><br>При отсутствии любой из них скрипт завершится с ошибкой.                                                               |
| WEB-сервер | Windows:<br>Microsoft Internet Information Services (IIS) 10<br>URL Rewrite Module ( <a href="#">скачайте</a> и выберите версию x64)<br>Application Request Routing 3.0 (ARR) ( <a href="#">скачайте</a> и выберите версию x64),<br>Linux:<br>web-сервер Nginx 1.20.1 и выше |
| PowerShell | Windows:<br>PowerShell 5.1                                                                                                                                                                                                                                                   |

### Особенности дистрибутивов с нестандартным nginx

Скрипт определяет ОС по `/etc/os-release` и подстраивает пути конфигурации nginx:

| ОС          | Каталоги сайтов nginx                                                                |
|-------------|--------------------------------------------------------------------------------------|
| ALT Linux   | `/etc/nginx/sites-available`,`<br>`/etc/nginx/sites-enabled`<br>(создаются скриптом) |
| Astra Linux | `/etc/nginx/sites-available`,`<br>`/etc/nginx/sites-enabled`<br>(создаются скриптом) |

|        |                                                                                                                                                                                                                              |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| РЕД ОС | <code>`/etc/nginx/sites-available`,<br/>`/etc/nginx/sites-enabled`;     при<br/>отсутствии в `nginx.conf` скрипт<br/>добавляет в блок `http` строку<br/>`include               /etc/nginx/sites-<br/>enabled/*.conf`;</code> |
| Прочие | <code>`/etc/nginx/sites-available`,<br/>`/etc/nginx/sites-enabled`</code>                                                                                                                                                    |

Также скрипт проверяет, что в ``/etc/nginx/nginx.conf`` есть блок ``http {`` — иначе установка не начнется.

### 3.3. Требования к аппаратному обеспечению

#### Требования к аппаратному обеспечению:

| Минимальные требования | Рекомендуемые требования |
|------------------------|--------------------------|
| 4 vCPU                 | 8 vCPU                   |
| 32 ОЗУ                 | 64 ОЗУ                   |
| 512 Гб HDD             | 512 Гб HDD               |

#### Поддерживаемые ОС

- Windows Server 2019
- Альт СП Сервер релиз 10
- Альт Сервер 10.4
- Astra Linux Special Edition «Воронеж» 1.8.3
- РЕД ОС Сервер 7.3.7
- РЕД ОС Сервер 8

## 3.4. Требования к серверу LLM (vLLM + LLM)

### Ограничения и прочие требования

Сервер только физический

Развертывание через CPU не поддерживается

Сервис поставляется в виде Docker-контейнера

### Рекомендуемые LLM

Qwen3.8-27B

<https://huggingface.co/Qwen/Qwen3.8-27B>

### Требования к аппаратному обеспечению:

#### Компьютер с доступом в интернет

##### Windows

1. Windows 11 и выше
2. Docker Desktop версии 4.30 и выше
3. WSL2
4. Hugging Face Hub CLI (hf) (<https://pypi.org/project/huggingface-hub/>)
5. registry:2 или tar

##### Ubuntu 25.10

1. Ubuntu 25.10
2. Docker Engine версии 29.1.3 (docker.io)
3. python3
4. python3-pip
5. python3-venv
6. Hugging Face Hub CLI (hf) (<https://pypi.org/project/huggingface-hub/>)
7. registry:2 или tar

#### Сервер ИИ без доступа в интернет

##### Windows

1. Windows 11 и выше
2. Docker Desktop версии 4.30 и выше

- 3. WSL2
- 4. Driver NVIDIA Studio версии 595.79 и выше
- 5. CUDA Toolkit 12.9
- 6. registry:2 или tar

**Ubuntu 25.10**

- 1. Ubuntu 25.10
- 2. Docker Engine версии 29.1.3 ([docker.io](https://docs.docker.com/engine/install/ubuntu/))
- 3. Driver NVIDIA Studio версии 595.79 и выше
- 4. NVIDIA Container Toolkit (nvidia-container-toolkit, libnvidia-container1, libnvidia-container-tools)
- 5. CUDA Toolkit 12.9
- 6. registry:2 или tar

| Минимальные требования к аппаратной конфигурации | Рекомендуемые требования к аппаратной конфигурации |
|--------------------------------------------------|----------------------------------------------------|
| 16 vCPU                                          | 20 vCPU                                            |
| 128 ОЗУ                                          | 256 ОЗУ                                            |
| 512 Гб HDD                                       | 1 Тб HDD                                           |

**Поддерживаемые ОС**

Linux ОС  
Windows ОС

**Требования к доступной видеопамяти:**

| Количество одновременных запросов | Общая VRAM |
|-----------------------------------|------------|
| 10                                | ~44 Гб     |
| 50                                | ~58 Гб     |
| 80                                | ~68 Гб     |
| 100                               | ~75 Гб     |

**Требования к количеству видеокарт:**

| Количество<br>одновременных<br>запросов | H 100<br>(минимум) | H 100<br>(комфортно) | RTX PRO<br>6000<br>(минимум) | RTX PRO<br>6000<br>(комфортно) |
|-----------------------------------------|--------------------|----------------------|------------------------------|--------------------------------|
| 10                                      | 1                  | 2                    | 1                            | 2                              |
| 50                                      | 3                  | 4                    | 5                            | 6                              |
| 80                                      | 5                  | 6                    | 8                            | 10                             |
| 100                                     | 6                  | 8                    | 9                            | 12                             |

**Необходимые библиотеки:**

| Способ<br>переноса<br>vim-<br>контейнера | Компьютер с доступом<br>в интернет                                                                                                                                                                                                                                                                 | Сервер ИИ без доступа в<br>интернет                                                                                                                                                                                                                                                                                                     |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Через registry                           | <ul style="list-style-type: none"><li>• Docker Engine 24+ или новее (Ubuntu 25.10: <a href="https://docs.docker.com/engine/install/ubuntu/">docker.io</a>)</li><li>• Python3 (python3, python3-pip, python3-venv)</li><li>• huggingface-hub (hf)</li><li>• Docker Registry (registry:2)*</li></ul> | <ul style="list-style-type: none"><li>• NVIDIA Driver</li><li>• Docker Engine 24+ или новее (Ubuntu 25.10: <a href="https://docs.docker.com/engine/install/ubuntu/">docker.io</a>)</li><li>• NVIDIA Container Toolkit (nvidia-container-toolkit, libnvidia-container1, libnvidia-container-tools)</li><li>• CUDA Toolkit 12.9</li></ul> |

|           |                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                       |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Через tar | <ul style="list-style-type: none"><li>• Docker Engine 24+ или новее (Ubuntu 25.10: <a href="https://docs.docker.com/engine/install/ubuntu/">docker.io</a>)</li><li>• Python3 (python3, python3-pip, python3-venv)</li><li>• huggingface-hub (hf)</li><li>• tar</li></ul> | <ul style="list-style-type: none"><li>• NVIDIA Driver</li><li>• Docker Engine 24+ или новее (Ubuntu 25.10: <a href="https://docs.docker.com/engine/install/ubuntu/">docker.io</a>)</li><li>• NVIDIA Container Toolkit (nvidia-container-toolkit, libnvidia-container1, libnvidia-container-tools)</li><li>• CUDA Toolkit 12.9</li><li>• tar</li></ul> |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

\* registry:2 нужен только если используется временный локальный Docker Registry. Если применяется корпоративный Registry, этот пункт пропускается.

**Примечание:**

Библиотеки и фреймворки для инференса модели (vLLM, PyTorch, Transformers, CUDA Runtime, NCCL, Safetensors, Tokenizers и др.) входят в состав Docker-образа и не требуют отдельной установки на целевых серверах.

## 4. Установка vLLM и LLM

Установка vLLM и LLM производится в 3 этапа:

1. Загрузка компонентов из интернета
2. Перенос LLM
3. Проверка

Ниже подробнее описано об этих этапах.

### 4.1. Загрузка компонентов из интернета

#### Шаг 1. Загрузить образ vLLM

Из Терминала запустите команду:

```
docker pull vllm/vllm-openai:v0.19.0
```

#### Шаг 2. Скачать модель

- Определите локальную директорию, в которую будут загружены файлы модели. Рекомендуется использовать отдельный каталог для хранения моделей, например: D:\vllm-packages\models\  
Для каждой модели рекомендуется создавать отдельный подкаталог, например: D:\vllm-packages\models\Qwen3.8-27B-MTP-NVFP4
- Для загрузки модели выполните команду:

```
hf download sakamakismile/Qwen3.8-27B-MTP-NVFP4 --local-dir D:\vllm-packages\models\Qwen3.8-27B-MTP-NVFP4
```

где:

sakamakismile/Qwen3.8-27B-MTP-NVFP4 - идентификатор модели на Hugging Face;

D:\vllm-packages\models\Qwen3.8-27B-MTP-NVFP4 - локальная директория для сохранения файлов модели.

После завершения загрузки в указанном каталоге должны появиться файлы модели, например:

D:\vllm-packages\models\Qwen3.8-27B-MTP-NVFP4

```
├─ config.json
├─ tokenizer.json
├─ tokenizer_config.json
├─ generation_config.json
├─ model-00001-of-00008.safetensors
├─ model-00002-of-00008.safetensors
└─ ...
```

## 4.2. Перенос LLM

### Способ 1. Перенос через реестр Docker

#### Шаг 3. Создать собственный образ

- Создайте в директории рядом с папкой models файл Dockerfile со следующим наполнением:

```
FROM vllm/vllm-openai:v0.19.0
COPY models/ /models/
```

- В Терминале перейдите в директорию, в которой располагаются Dockerfile и models:

```
cd D:\vllm-packages
```

```
cd D:\vllm-packages:
```

```
docker build -t corp-vllm:1.0 .
```

#### **Шаг 4. Опубликовать собственный образ в реестр образов**

Если у вас нет корпоративного реестра, вы можете временно поднять его на своей машине одной командой:

```
docker run -d -p 5000:5000 --name local-registry registry:2
```

- Выполните сборку тега и отправку образа. Вместо <АДРЕС\_РЕЕСТРА> подставьте нужный вариант:
  - Если используете временный локальный реестр: localhost:5000
  - Если реестр на другой машине в сети: IP\_АДРЕС\_МАШИНЫ:5000
  - Если используете корпоративный реестр: ваш рабочий адрес (например, registry.company.local/ai)

# 1. (При необходимости) Авторизуйтесь в реестре

```
docker login <АДРЕС_РЕЕСТРА>
```

# 2. Присвойте образу тег реестра

```
docker tag corp-vllm:1.0 <АДРЕС_РЕЕСТРА>/corp-vllm:1.0
```

# 3. Отправьте образ в реестр

```
docker push <АДРЕС_РЕЕСТРА>/corp-vllm:1.0
```

### Шаг 5. Запустить vLLM на сервере без доступа в интернет

Важно: Если на Шаге 4 вы использовали временный реестр на базе registry:2 без SSL, перед выполнением команд ниже убедитесь, что адрес вашего реестра внесен в список insecure-registries в настройках Docker на целевом сервере.

