Руководство пользователя SmartMLOps

Версия SmartMLOps: 0.13.0.

Аннотация

Руководство описывает вход, рабочий контекст, поиск и развертывание сервисов, запросы, сценарии больших языковых моделей, конвейеры и мониторинг. Для внешнего вызова откройте карточку сервиса и отправьте запрос на адрес, опубликованный в ее разделе `API`.

С чего начать

- Войти и проверить защищенный раздел: «Войти в SmartMLOps».
- Выбрать сервис с понятным результатом: «Выбрать один проверяемый пример».
- Выбрать способ запроса: «Выбрать способ запроса к сервису».
- Завершить работу с временными объектами: «Завершить работу с временным объектом».

Содержание

Руководство пользователя SmartMLOps 1
Аннотация 1
С чего начать 1
Назначение программы 6
Назначение 6
Пользовательские результаты 6
Границы роли 6
Ограничения текущей версии 6
Условия выполнения программы 6
Обязательные условия 6
Проверка и граница 7
Войти в SmartMLOps 7
Перед началом 7
Выполните 7
Проверьте результат 7
Если не получилось 7
Завершите работу 8
Проверить рабочий контекст 8
Перед началом 8
Выполните 8
Проверьте результат 8
Если не получилось 8
Завершите работу 8
Выполнение программы 9
Выберите результат 9
Разделы пользовательского портала 9
Назначение 9
Разделы 9
Границы 9
Связанные действия 10
Выбрать один проверяемый пример 10
Перед началом 10
Выполните 10
Проверьте результат 10

Если не получилось 10
Завершите работу 11
Запросить доступ к сервису 11
Перед началом 11
Выполните 11
Проверьте результат 11
Если не получилось 11
Завершите работу 12
Выбрать источник модели машинного обучения 12
Выберите один путь 12
Развернуть модель из загружаемых файлов 12
Перед началом 12
Выполните 12
Проверьте результат 12
Если не получилось 12
Завершите работу 13
Развернуть модель по разрешенной ссылке 13
Перед началом 13
Выполните 13
Проверьте результат 13
Если не получилось 13
Завершите работу 14
Выбрать среду выполнения большой языковой модели 14
Выберите один путь 14
Развернуть модель в среде `llama.cpp` 14
Перед началом 14
Выполните 14
Проверьте результат 15
Если не получилось 15
Завершите работу 15
Развернуть модель в среде `vLLM` 15
Перед началом 15
Выполните 15
Проверьте результат 15
Если не получилось 16
Завершите работу 16
Развернуть собственный сетевой сервис 16
Перед началом 16
Выполните 16
Проверьте результат 16
Если не получилось 17
Завершите работу 17
Выбрать способ запроса к сервису 17
Выберите один результат 17
Выполнить запрос к опубликованному сервису 17
Перед началом 17
Выполните 17
Проверьте результат 18
Если не получилось 18

Завершите работу ... 18
Выполнить запрос к собственному сервису ... 18
Перед началом ... 18
Выполните ... 18
Проверьте результат ... 19
Если не получилось ... 19
Завершите работу ... 19
Получить пример вызова программного интерфейса ... 19
Перед началом ... 19
Выполните ... 19
Проверьте результат ... 19
Если не получилось ... 19
Завершите работу ... 20
Создать сценарий большой языковой модели ... 20
Перед началом ... 20
Выполните ... 20
Проверьте результат ... 20
Если не получилось ... 20
Завершите работу ... 20
Добавить шаг в сценарий ... 21
Перед началом ... 21
Выполните ... 21
Проверьте результат ... 21
Если не получилось ... 21
Завершите работу ... 21
Выполнить тестовый запуск сценария ... 21
Перед началом ... 21
Выполните ... 21
Проверьте результат ... 22
Если не получилось ... 22
Завершите работу ... 22
Открыть внешний доступ к сценарию ... 22
Перед началом ... 22
Выполните ... 22
Проверьте результат ... 22
Если не получилось ... 22
Завершите работу ... 23
Выдать ключ сценария ... 23
Перед началом ... 23
Выполните ... 23
Проверьте результат ... 23
Если не получилось ... 23
Завершите работу ... 23
Вызвать сценарий через программный интерфейс ... 23
Перед началом ... 23
Выполните ... 23
Проверьте результат ... 24
Если не получилось ... 24
Завершите работу ... 24