- Замените <АДРЕС\_РЕЕСТРА> на сетевой адрес машины, куда вы отправили образ на Шаге 4 (например, 192.168.1.50:5000 или registry.company.local/ai).
- Загрузите образ из реестра:

```
docker pull <АДРЕС_РЕЕСТРА>/corp-vllm:1.0
```

- Запустите контейнер vLLM на порту 8000:

```
docker run --gpus all -p 8000:8000 -v D:\vllm-packages\models:/models  
vllm/vllm-openai:v0.28.0 --model /models/Qwen3.8-27B-MTP-NVFP4 --  
served-model-name Qwen3.8-27B
```

### Способ 2. Перенос через tar-файл

#### Шаг 3. Создать tar-архив

- Создайте в директории рядом с папкой models директорию docker-images

- В Терминале перейдите в директорию, в которой располагаются Dockerfile и models:

```
cd D:\vllm-packages
```

- Сохраните Docker-образ:

```
docker save vllm/vllm-openai:v0.19.0 -o docker-images/vllm-openai.tar
```

- Упакуйте в архив:

```
tar -czf offline-package.tar.gz D:\vllm-packages
```

#### **Шаг 4. Перенести архив на сервер**

#### **Шаг 5. Запустить vLLM на сервере без доступа в интернет**

- Определите локальную директорию для распаковки архива, например: D:\vllm-packages\
- Распакуйте архив в локальную директорию
- Импортируйте образ

```
docker load -i D:\vllm-packages\docker-images\vllm-openai.tar
```

- Запустите контейнер, выполнив команду:

```
docker run --gpus all -p 8000:8000 -v D:\vllm-packages\models:/models  
vllm/vllm-openai:v0.28.0 --model /models/Qwen3.8-27B-MTP-NVFP4 --  
served-model-name Qwen3.8-27B
```

### 4.3. Проверка

Проверьте, что следующая команда выводит список скачанных ранее моделей:

```
curl http://localhost:8000/v1/models
```

Проверить работу непосредственно модели можно при помощи тестового запроса:

(cmd)

```
curl http://localhost:8000/v1/chat/completions ^  
-H "Content-Type: application/json" ^  
-d '{"model":"'Qwen3.8-  
27B"', "messages": [{"role": "user", "content": "Привет"}]}'
```

(PowerShell)

```
curl http://localhost:8000/v1/chat/completions `  
-H "Content-Type: application/json" `  
-d '{  
  "model": "Qwen3.8-27B",  
  "messages": [  
    {  
      "role": "user",  
      "content": "Привет"  
    }  
  ]  
}'
```

## 5. Установка агента на Windows сервер

### 5.1. Установка и настройка агента на Windows

Для установки выполните следующие шаги:

#### 1. Получение и подготовка дистрибутива

- 1.1. Скачайте архив eosai\_win.zip (имя отличается в зависимости от ОС)
- 1.2. Распакуйте его в любую временную папку, например C:\Temp\EOSAI
- 1.3. Внутри вы увидите:
  - python – портативный Python со всеми зависимостями,
  - backend – серверная часть,
  - Front и wwwroot – фронтенд и собранный фронтенд,
  - scripts – скрипты запуска, остановки и удаления,
  - install.ps1 - установочный скрипт для установки и обновления,
  - deploy – инструменты для деплоя (nssm.exe для регистрации Windows-службы).

#### 2. Запуск скрипта

Откройте PowerShell от имени администратора и перейдите в папку дистрибутива. Затем выполните:

```
cd "C:\путь\к\папке\дистрибутива"  
.\install.ps1
```

Если политика выполнения скриптов не позволяет запуск, временно разрешите:

*Set-ExecutionPolicy -Scope Process -ExecutionPolicy Bypass*  
*.\install.ps1*

### **3. Процесс исполнения скрипта**

Скрипт началом установки скрипт автоматически выполнит ряд проверок:

- права администратора - если скрипт запущен без прав администратора, он сразу завершится с ошибкой.
- наличие IIS – проверка службы `W3SVC`. Если IIS не установлен, скрипт завершится с ошибкой.
- URL Rewrite Module – проверка наличия в реестре. Если установлен, выведет версию. Если нет — скрипт завершится с ошибкой.
- ARR (Application Request Routing) – проверка наличия версии 3.0+. Если не установлен или версия ниже 3.0 — скрипт завершится с ошибкой.
- включение ARR Proxy – скрипт автоматически включит ARR Proxy, если он еще не включен.

После проверок скрипт прочитает список уже установленных серверов EOSAI из реестра:

- если серверов нет — скрипт сразу перейдет к чистой установке без лишних вопросов
- если серверы есть — выведет их список и предложит выбрать режим. (см. Рис. 5.1.1)

```
=====
EOSAI – установка / обновление
=====

Найдите установленный сервер:
eosa1 -> C:\Program Files\EOS\EOSAI
Выбор режима установки

[1]. Обновить существующий
[2]. Установить новый сервер

Введите номер [1,2] По умолчанию – Обновить существующий:
```

Рис. 5.1.1 – Выбор режима установки

- «1» – обновление выбранного сервера (конфигурационные файлы сохраняются, код заменяется новым из дистрибутива),
- «2» – чистая установка нового экземпляра (будут запрошены папка, имя, порт, а затем настройки провайдера и модели).

По умолчанию (кнопка «Enter») выбран вариант **[1] — Обновить существующий**. Если серверов еще нет, скрипт сразу перейдет к чистой установке.

## 4. Первоначальная установка

### 4.1. Основные параметры (см. Рис. 5.1.2)

- «Папка установки» – по умолчанию C:\Program Files\EOS\EOSAI. Можно указать любой другой путь, папка должна быть пустой или не существовать. При конфликте предложит ввести другую папку,
- «Имя сервера» – по умолчанию eosa1. Уникальное слово. Повторять имена уже существующих серверов нельзя – скрипт проверит и попросит ввести другое при совпадении,
- «Порт фронтенда» (FrontPort) – по умолчанию 3333. Скрипт проверит, свободен ли этот порт в системе, а также не используется ли он другими серверами EOSAI (читает их файлы iisconfig.json).). При конфликте предложит выбрать другой порт,
- «Порт бэкенда» (BackendPort) – по умолчанию 3334. Скрипт проверит, свободен ли этот порт в системе, а также не

используется ли он другими серверами EOSAI (читает их файлы `iisconfig.json`). При конфликте предложит выбрать другой порт.

```
Введите директорию для новой установки (по умолчанию C:\Program Files\EOS\EOSAI):  
Путь сохранен: C:\Program Files\EOS\EOSAI  
Создана новая директория: C:\Program Files\EOS\EOSAI  
Выбрана директория установки: C:\Program Files\EOS\EOSAI  
Введите имя сервера (по умолчанию eosai):  
Введите порт для фронта сервера (по умолчанию 3333):  
Введите порт для бэкенда сервера (по умолчанию 3334):  
Будет выполнена установка в "C:\Program Files\EOS\EOSAI" с портом для сайта 3333 и портом для серверной части 3334, имя сервера "eosai".  
==> ЧИСТАЯ УСТАНОВКА ==>  
Собираем файлы на C:\AI\38208901-3\BuildResult\ecosai_min\ecosai_win в C:\Program Files\EOS\EOSAI ...
```

Рис. 5.1.2 – Основные параметры при установке

## 4.2 Интерактивная настройка провайдера и модели

После копирования файлов запустится мастер настройки. Вам нужно создать **одного провайдера и одну модель** (см. Рис. 5.1.3).

**Провайдер (обязательно один):**

- Тип: `vllm` или `ollama`. Для выбора необходимо нажать соответствующую типу цифру.



### Примечание:

Поддержка типа `ollama` сделана в ознакомительных целях. Корректная работа функционала поддерживается при выборе типа `vllm`.

- URL. Скрипт автоматически добавит нужный суффикс (`/v1` для `vllm`, `/api` для `ollama`).
- API-ключ (можно оставить пустым, нажав кнопку «Enter»).

**Примечание:**

Поддержка типа ollama сделана в ознакомительных целях. Корректная работа функционала поддерживается при выборе типа vllm.

**Модель (только одна):**

- Введите название. Отображаемое имя будет сгенерировано автоматически (точки, тире и т.д. заменяются на \_).
- Модель привязывается к созданному провайдеру.

```
=====
НАСТРОЙКА КОНФИГУРАЦИИ
=====

-- Создание провайдера и модели --
Выбор типа провайдера
[1]. vllm
[2]. ollama

Введите номер (1,2) По умолчанию - vllm
Выбор тип vllm
Адрес провайдера модели: http://mkrcai.eos.lcc:3336
Адрес провайдера токенов: http://mkrcai.eos.lcc:3336/m1
API key (оставьте пустым, если не требуется)
Провайдер '_vllm_' добавлен.

-- Добавление модели --
Введите название модели: glltensor-model-hub/Qwen3.8-72B-NVFP4-RTX5090
Модель '_vllm_glltensor_model_hub_Qwen3.8-72B-NVFP4-RTX5090' (id: glltensor-model-hub/Qwen3.8-72B-NVFP4-RTX5090) добавлена.

Добавлен провайдер: 1
Добавлено моделей: 1

Конфигурация сохранена в C:\Program Files\EasyLLM\backend\config\settings2.json
prompts.json дополнен указанием на модель '_vllm_glltensor_model_hub_Qwen3.8-72B-NVFP4-RTX5090' и провайдера '_vllm_'
сервер 'local' зарегистрирован в реестре.
```

Рис. 5.1.3 – Настройка конфигурации

После завершения настройки (см. Рис. 5.1.4):

- конфигурация сохраняется в backend\config\settings2.json
- данные о сайте в IIS, пуле в IIS и службе Windows, которые создаются в процессе, сохраняются в backend\config\iisconfig.json (включая FrontPort и BackendPort)
- файл с предустановленными промптами (prompts.json) автоматически обновляется: для всех промптов поле модели заполняется именем добавленной вами модели и идентификатором провайдера

- сервер регистрируется в реестре Windows
- фронтенд регистрируется как сайт в IIS (в отдельном новом пуле), а бэкэнд как Windows-служба через NSSM (фоновый процесс, не привязанный к консоли).

```

Конфигурация сохранена в C:\Program Files\EOS\EOSAI\backend\config\settings1.json
scripts\start.ps1 должен указывать на модуль "vlls_gittensor_cuda3_run_backend_8_370_WTP4 RTX890" и провайдер "vlls"
Сервер "eosai" зарегистрирован в реестре.

Name      State      Applications
----      -
eosai     Started

Name      : eosai
ID        : 8
State     : Started
PhysicalPath : C:\Program Files\EOS\EOSAI\wwwroot
Bindings  : Microsoft.IIS.PowerShell.Framework.ConfigurationElement

Назначаем пул в IIS, создаем C:\Program Files\EOS\EOSAI\wwwroot...
IIS сайт "eosai" успешно создан и получен права.
Сохранение информации о IIS (в том числе, портах)
web.config успешно обновлен: C:\Program Files\EOS\EOSAI\wwwroot\web.config
Регистрация службы "eosai"...
2025-09-04 09:42:06,824 INFO - Installing service "eosai-Служба сервера EOS ИИ-агента (eosai)"...
2025-09-04 09:42:06,875 INFO - Service "eosai-Служба сервера EOS ИИ-агента (eosai)" was installed successfully.
Служба "eosai" успешно зарегистрирована.

=====
ОСТАВШАЯ/ПОСЛЕДНЯЯ ЗАДАЧА:
=====

Для дальнейшей работы с приложением (остановка, старт, удаление) перейти в директорию с установленным кодом: C:\Program Files\EOS\EOSAI и выполнить соответствующий скрипт.
Для открытия в браузере: http://DELOECSMOBILE:3333
Запускаем приложение...