Создать конвейер из готового шаблона 24
Перед началом 24
Выполните 24
Проверьте результат 25
Если не получилось 25
Завершите работу 25
Создать конвейер с пустого графа 25
Перед началом 25
Выполните 25
Проверьте результат 25
Если не получилось 26
Завершите работу 26
Запустить сохраненный конвейер 26
Перед началом 26
Выполните 26
Проверьте результат 26
Если не получилось 26
Завершите работу 26
Прочитать качество и дрейф модели 27
Перед началом 27
Выполните 27
Проверьте результат 27
Если не получилось 27
Завершите работу 27
Загрузить правильные ответы модели 28
Перед началом 28
Выполните 28
Проверьте результат 28
Если не получилось 28
Завершите работу 28
Прочитать показатели большой языковой модели 28
Перед началом 28
Выполните 28
Проверьте результат 29
Если не получилось 29
Завершите работу 29
Завершить работу с временным объектом 29
Выберите один результат 29
Границы текущей версии 30
Архивировать временный сервис 30
Перед началом 30
Выполните 30
Проверьте результат 30
Если не получилось 30
Завершите работу 30
Восстановить сервис из архива 31
Перед началом 31
Выполните 31
Проверьте результат 31

Если не получилось ... 31
Завершите работу ... 31
Удалить архивный сервис навсегда ... 31
Перед началом ... 31
Выполните ... 31
Проверьте результат ... 32
Если не получилось ... 32
Завершите работу ... 32
Отменить ожидающую заявку ... 32
Перед началом ... 32
Выполните ... 32
Проверьте результат ... 32
Если не получилось ... 32
Завершите работу ... 32
Отозвать ключ собственного сервиса ... 32
Перед началом ... 32
Выполните ... 33
Проверьте результат ... 33
Если не получилось ... 33
Завершите работу ... 33
Отозвать один ключ сценария ... 33
Перед началом ... 33
Выполните ... 33
Проверьте результат ... 33
Если не получилось ... 33
Завершите работу ... 34
Закрыть внешний доступ к сценарию ... 34
Перед началом ... 34
Выполните ... 34
Проверьте результат ... 34
Если не получилось ... 34
Завершите работу ... 34
Отменить выполняющийся запуск конвейера ... 34
Перед началом ... 34
Выполните ... 34
Проверьте результат ... 35
Если не получилось ... 35
Завершите работу ... 35
Завершить пользовательский сеанс ... 35
Перед началом ... 35
Выполните ... 35
Проверьте результат ... 35
Если не получилось ... 35
Завершите работу ... 36
Сообщения пользователю ... 36
Выберите страницу ... 36
Сообщения пользователя ... 36
Как пользоваться справкой ... 36
Записи ... 36

Границы 37
Минимальная запись 37
Кому передать пользовательскую проблему 37
Назначение 37
Адресаты 38
Отказы при передаче 38
Минимальная запись 38
Лист регистрации изменений 39

Назначение программы

Назначение

SmartMLOps дает пользователю портал и разрешенный программный интерфейс для работы с опубликованными сервисами искусственного интеллекта. Доступное действие зависит от учетной записи и рабочего контекста арендатора.

Пользовательские результаты

Через портал пользователь может:

- найти сервис в разделе `Сервисы` на вкладке `Каталог`, отправить заявку и проверить ее в разделе `Мои заявки`;
- посмотреть выданный доступ и потребление в разделе `Мой доступ`;
- создать поддерживаемый сервис модели машинного обучения, большой языковой модели или собственного сетевого сервиса;
- отправить тестовый запрос и получить пример внешнего вызова;
- создать и запустить сценарий большой языковой модели или конвейер;
- прочитать качество, дрейф, события, трассировки, число токенов и задержку, если сервис передает эти данные.

Результат развертывания проверяется по карточке сервиса и состоянию операции. Результат запроса - по конечному состоянию и ответу. Для конвейера портал показывает состояние запуска и его шагов. Для мониторинга нужны значение показателя, время расчета и состояние источника.

Границы роли

Пользователь не устанавливает платформу, не настраивает кластер, сеть, хранилище и поставочные секреты. Он также не меняет общие роли и квоты и не правит исходный код. Адресат для такой работы указан на странице «Кому передать пользовательскую проблему».

Ограничения текущей версии

- Рабочий контекст поступает из сеанса входа. Отдельного переключателя для роли `user` нет.
- Пункт меню не подтверждает наличие объекта или разрешения.
- Примеры `tabular`, классификации текста и `OCR` обозначают тип результата. Они не подтверждают, что в организации опубликован отдельный готовый продукт.
- Удалить сценарий большой языковой модели или конвейер через портал нельзя.