```

Рис. 5.1.4 – Завершение настройки

## 5. Запуск и остановка приложения

Сразу после завершения установки или обновления происходит запуск нового/обновленного сервера (IIS-пул, IIS-сайт, Windows-служба бэкэнда).

**Также существует способ запуска после остановки приложения:** Перейдите в папку установки (например, C:\Program Files\EOS\EOSAI, куда вы установили код) и выполните:

`.\scripts\start.ps1`

**Важно:** для запуска скрипта старта требуются права администратора.

Сервер будет доступен по адресу `http://localhost:<FrontPort>` (FrontPort указан в `backend\config\iisconfig.json`, по умолчанию 3333). Откройте его в браузере.

**Примечание:**

При первоначальной установке по умолчанию задан пароль «admin». При необходимости следует изменить пароль (подробнее см. п. «Раздел «Настройки» текущего документа).

**Для остановки** используйте скрипт:

```
.\scripts\stop.ps1
```

Скрипт остановит IIS-пул, IIS-сайт и Windows-службу бэкенда.

## 5.1. Обновление существующего сервера

Чтобы обновить уже установленный экземпляр сервера (см. Рис. 5.2.1):

1. Получите **новый** дистрибутив (распакованный архив с новой версией) и разместите его в отдельной папке (не в папке текущей установки!).
2. Откройте PowerShell от имени администратора, перейдите в папку с кодом нового дистрибутива и выполните:

```
.\install.ps1
```

```
PS C:\AI1\20260906.1\BuildResult\cosai_min\cosai_win> .\install.ps1
[Info] Кем-то установлен. Версия: 0.2.1395
[Info] Установлен. Версия: 1.0-05311 (защита: 0.0)

Включение АИИ Проху...
[Info] Проху уже включен.
Тогда
Используем дистрибутив из: C:\AI1\20260906.1\BuildResult\cosai_min\cosai_win
=====
COSAI - установка / обновление
=====

Найдены установленные серверы:
cosai -> C:\Program Files\EOS\EOSAI
Выбор режима установлен

[1]: Обновить существующий
[2]: Установить новый сервер

Введите номер (1..2) По умолчанию - Обновить существующий: 1
```

Рис. 5.2.1 – Начало исполнения скрипта

3. В появившемся меню нажмите кнопку «Enter» на клавиатуре или выберите «1» и нажмите кнопку «Enter»
4. Если в системе зарегистрировано несколько серверов, введите имя нужного. Если сервер один, он будет выбран автоматически.

Скрипт выполнит (см. Рис. 5.2.2):

- остановку работающего сервера (IIS-пул, сайт, Windows-служба),
- создание резервной копии старой папки,
- копирование новых файлов из дистрибутива,
- восстановление **всех ваших конфигурационных файлов** из бэкапа (вся папка *backend\config*, включая *settings2.json*, *prompts.json*, *iisconfig.json*),
- удаление бэкапа.

Интерактивная настройка при обновлении не запускается – ваши провайдеры, модели и порты остаются прежними.

```
Автоматически выбран единственный сервер 'eosai' для обновления
будет обновлен сервер 'eosai' в папке 'C:\Program Files\EOSAI'

=== РЕЖИМ ОБНОВЛЕНИЯ ===
Дви 'eosai' остановлен.
Служба 'eosai' остановлена.
Сайт 'eosai' остановлен.
Остановка всех компонентов завершена.
Ожидаем освобождения файлов...
Создаем файл: переименовываем C:\Program Files\EOSAI -> C:\Program Files\EOSAI.backup_20230808_090126
Копируем новые файлы из C:\AI1\20260901\Build\Result\eosai_win\eosai_win в C:\Program Files\EOSAI ...
Берем полную информацию из старого кода (инструменты, провайдеры, создание бэкаupa, создание бэкаupa и т.д.)...
Записи в сервере в реестре и IIS оставлены без изменений.

=====
ОБНОВЛЕНИЕ: конфигурация сохранена на предыдущий уровень
=====
Бэкап удален
web.config успешно обновлен: C:\Program Files\EOSAI\wwwroot\web.config
=====
Остановка/обновление завершено
=====
```

Рис. 5.2.2 – Обновление сервера

### Принудительное обновление (без меню)

Скрипт поддерживает запуск без интерактивных вопросов (удобно для скриптов централизованного развертывания или планировщика):

```
.\install.ps1 -UpdateMode
```

Если серверов несколько, необходимо указать имя через -ServerName:

```
.\install.ps1 -UpdateMode -ServerName "prod"
```

## 5.2. Управление несколькими серверами

Вы можете установить несколько независимых экземпляров EOSAI на одном компьютере:

- Каждый раз запускайте `.\install.ps1` из свежего дистрибутива
- Выбирайте «2» (Установить новый сервер)
- Указывайте разные папки, уникальные имена и свободные порты (и фронтенда, и бэкенда).

Каждый сервер регистрируется в реестре, и вы сможете увидеть их список при следующем запуске установщика.

## 5.3. Удаление сервера

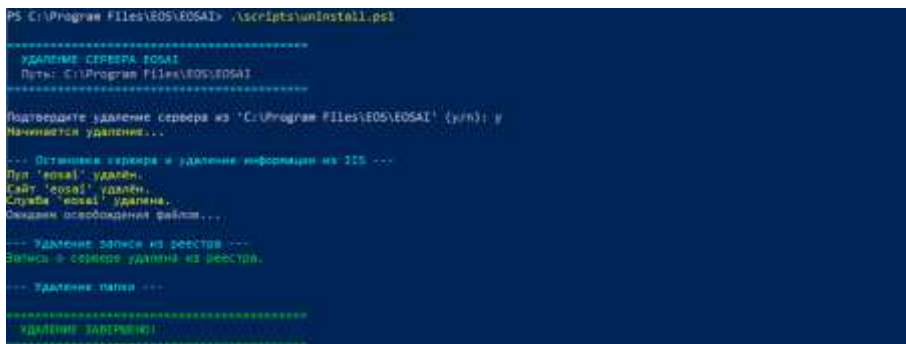
Для удаления запустите PowerShell 5 от имени администратора, перейдите в папку с установленным EOSAI, например, C:\Program Files\EOS\EOSAI

```
cd "C:\Program Files\EOS\EOSAI"
```

Выполните команду:

```
.\scripts\uninstall.ps1
```

Подтвердите удаление сервера: нажмите кнопку «у», а затем «Enter» на клавиатуре (см. Рис. 5.4.1).



```
PS C:\Program Files\EOS\EOSAI> .\scripts\uninstall.ps1

=====
УДАЛЕНИЕ СЕРВЕРА EOSAI
Путь: C:\Program Files\EOS\EOSAI
=====

Подтвердите удаление сервера из 'C:\Program Files\EOS\EOSAI' (y/n): y
Начинается удаление...

--- Остановка сервера и удаление информации из IIS ---
Пул 'eosai' удален.
Сайт 'eosai' удален.
Служба 'eosai' удалена.
Ожидание освобождения файлов...

--- Удаление записей из реестра ---
Записи в реестре удалены из реестра.

--- Удаление файлов ---

=====
УДАЛЕНИЕ ЗАВЕРШЕНО
=====
```

Рис. 5.4.1 – Удаление сервера

Скрипт последовательно выполнит:

- остановку сервера - остановка IIS-пула, IIS-сайта и Windows-службы бэкенда
- удаление из IIS - удаление сайта и пула приложений из IIS
- удаление Windows-службы - удаление службы бэкенда (NSSM)

- удаление из реестра - очистка записи о сервере из HKLM:\SOFTWARE\EOS\EOSAI
- удаление папки - удаление всей папки установки (включая код, конфиги и данные).

**Внимание:** удаление необратимо. Все данные сервера (конфигурация, промпты, логи) будут безвозвратно удалены. Перед удалением рекомендуется сделать резервную копию папки backend\config, если вы планируете переустановить сервер с сохранением настроек.

## 6. Установка агента на Linux сервер

### 6.1. Установка и настройка агента на Linux

Для установки выполните следующие шаги:

#### 1.Получение и подготовка дистрибутива

- 1.1. Скачайте архив `eosai-distribution-*.tar.gz` (имя может отличаться)
- 1.2. Распакуйте его в папку, из которой разрешено выполнение файлов (например, `/var/eosai_install` или `/opt`):

создание папки для распаковки архива:

```
mkdir -p /var/eosai_install
```

распаковка в созданную папку:

```
tar -xzf /var/<название_архива_дистрибутива> -C /var/eosai_install
```

например:

```
tar -xzf /var/eosai-distribution-3.13.15-x86_64.tar.gz -C /var/eosai_install
```



#### Примечание:

Не распаковывайте дистрибутив в папку `/tmp`. На многих системах каталог `/tmp` смонтирован с опцией ``noexec`` (или выполнение файлов из `/tmp` запрещено политикой безопасности) — тогда портативный Python из дистрибутива не запускается: при старте бэкенда будет ошибка «Permission denied». Используйте папку, где разрешено выполнение, например `/var/eosai_install` или `/opt`.

Назначьте права файлам дистрибутива:

```
chmod --recursive u=rx "/var/eosai_install"
```

### 1.3. Внутри вы увидите:

- `install.sh` – установщик (в корне дистрибутива),
- `python` – портативный Python со всеми зависимостями,
- `backend` – серверная часть (FastAPI),
- `wwwroot` – собранный фронтенд (именно его отдает nginx),
- `Front` – исходники фронтенда (в работе не участвуют),
- `scripts` – скрипты запуска, остановки и удаления.



#### Примечание:

Скрипт не умеет распаковывать архив самостоятельно – это необходимо сделать вручную перед запуском.

## 2. Запуск установки

Откройте терминал, перейдите в корневую папку распакованного дистрибутива и выполните **от root**:

`./install.sh`

Скрипт выполнит предпроверки (root, наличие ``jq`/`nginx`/`systemctl`/`ss``, блок ``http`` в ``nginx.conf``) и начнет работу. Все действия дублируются в лог-файл ``/tmp/eosai_<дата_время>.log``.

Если серверов еще нет, скрипт сразу перейдет к чистой установке (см. Рис. 6.1.2).



Рис. 6.1.2 – Выбор режима установки

### 3. Чистая установка (первый раз)

#### 3.1. Основные параметры

- «Директория установки» – по умолчанию /opt/eosai. Путь должен быть абсолютным; папка должна быть пустой или не существовать (создается автоматически),
- «Имя сервера» – по умолчанию `eosai`. Имя должно быть уникальным среди уже установленных серверов (проверяется по `/etc/eosai/instances/`),
- «Порт для фронта» - по умолчанию 3333. Скрипт проверит, что порт свободен в системе (`ss -lnt`) и не используется другими серверами EOSAI (читает `/etc/eosai/instances/\*.conf`). При конфликте попросит выбрать другой.
- Порт для бэкенда - по умолчанию `3334`. Проверяется так же, кроме того, не должен совпадать с портом фронта.

После этого скрипт скопирует файлы дистрибутива в директорию установки (см. Рис. 6.1.3).

```
Введите директорию для новой установки (по умолчанию /opt/eosai):
Путь корректен: /opt/eosai
Создана новая директория: /opt/eosai
Выбрана директория установки: /opt/eosai
Введите имя сервера (по умолчанию eosai):
Введите порт для фронта сервера (по умолчанию 3333):
Введите порт для бэкенда сервера (по умолчанию 3334):

Будет выполнена чистая установка в '/opt/eosai' с портом для сайта 3333 и портом для серверной части 3334, имя сервера 'eosai'

=== ЧИСТАЯ УСТАНОВКА ===
Копируем файлы из /tmp/eos в /opt/eosai ...
```

Рис. 6.1.3 – Основные параметры при установке

#### 3.2. Интерактивная настройка провайдера и модели

После копирования файлов запустится интерактивная настройка. Вам нужно создать **одного провайдера и одну модель** (см. Рис. 6.1.3).

**Провайдер (обязательно один):**

Тип: `vllm` или `ollama`. Для выбора необходимо нажать соответствующую цифру.

**Примечание:**

Поддержка типа `ollama` сделана в ознакомительных целях. Корректная работа функционала поддерживается при выборе типа `vllm`.

- URL. Скрипт автоматически добавит нужный суффикс (`/v1` для `vllm`, `/api` для `ollama`). Доступность провайдера по сети при установке **не проверяется** — достаточно корректного формата
- API-ключ (можно оставить пустым, нажав кнопку «Enter»)

Идентификатор провайдера (`id`) формируется автоматически как `_тип_` (например, `_ollama_`, `_vllm_`).

**Модель (только одна):**

- Введите название модели
- Модель привязывается к созданному провайдеру

```
НАСТРОЙКА КОНФИГУРАЦИИ
=====
Выбор типа провайдера
[1]. vllm
[2]. ollama

Введите номер (1..2) По умолчанию - vllm:
> 1
Выбран тип: vllm
Адрес провайдера модели: http://mlrpcal.eos.loc:3336
Адрес провайдера инференса: http://mlrpcal.eos.loc:3336/v1
API key (оставьте пустым, если не требуется):

— Добавление модели —
Введите название модели: gpt4o-mini
Конфигурация сохранена в /opt/coolai/backend/config/settings2.json
Готово! Для запуска используйте команду: vllm --vllm gpt4o-mini gpt4o-mini --provider vllm
```

Рис. 6.1.3 – Настройка конфигурации

Настройки сохраняются в `backend/config/settings2.json` (создается из шаблона `settings2.release.json`), промпты — в `backend/config/prompts.json` (создается из `prompts.release.json`) (см. Рис. 6.1.4).

```
Backend Python: /opt/eosai/python/bin/python3.11
Created symlink /etc/systemd/system/multi-user.target.wants/eosai-eosai.service → /etc/systemd/system/eosai-eosai.service.
База данных 'eosai' зарегистрирована (eosai-eosai)
nginx: the configuration file /etc/nginx/nginx.conf syntax is ok
nginx: configuration file /etc/nginx/nginx.conf test is successful
Created symlink /etc/systemd/system/multi-user.target.wants/nginx.service → /usr/lib/systemd/system/nginx.service.
Фронтенд 'eosai' зарегистрирован на порту 3333
Экземпляр 'eosai' полностью зарегистрирован
Фронт: http://<имя-сервера>:3333
API: http://<имя-сервера>:3333/api/

УСТАНОВКА ЗАВЕРШЕНА!
```

Рис. 6.1.4 – Завершение настройки



#### Примечание:

При чистой установке в `backend/config/settings2.json` копируется шаблон `settings2.release.json`, который содержит пользователя «admin» с паролем «admin».

Если вам нужен уникальный пароль — измените его сразу после установки:

1. Зайдите в систему под пользователем «admin» (пароль: «admin»).
2. Откройте «Настройки» в левом меню.
3. В разделе «Авторизация» нажмите кнопку «ИЗМЕНИТЬ» напротив строки «Пароль».
4. Введите текущий пароль, новый пароль и его подтверждение.

### 3.3. Регистрация службы и сайта

После настройки конфигурации скрипт (см. Рис. 6.1.5):

1. Регистрирует экземпляр — создает `/etc/eosai/instances/<имя>.conf` с параметрами: `INSTANCE_NAME`, `INSTANCE_PATH`, `BACKEND_PORT`, `FRONTEND_PORT`, `FRONTEND_DIR`
2. Создает systemd-службу `eosai-<имя>` (`/etc/systemd/system/eosai-<имя>.service`):
  - запускает бэкэнд портативным Python: `python backend/app/main.py --port <BACKEND_PORT> --workers <N>` (где `N` = количество ядер × 2)
  - работает от имени пользователя, от которого запущен скрипт (`root`)
  - автоматически перезапускается при сбое (`Restart=on-failure`)
  - включается при загрузке системы (`systemctl enable`) и **запускается сразу**
3. Создает сайт nginx `/etc/nginx/sites-available/eosai-<имя>.conf`:
  - слушает `FRONTEND_PORT`
  - отдает статические файлы из `<директория_установки>/wwwroot`
  - проксирует запросы `/api/` на `127.0.0.1:<BACKEND_PORT>`
  - для SPA-маршрутов (все остальные запросы) отдает `/assets/index.html`
  - создает симлинк в каталоге `sites-enabled`, проверяет конфигурацию (`nginx -t`) и перезапускает nginx.

SELinux: перед работой с nginx скрипт выполняет `setsebool -P httpd_can_network_connect 1` (разрешает nginx делать исходящие соединения для `proxy_pass`), а после копирования файлов — `restorecon -R -v <директория_установки>` (восстанавливает стандартные SELinux-контексты). Команды выполняются только если `setsebool`/`restorecon` доступны в системе; иначе выводится предупреждение и установка продолжается.

Если на любом этапе что-то пойдет не так, скрипт **откатит** установку: удалит директорию установки и зарегистрированные записи (systemd, nginx, `/etc/eosai/instances/`).



```
Для дальнейшего работы с приложением перейдите в /opt/eosai.  
Для открытия в браузере: http://red-devel.loc:3333
```

Рис. 6.1.5 – Завершение установки

Сервер уже запущен — можно сразу открывать браузер.

## РЕД ОС

В системе по умолчанию включен брандмауэр `firewalld` с SELinux, который блокирует входящие соединения — даже при запущенном сервере браузер не откроет страницу. Разрешите порты сервера (по умолчанию фронтенд `3333` и бэкэнд `3334`):

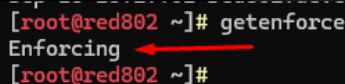
```
semanage port -a -t http_port_t -p tcp 3333  
semanage port -a -t http_port_t -p tcp 3334
```

(вместо 3333/3334 — свои значения `FRONTEND\_PORT`/`BACKEND\_PORT` из `/etc/eosai/instances/<имя>.conf`).

Если у вас не выполняется *semanage*, то проверьте включен или нет SELinux:

```
getenforce
```

Если он включен (ответ на команду enforce)



```
[root@red802 ~]# getenforce  
Enforcing  
[root@red802 ~]#
```

выполните:

```
dnf update sudo  
dnf install polycoreutils-python-utils
```

## 4.Запуск и остановка приложения

После установки (и после каждого обновления) бэкенд работает как systemd-служба ``eosai-<имя>``, а фронтенд отдается nginx.

### 4.1. Скрипты управления

Перейдите в папку установки (например, `/opt/eosai`) и выполните **от root** (скрипты управляют systemd-службой и каталогами ``/etc/nginx``):

```
./scripts/install.sh    # запустить  
./scripts/stop.sh     # остановить
```

Скрипты определяют текущий экземпляр по пути (сравнивают с ``INSTANCE_PATH`` в ``/etc/eosai/instances/*.conf``), затем:

- ``start.sh`` — запускает systemd-службу ``eosai-<имя>``, создает симлинк сайта nginx, проверяет конфигурацию (``nginx -t``) и перезагружает nginx (фронтенд снова доступен),
- ``stop.sh`` — останавливает службу, убирает симлинк сайта nginx, проверяет конфигурацию (``nginx -t``) и перезагружает nginx (фронтенд недоступен, бэкенд остановлен).

Перед выполнением действий скрипты проверяют наличие ``nginx`` и ``systemctl`` — при отсутствии завершаются с ошибкой. На ALT Linux пути каталогов сайтов nginx подстраиваются автоматически (``sites-available.d`/`sites-enabled.d``).

Скрипты нужно запускать именно из папки установки этого экземпляра — именно по текущему пути они находят свой конфиг.

### 4.2. Через systemctl напрямую

```
systemctl status eosai-<имя> # состояние  
systemctl start eosai-<имя>  # запустить бэкенд  
systemctl stop eosai-<имя>   # остановить бэкенд  
systemctl restart eosai-<имя> # перезапустить
```

systemctl start/stop` управляет только бэкендом. Симлинк сайта nginx (доступность фронта) при этом не меняется — для полного «включения/выключения» используйте `start.sh`/`stop.sh`.

### 4.3. Адрес приложения

Откройте в браузере: `http://<адрес-машины>:<FRONTEND\_PORT>` (по умолчанию `http://localhost:3333`).

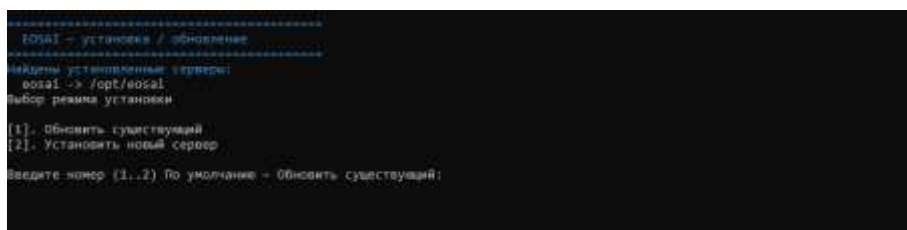
## 6.2. Обновление существующего сервера

### 1. Интерактивное обновление (с ручным выбором)

1. Получите **новый архив** дистрибутива
2. Распакуйте его во временную папку
3. Запустите установщик от root из распакованного архива:

`./install.sh`

4. В появившемся меню нажмите кнопку «Enter» на клавиатуре или выберите «1» и нажмите кнопку «Enter» (см. Рис. 6.2.1)



```
=====
eoas1 - установка / обновление
=====
Нажмите установленные серверы:
eoas1 -> /opt/eoas1
Выбор режима установки:

[1]. Обновить существующий
[2]. Установить новый сервер

Введите номер (1..2) По умолчанию - Обновить существующий:
```

Рис. 6.2.1 – Выбор режима установки

5. Если в системе зарегистрировано несколько серверов, введите имя нужного. Если сервер один, он будет выбран автоматически

## 6. Скрипт (см. Рис. 6.2.2):

- остановит systemd-службу бэкенда,
- временно уберет симлинк сайта nginx и перезапустит nginx,
- создаст резервную копию старой папки (переименует ее в ``папка>.backup_<timestamp>``),
- скопирует новые файлы из дистрибутива,
- восстановит ваши конфигурации: ``settings2.json`` (провайдер, модель), ``prompts.json`` (ваши пользовательские промпты), каталог ``backend/traces``,
- перезапустит службу и вернет сайт nginx,
- удалит бэкап.



```
root@kali:~/scripts# ./update_server.sh
Создаем резервную копию папки...
Создаем бэкап: переименовываем /opt/musai -> /opt/musai_backup_20260917_104123
Скопируем новые файлы из /etc/nginx/nginx.conf в /opt/musai...
Скопируем новую информацию из старого кода (настройки модели, провайдера, созданные нами промпты и т.д.)...
Вывод о статусе в systemd/nginx оставим без изменений.
nginx: the configuration file /etc/nginx/nginx.conf syntax is ok
nginx: configuration file /etc/nginx/nginx.conf test is successful
Сайт обновлен
```

Рис. 6.2.2 – Обновление сервера

**SELinux:** при обновлении скрипт также выполняет ``setsebool -P httpd_can_network_connect 1` и `restorecon -R -v <директория_установки>`.`

Ваши настройки (провайдер, модель, порты) останутся без изменений. Интерактивная настройка провайдера при обновлении **не запускается**.

Если на любом этапе произойдет сбой, скрипт **автоматически восстановит** старую папку из бэкапа и перезапустит службу.

## 2. Автоматическое обновление (для CI/CD)

Скрипт поддерживает **неинтерактивный режим** с параметрами командной строки:

*./install.sh --update --server-name <имя\_сервера>*

- update (или ``-u``) — принудительно включает режим обновления (без запроса «обновить/новый»),
- name (или ``-n``) — имя сервера, который нужно обновить.

**Обязателен**, если у вас несколько серверов; если сервер один, его можно опустить (будет выбран автоматически).

Примеры:

- Если сервер один:  
*./install.sh --update*
- Если серверов несколько:  
*./install.sh --update --name prod*

В автоматическом режиме:

- Скрипт не задает вопросов, кроме случая, когда серверов несколько и не указано имя – тогда он попросит ввести имя. В неинтерактивном режиме (CI/CD, stdin не является терминалом) этот запрос **не завершится**, а будет ждать ввода бесконечно, поэтому при автоматическом запуске параметр ``--name`` **обязателен**.
- Все остальное происходит так же, как при интерактивном обновлении.

## 6.3. Управление несколькими серверами

Вы можете установить несколько независимых экземпляров EOSAI на одном компьютере:

- Запускайте *./install.ps1* из свежего дистрибутива
- Выбирайте «2» (Установить новый сервер)
- Указывайте разные папки, уникальные имена и свободные порты.

Каждый сервер регистрируется в `/etc/eosai/instances/<имя>.conf` (по одному файлу на сервер), получает свою systemd-службу `'eosai-<имя>'` и свой сайт nginx `'eosai_<имя>.conf'`.

Список всех серверов можно посмотреть:

- в каталоге `/etc/eosai/instances/` (файлы `'<имя>.conf'`),
- или запустив `'install.sh'` из любого дистрибутива — он сам покажет список в меню.

## 6.4. Удаление сервера

Для полного удаления экземпляра EOSAI с машины выполните команду **из папки установки** этого экземпляра:

`./scripts/install.sh`

Подтвердите удаление сервера: нажмите кнопку «у», а затем «Enter» на клавиатуре (см. Рис. 6.4.1).

```
root@devastral83:~# /opt/eosai/scripts/uninstall.sh
Удаляем проект...
Подтвердите удаление экземпляра 'eosai' из '/opt/eosai' (y/n) y
Removed "/etc/systemd/system/multi-user.target.wants/eosai-eosai.service".
nginx: the configuration file /etc/nginx/nginx.conf syntax is ok
nginx: configuration file /etc/nginx/nginx.conf test is successful
Экземпляр 'eosai' успешно удален
root@devastral83:~#
```

Рис. 6.4.1 – Удаление сервера

Скрипт:

- находит конфиг экземпляра по текущему пути (`'INSTANCE_PATH'` в `/etc/eosai/instances/*.conf`)
- удаляет конфиг `/etc/eosai/instances/<имя>.conf`

- останавливает и отключает systemd-службу `eosai-<имя>`, удаляет ее файл
- убирает симлинк сайта nginx и удаляет конфиг сайта `/etc/nginx/sites-available/eosai\_<имя>.conf`, проверяет nginx (`nginx -t`) и перезагружает его
- удаляет саму папку установки

**Внимание:** удаление необратимо. Папка установки (вместе с `backend/config`, промптами и трейсами) будет удалена. Скрипт не запрашивает подтверждение — удаление выполняется сразу. При необходимости сохраните нужные данные заранее.

## 7. Настройка сайта агента

### 7.1. Авторизация в модуле «ИИ Агент»

Для открытия сайта модуля «ИИ Агент» введите в адресной строке путь к нему «`http://имя_сервера /имя _приложения`». Адрес содержит `<имя_сервера>:<порт_фронта*> *` который указали при установке, например : `http://hqv-mobfe2026.sup.loc:3335/`.

Откроется страница авторизации (см. Рис. 7.1.1) со следующими реквизитами:

- версия - номер версии ИИ Агента,
- поле «Имя» - логин пользователя,
- поле «Пароль» - пароль пользователя,
- кнопка «ВОЙТИ» - вход в систему,
- кнопка «ВИТРИНА ИИ» - переход в демо-режим (витрина) без полного входа.



Рис. 7.1.1 - Окно авторизации

По нажатию на кнопку «ВИТРИНА ИИ» будет открыто окно с демонстрацией функций ИИ-агента (см. Рис. 7.1.2). Подробнее см. п. «Витрина ИИ» текущего раздела.

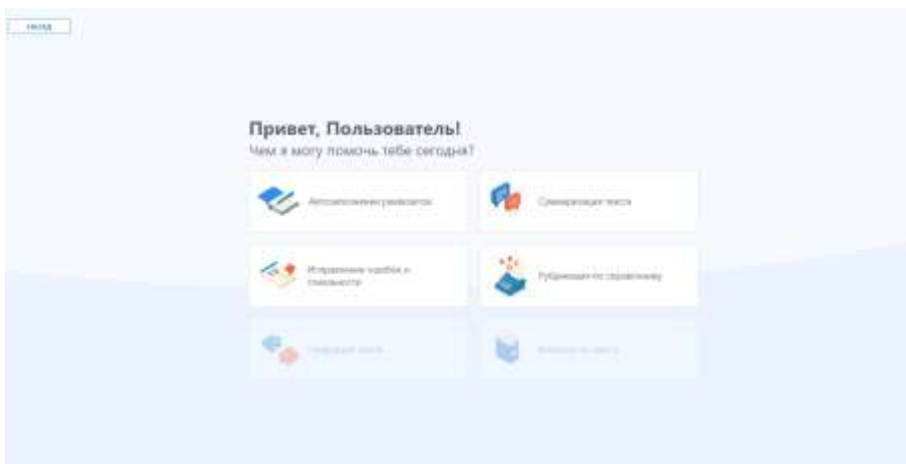


Рис. 7.1.2 – Витрина ИИ

Чтобы выйти из демонстрационного окна нажмите кнопку «НАЗАД»

НАЗАД

, пользователь будет возвращен на экран авторизации.

Для авторизации введите имя и пароль. Нажмите кнопку «ВОЙТИ». Откроется страница «Промпты» (см. Рис. 7.1.3). Подробнее см. п. «Раздел «Промпты»» текущего раздела.

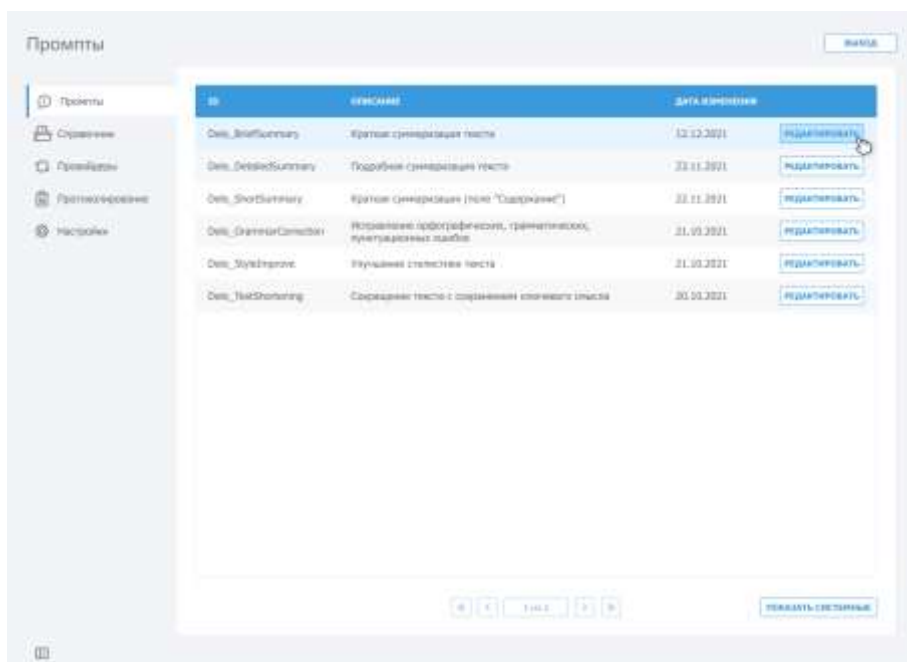


Рис. 7.1.3 – Раздел «Промпты»

## 7.2. Витрина ИИ

Витрина ИИ демонстрирует пользователю функции ИИ-агента - это beta-раздел для ознакомления с функционалом (см. Рис. 7.2.1).

Витрина ИИ состоит из перечня подразделов:

- Автозаполнение реквизитов,
- Суммаризация текста,
- Исправление ошибок и тональности,
- Рубрикация по справочнику,
- Генерация контента — отображается в заблокированном виде,
- Вопросы по тексту — отображается в заблокированном виде.

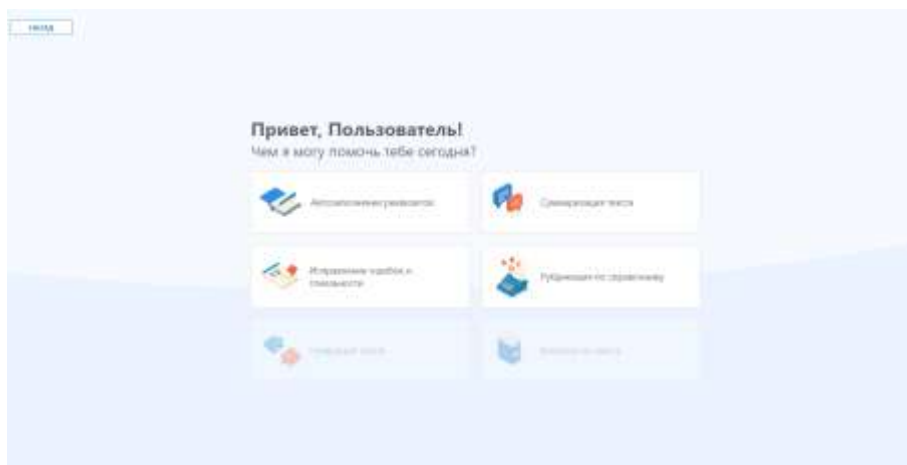


Рис. 7.2.1 – Витрина ИИ

## 7.3. Структура сайта агента

Сайт состоит из следующих разделов:

- Промпты,
- Справочники,
- Провайдеры,
- Протоколирование,
- Настройки.

Ниже будет подробнее описано про разделы сайта.

### 7.3.1. Раздел «Промпты»

В разделе «Промпты» отображается список промптов (запросов). Раздел включает в себя следующие реквизиты (см. Рис. 7.3.1.1):

- меню – перечень разделов,
- кнопка «Показать/Скрыть панель меню» ,
- кнопка «ВЫХОД» . По нажатию произойдет разлогирование пользователя. Пользователю будет показана страница «Авторизация» (подробнее см. п. «Авторизация в модуле «ИИ Агент»),
- ID - идентификатор промпта,
- Описание – описание промпта,
- Дата изменения - дата последнего изменения промпта. По умолчанию при установке ИИ Агента это дата создания конфиг-файла,
- кнопка «РЕДАКТИРОВАТЬ» ,
- блок навигации между страницами,
- кнопка «ПОКАЗАТЬ СИСТЕМНЫЕ».

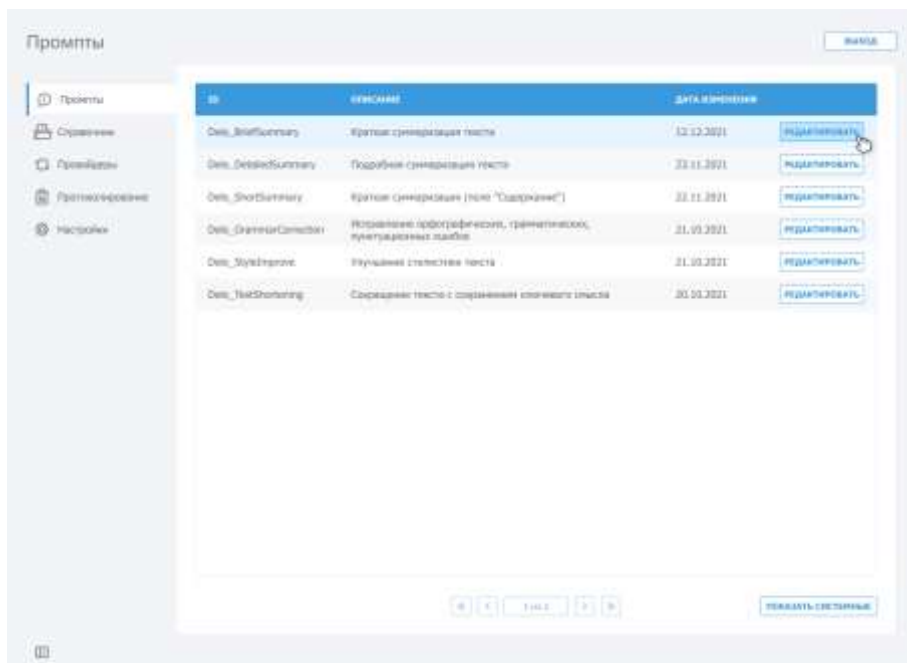


Рис. 7.3.1.1 – Раздел «Промпты»

По нажатию на кнопку «РЕДАКТИРОВАТЬ» открывается окно «Описание промпта» выбранного промпта, в котором пользователь может отредактировать промпт (см. Рис. 7.3.1.2).

Рис. 7.3.1.2 – Окно «Описание промпта»

Окно «Описание промпта» включает в себя следующие реквизиты:

- кнопка «Назад» . По нажатию пользователь вернется в раздел «Промпты»,
- кнопка «Показать/Скрыть панель меню» ,
- поле «Описание» обязательно для заполнения, заполняется администратором (максимум 200 символов) – краткое смысловое описание назначения промпта, позволяющее администратору идентифицировать его в общем списке,

- поле «ID, Идентификатор» - уникальный глобальный идентификатор записи в рамках системы. Не редактируется пользователем (максимум 50 символов), задается в конфиг-файле промпта. Используется для API-операций, логирования,
- кнопка «Копировать»  в поле «ID, Идентификатор». По нажатию на кнопку ID промпт будет скопирован в буфер, отображается сообщение «Идентификатор скопирован в буфер обмена.» (кн. «ЗАКРЫТЬ»),

**Примечание:**

Копирование ID,Идентификатора в буфер обмена может быть не доступно в некоторых браузерах/версиях браузеров.

- поле «Провайдер» с выпадающим списком названий провайдеров, добавленных в разделе «Провайдеры» (подробнее см. пп. «Провайдеры» текущего пункта документа). Указывает среду исполнения модели, через который будет осуществляться инференс.

Провайдер

Модель

**Примечание:**

Значение провайдера задается в конфиг-файле. Если его удаляют и не создают новый, то при переходе на страницу «Описание промпта» в поле «Провайдер» подставляется значение «Автоматически».

- поле «Модель» с выпадающим списком моделей - идентификатор конкретной LLM, развернутой у выбранного провайдера,

**Примечание:**

Выбор модели невозможен без выбора провайдера.

- кнопка «Проверить». Кнопка доступна, если в поле «Провайдер» выбрано конкретное значение и в поле «Модель» выбрано значение из выпадающего списка (поле не пустое),



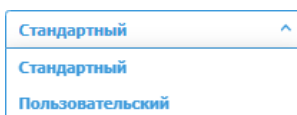
По нажатию выполняется проверка доступности модели. В зависимости от условий будет показано соответствующее сообщение, например, «Модель доступна», «Модель не найдена у провайдера» и др.

- поле «Промпт» с выпадающим списком типов промптов:
  - «Стандартный» выбран по умолчанию - предварительно сконфигурированный шаблон. Поле не предназначено для редактирования администратором, при этом поле «Текст промпта» заблокировано (нельзя редактировать),

**Примечание:**

Текст стандартного промпта может быть изменен через конфиг-файл промпта.

- «Пользовательский»,



Редактируемое текстовое поле. По умолчанию при выборе «Пользовательский» в поле «Текст промпта» будет скопирован текст промпта «Стандартный». Далее пользователь может изменить/редактировать текст промпта.

Если в поле «Промпт» выбран параметр «Пользовательский», тогда поле «Текст промпта» пользователь может редактировать, при этом поле обязательно для заполнения.

- поле «Текст промпта» - тело системного или пользовательского промпта, которое будет отправляться в модель. Обязательно для заполнения (максимум 8000 символов) и зависит от выбранного параметра в поле «Промпт». Если в поле «Промпт» выбран параметр «Стандартный», тогда поле «Текст промпта» нет возможности редактировать, отображается системный промпт.



#### Примечание:

Текст стандартного промпта может быть изменен через конфиг-файл промпта.

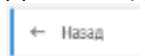
- блок «Проверка промпта» состоит из:
  - поле «Исходный текст»,
  - поле «Результат выполнения промпта»,
  - кнопка «ПРОТЕСТИРОВАТЬ»  доступна когда введен текст в поле «Исходный текст».

Для того, чтобы протестировать текст, пользователю необходимо ввести текст в поле «Исходный текст» и нажать на кнопку «ПРОТЕСТИРОВАТЬ». Если результат тестирования успешный, то будет показано сообщение «Тестирование промпта завершилось успешно.» (кн. «ЗАКРЫТЬ»), если с ошибкой, то будет показано сообщение «Тестирование завершилось с ошибкой: <текст ошибки>.» (кн. «ЗАКРЫТЬ»). Текст ошибки будет записан в протокол.

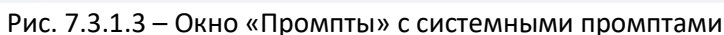
Результат будет отображен в поле «Результат проверки промпта».

- кнопка «ОТМЕНИТЬ» . По нажатию изменения не будут сохранены и окно «Описание промпта» закроется, произойдет возврат в окно «Промпты»,
- кнопка «СОХРАНИТЬ» . По нажатию корректные данные в полях сохранятся, в противном случае будет показано соответствующее сообщение, например «Необходимо выбрать провайдера.», «Укажите тип промпта.» и др.

Для выхода из окна «Описание промпта» нажмите кнопку «Назад»



По нажатию на кнопку «ПОКАЗАТЬ СИСТЕМНЫЕ» дополнительно отображаются системные промпты в окне «Промпты», в которых пользователь может изменить только применяемую модель(см. Рис. 7.3.1.3).



Раздел «Справочники» включает в себя следующие реквизиты (см. Рис. 5.3.2.1):

- меню – перечень разделов,
- кнопка «Показать/Скрыть панель меню» ,
- кнопка «ВЫХОД» . По нажатию произойдет разлогирование пользователя. Пользователю будет показана страница «Авторизация» (подробнее см. п. «Авторизация в модуле «ИИ Агент»),
- Описание
  - Справочник ССТУ

- Дата изменения – дата последнего изменения поля "Семантическое представление для модели". По умолчанию при установке ИИ Агента это дата создания конфиг-файла,
- кнопка «РЕДАКТИРОВАТЬ» . По нажатию открывается окно «Дерево рубрик» (см. Рис. 7.3.2.2),
- блок навигации между страницами.

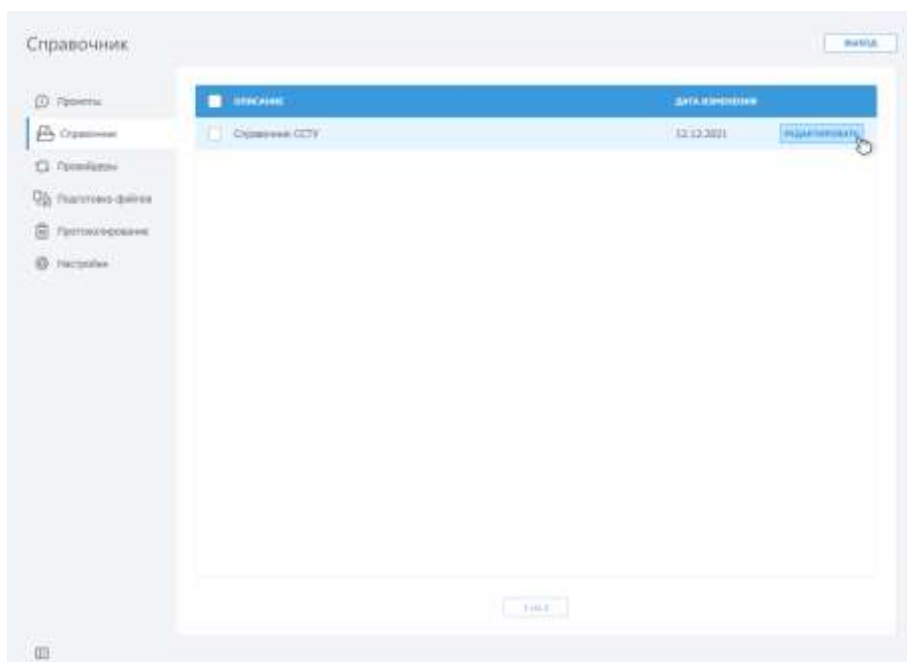


Рис. 7.3.2.1. Окно «Справочники»

В окне «Дерево рубрик» администратор может настроить семантическое представление для модели, нажав на кнопку «РЕДАКТИРОВАТЬ» в окне «Дерево рубрик справочника» и выбрав нужную рубрику, дополнив ее текстом. Текст вносится в поле

«Семантическое представление для модели». После внесения текста кнопка «СОХРАНИТЬ» станет активной.

После редактирования следует нажать на кнопку «СОХРАНИТЬ», при корректном заполнении отобразится сообщение «Описание успешно сохранено.» (кнопка «ЗАКРЫТЬ»), в противном случае будет показано соответствующее сообщение, например «Для рубрики описание обязательно.», «Не удалось сохранить описание. Проверьте подключение и попробуйте еще раз.» и др.

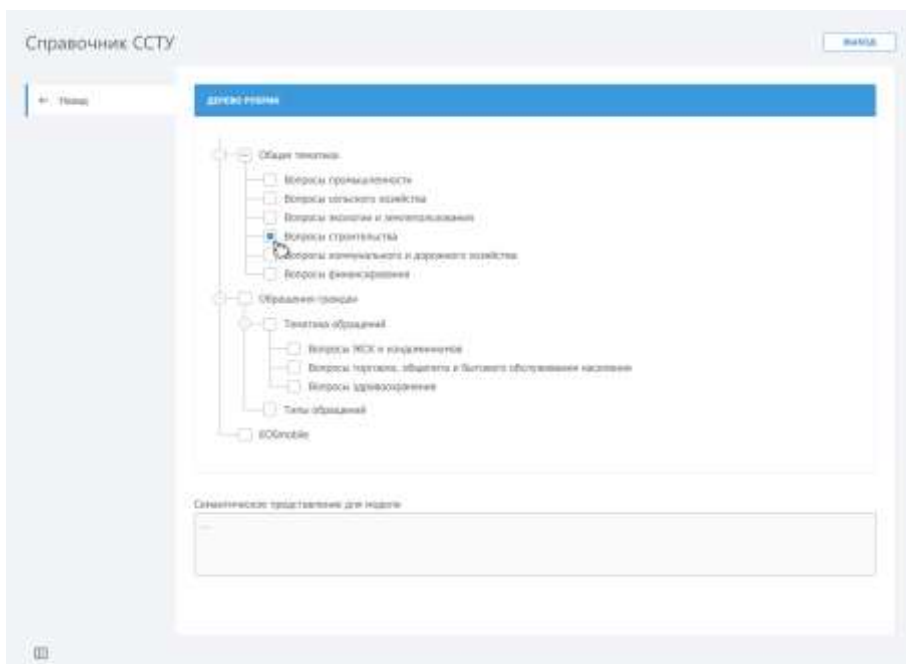


Рис. 7.3.2.2 – Дерево рубрик справочника

### 7.3.3. Раздел «Провайдеры»

В разделе «Провайдеры» отображается список провайдеров (см. Рис. 7.3.3.0.1). Раздел включает в себя следующие реквизиты:

- меню – перечень разделов,
- кнопка «Показать/Скрыть панель меню» ,
- кнопка «ВЫХОД» . По нажатию произойдет разлогирование пользователя. Пользователю будет показана страница «Авторизация» (подробнее см. п. «Авторизация в модуле «ИИ Агент»),
- флаговая кнопка ☐ – дает возможность выбрать записи из перечня,
- название провайдера,
- тип провайдера,
- кнопка «РЕДАКТИРОВАТЬ» . По нажатию откроется окно «Описание провайдера» (подробнее см. пп. «Окно «Описание провайдера» текущего пункта документа),
- кнопка «УДАЛИТЬ». Отображается неактивной (серой), если ни один провайдер не выбран. По нажатию на активную кнопку  будет показано сообщение «Внимание! Выбранные провайдеры будут удалены. Удалить?» (кнопки «ОТМЕНИТЬ» и «УДАЛИТЬ»),
- кнопка «ДОБАВИТЬ ПРОВАЙДЕРА» . По нажатию откроется окно «Описание провайдера» (подробнее см. п. «Окно «Описание провайдера» текущего пункта документа),
- блок навигации между страницами.

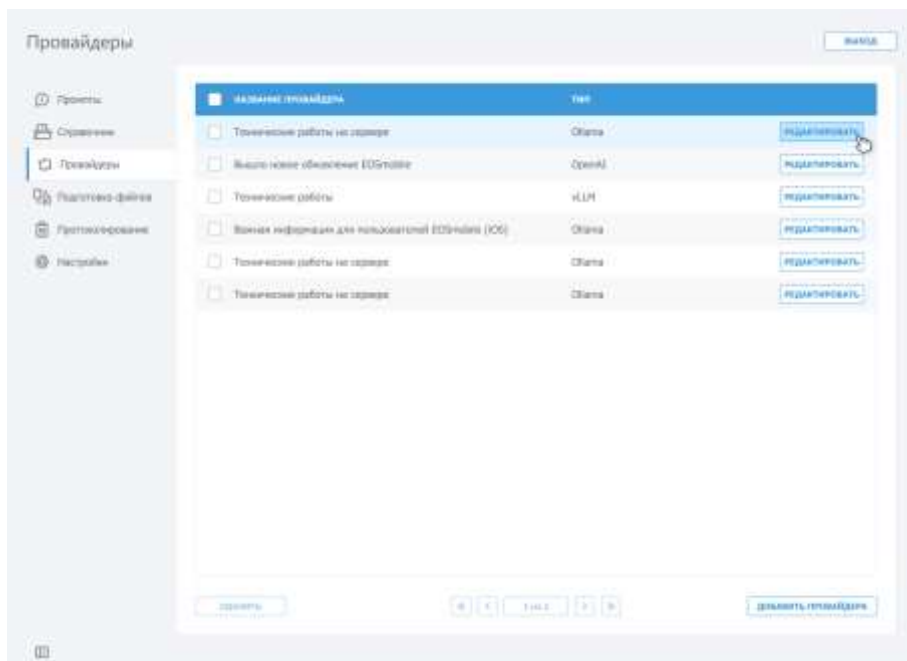


Рис. 7.3.3.0.1 – Раздел «Провайдеры»

### 7.3.3.1. Окно «Описание провайдера»

По нажатию на кнопку «ДОБАВИТЬ ПРОВАЙДЕРА» **ДОБАВИТЬ ПРОВАЙДЕРА** или «РЕДАКТИРОВАТЬ» **РЕДАКТИРОВАТЬ** в разделе «Провайдеры» отображается окно «Описание провайдера» (см. Рис. 7.3.3.1.1 и Рис. 7.3.3.1.2), включающее в себя следующие реквизиты:

- кнопка «Назад» **← назад**. По нажатию произойдет возврат в раздел «Провайдеры»,
- меню – перечень разделов,
- кнопка «Показать/Скрыть панель меню» **☰**,

- кнопка «ВЫХОД» . По нажатию произойдет разлогирование пользователя. Пользователю будет показана страница «Авторизация» (подробнее см. п. «Авторизация в модуле «ИИ Агент»),
- блок «Название провайдера» (максимально 100 символов). Обязательное поле,
- флаговая кнопка ☐ «Использовать провайдер по умолчанию». Если установлен флаг, то данный провайдер автоматически подставляется при создании нового промпта, если пользователь не выбрал другой,
- блок «Модель по умолчанию» с выпадающим списком. Список моделей обновляется динамически при смене типа провайдера и/или URL,
- блок «ID, Идентификатор» - уникальный глобальный идентификатор записи в рамках системы. После создания провайдера не редактируемое. Обязательное поле,
- блок «Тип» с выпадающим списком. Обязательное поле,



### Примечание:

Поддержка типа ollama сделана в ознакомительных целях. Корректная работа функционала поддерживается при выборе типа vllm.

- блок «URL» с текстовым полем «Введите URL» - базовый адрес API провайдера, используемый для отправки запросов. Может содержать путь и путь. Обязательное поле,
- блок «API\_KEY» с текстовым полем «В случае необходимости введите API\_KEY» - секретный ключ для авторизации запросов к провайдеру,
- «Количество токенов на входе» - максимально допустимая длина входного контекста, которое принимает провайдер, целое число до 10000. По умолчанию 32000. Обязательное поле,

- «Количество токенов на выходе» - максимальное количество токенов, которое разрешено сгенерировать модели в одном ответе, целое число до 10000. Значение может дополнительно ограничиваться на уровне самой LLM. По умолчанию 15000. Обязательное поле,
- кнопка «ОТМЕНИТЬ» . По нажатию изменения не будут сохранены и окно «Описание провайдера» закроется, произойдет возврат в окно «Провайдеры»,
- кнопка «СОХРАНИТЬ» . По нажатию при корректном заполнении отобразится сообщение «Описание успешно сохранено.» (кнопка «ЗАКРЫТЬ»), в противном случае будет показано соответствующее сообщение, например «Название провайдера обязательно.», «Необходимо выбрать тип провайдера.» и др.

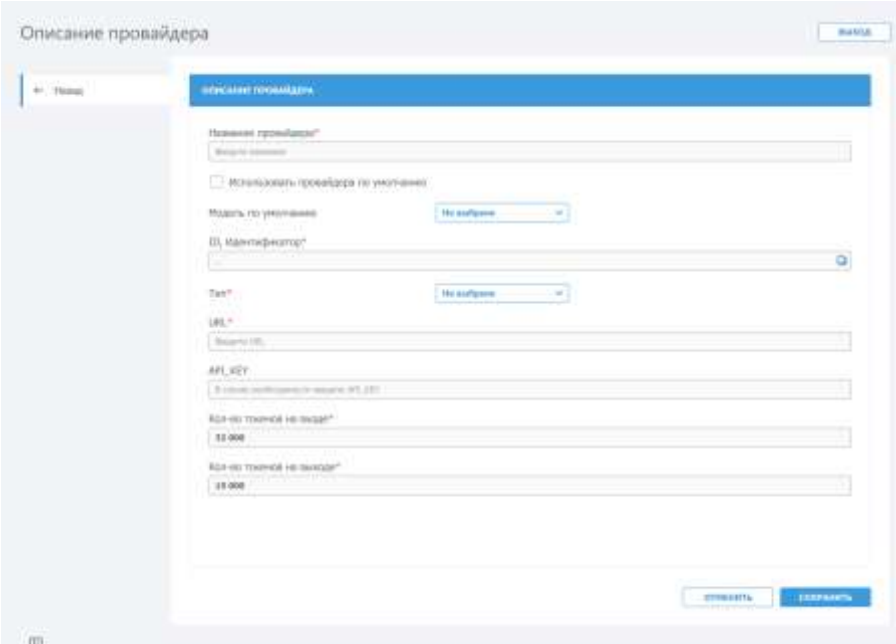


Рис. 7.3.3.1.1 – Окно «Описание провайдера» при нажатии на кнопку «ДОБАВИТЬ ПРОВАЙДЕРА»

The screenshot shows a web-based form titled 'Описание провайдера' (Description of provider). The form is divided into two main sections: a left sidebar with a 'Назад' (Back) button and a main content area. The main content area has a blue header bar with the text 'Описание провайдера'. Below this, there are several input fields and a checkbox. The fields are labeled: 'Название провайдера \*' (Provider name \*), 'ID провайдера \*' (Provider ID \*), 'URL \*' (URL \*), 'API\_KEY' (API key), 'Матрица токенов на входе' (Token matrix on input), and 'Матрица токенов на выходе' (Token matrix on output). The checkbox is labeled 'Использовать провайдера по умолчанию' (Use provider by default). There are also two buttons at the bottom right: 'Отменить' (Cancel) and 'Сохранить' (Save). The form is set against a light blue background with a sidebar on the left.

Рис. 7.6.3.1.2 – Окно «Описание провайдера» при нажатии на кнопку «РЕДАКТИРОВАТЬ»

## 7.3.4. Раздел «Протоколирование»

Раздел «Протоколирование» включает в себя следующие реквизиты (см. Рис. 7.3.4.1):

- меню – перечень разделов,
- кнопка «Показать/Скрыть панель меню» ,

- кнопка «ВЫХОД» . По нажатию произойдет разлогирование пользователя. Пользователю будет показана страница «Авторизация» (подробнее см. п. «Авторизация в модуле «ИИ Агент»),
- поле «Введите запрос» - поиск по всем видимым полям,
- кнопка «Очистить»  – по нажатию очищает строку поиска и сбрасывает поисковый запрос. Отображается при вводе значений,
- кнопка «ИСКАТЬ»  – по нажатию осуществлять поиск (обновление) поискового запроса,
- флаговая кнопка ☐ «Показать только ошибки». По нажатию будут отображены только ошибки,
- флаговая кнопка ☐ «Логирование». По умолчанию ВЫКЛ. При включении опции начинают логироваться отладочные данные,
- флаговая кнопка ☐ – дает возможность выбрать записи из перечня,
- событие – перечень событий,
- время – время событий протокола,
- Выбрано записей  - располагается над кн. «СОХРАНИТЬ» и показывает количество выбранных записей из перечня,
- кнопка «СОХРАНИТЬ»  – по нажатию предлагает сохранить выбранное событие или группу событий. При сохранении будет сформирован файл согласно выбранному в меню:
  - ВЫГРУЗИТЬ С УЧЕТОМ ФИЛЬТРОВ - выгружает записи протокола, попадающие под текущие значения фильтров,
  - ВЫГРУЗИТЬ ВЫДЕЛЕННОЕ - доступна только если выбрана хотя бы одна запись,
- блок навигации между страницами.

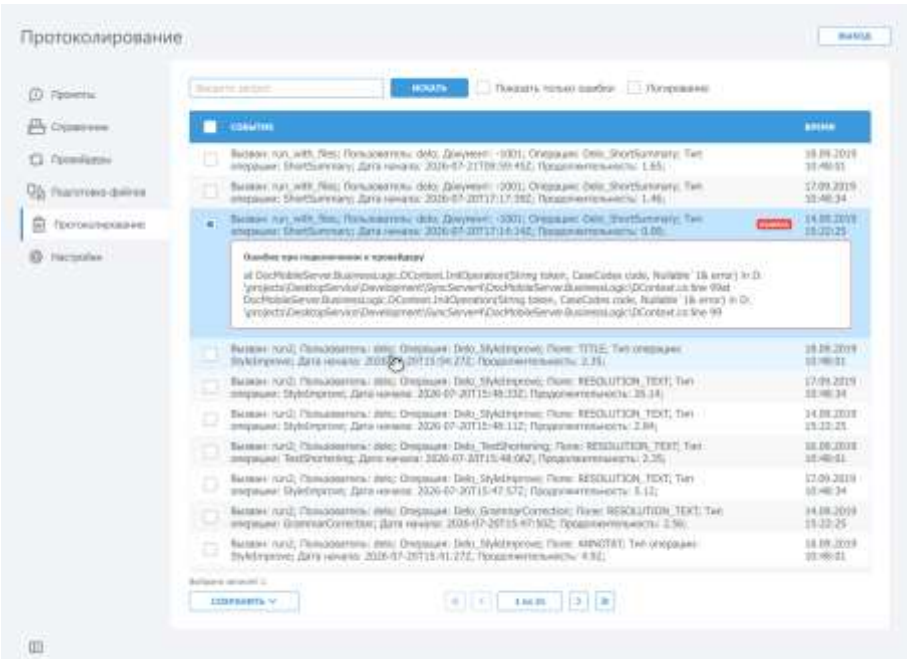


Рис. 7.3.4.1 – Окно «Протоколирование»

Если по событию произошла ошибка, то запись в протоколе будет окрашена красным с меткой «ОШИБКА» **ОШИБКА**. По нажатию на метку раскроется подробное описание ошибки.

7.3.5. Раздел «Настройки»

Раздел «Настройки» (см. Рис. 7.3.5.1) включает в себя следующие реквизиты:

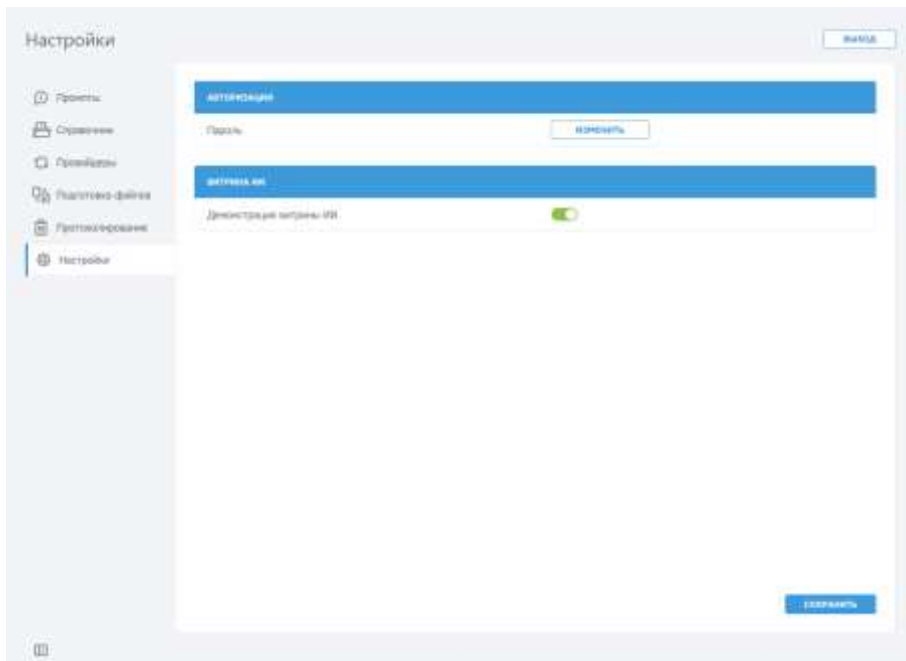
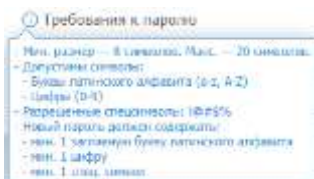


Рис. 7.3.5.1 – Окно «Настройки»

- меню – перечень разделов,
- кнопка «Показать/Скрыть панель меню» ,
- кнопка «ВЫХОД» . По нажатию произойдет разлогирование пользователя. Пользователю будет показана страница «Авторизация» (подробнее см. п. «Авторизация в модуле «ИИ Агент» ),
- блок «Авторизация»
  - кнопка «ИЗМЕНИТЬ»  в блоке «Авторизация». По нажатию отображается окно «ИЗМЕНИТЬ ПАРОЛЬ» с реквизитами:
    - поле «Текущий пароль» для ввода текущего пароля,
    - поле «Новый пароль» для ввода нового пароля (требования к паролю см. ниже),

- поле «Подтверждение нового пароля» для ввода нового пароля,
- подсказка «Требования к паролю». По нажатию отображаются необходимые требования к паролю,



- кнопка «ИЗМЕНИТЬ»  в окне «ИЗМЕНИТЬ ПАРОЛЬ». По нажатию новый корректный пароль сохраняется, окно «ИЗМЕНИТЬ ПАРОЛЬ» закрывается, отображается сообщение «Новый пароль сохранен.» (кнопка «ЗАКРЫТЬ»), в противном случае будет показано соответствующее сообщение, например «Пароль должен содержать как минимум одну заглавную букву.», «Новые пароли не совпадают.» и др.
- кнопка «ОТМЕНИТЬ»  . По нажатию не сохраняет данные, окно «ИЗМЕНИТЬ ПАРОЛЬ» закрывается.

Окно «ИЗМЕНИТЬ ПАРОЛЬ» с заголовком и кнопкой закрытия. В нем три поля ввода: «Текущий пароль», «Новый пароль» и «Подтверждение нового пароля», каждое со значком «показать/скрыть» (глазик). Ниже поле «Требования к паролю» со значком информации. Внизу две кнопки: «ИЗМЕНИТЬ» (синяя) и «ОТМЕНИТЬ» (серая).

Для изменения пароля введите текущий пароль, новый и подтвердите новый пароль повторным введением.

- блок «Витрина ИИ»
  - параметр «Демонстрация витрины ИИ». По умолчанию — ВКЛ. Если ВЫКЛ и нажать на кнопку «СОХРАНИТЬ», то кнопка «ВИТРИНА ИИ» будет скрыта в окне авторизации (см. Рис. 7.3.5.2),
- кнопка «СОХРАНИТЬ»  . По нажатию все изменения сохраняются.



Рис. 7.3.5.2 - Окно авторизации без кнопки «ВИТРИНА ИИ»