Условия выполнения программы

Обязательные условия

До начала операции проверьте:

- адрес портала получен через утвержденный канал организации и открывается по защищенному соединению;
- учетная запись активна, а вход возвращает пользователя в портал;
- сеанс содержит рабочий контекст арендатора;
- учетной записи разрешено выбранное действие;
- нужный сервис, модель или шаблон опубликован и доступен;
- тестовые данные не содержат персональных данных, секретов и сведений ограниченного распространения;
- безопасное название отличает временный объект от рабочего;
- результат можно проверить по состоянию, ответу, идентификатору или файлу результата.

Проверка и граница

Вход описан на странице «Войти в SmartMLOps». Наличие рабочего контекста проверяется на странице «Проверить рабочий контекст».

Если условие не выполнено, остановите выбранную операцию и определите адресата по странице «Кому передать пользовательскую проблему». Установка компонентов и изменение инфраструктуры не заменяют пользовательское предварительное условие.

Войти в SmartMLOps

Перед началом

- *Проверьте.* Адрес портала получен через утвержденный канал, открывается по защищенному соединению и показывает название SmartMLOps. Учетная запись активна. При отказе учетной записи обратитесь к роли `portal-admin`; при общей недоступности портала - к роли `system-admin`.
- *Подготовьте.* Адрес вида `<адрес-портала>`, учетные данные организации и второй фактор, если его требует поставщик удостоверений. Не записывайте пароль, одноразовый код или токен в обращение.

Выполните

1. Откройте `<адрес-портала>`.
2. В правой части шапки нажмите `Перейти ко входу`. Защищенный раздел может сначала показать страницу `Войдите, чтобы продолжить работу`.
3. Выполните вход на странице поставщика удостоверений по правилам организации.
4. Дождитесь завершения страницы `Подтверждаем вход`. Портал вернет вас на исходную страницу.
5. Откройте меню учетной записи в правой части шапки. Ожидаемый результат: в меню есть команда `Выйти`.

Проверьте результат

Откройте `Сервисы`, затем вкладку `Мои решения`. Страница должна показать список, пустое состояние или предметное сообщение о правах. Она не должна снова показывать `Войдите, чтобы продолжить работу`.

Если не получилось

- `Не удалось начать вход`. Повторите вход один раз. При повторе сохраните время и адрес страницы и передайте их роли `system-admin`.

- `Не удалось завершить вход`. Нажмите `Попробовать снова`. При повторе сохраните время и адрес без параметров запроса и передайте их роли `portal-admin`. Код возврата из адресной строки не копируйте.
- После возврата показан `Гость`. Разрешите обязательные файлы `cookie` для адреса портала и повторите вход. При повторе передайте время роли `system-admin`.
- Ответ `401`. Сеанс истек или не обновился. Выполните выход и вход заново.
- Ответ `403`. Вход выполнен, но право или рабочий контекст не подходит. Сохраните идентификатор обращения и передайте его роли `portal-admin`.

Завершите работу

На общем устройстве откройте меню учетной записи, нажмите `Выйти` и дождитесь подписи `Гость` и кнопки `Перейти ко входу`. Затем закройте вкладку браузера.

Проверить рабочий контекст

Перед началом

- *Проверьте.* Вход выполнен. Получите у роли `portal-admin` название ожидаемой организации или арендатора. Отдельного переключателя контекста для роли `user` нет: поле `tenant_id` поступает в сеанс при входе.
- *Подготовьте.* Название одного известного решения или опубликованного сервиса, если такой объект должен существовать в этом контексте.

Выполните

1. Откройте раздел `Сервисы`.
2. Выберите вкладку `Мои решения` и найдите известное решение, если оно должно существовать.
3. Откройте вкладку `Каталог` и найдите известный опубликованный сервис, если он должен быть доступен организации.

Ожидаемый результат: обе вкладки загружаются без сообщения `Активный арендатор не выбран` и без повторного запроса входа.

Проверьте результат

Успешная загрузка подтверждает наличие рабочего контекста, но не его точный идентификатор. Портал не показывает и не меняет `tenant_id` в шапке. Для проверки точного идентификатора передайте роли `portal-admin` имя учетной записи и ожидаемое название организации.

Если не получилось

- `Активный арендатор не выбран`. Выполните выход и вход один раз. При повторе передайте время входа роли `portal-admin`.
- Известного решения нет, но ошибки нет. Проверьте вкладку `Мои решения`, а не `Каталог`. Если решение должно принадлежать вам, передайте его название роли `portal-admin`.
- Каталог пуст. Это может означать, что в текущем контексте нет опубликованных сервисов; пустой каталог не подтверждает отказ платформы.
- `403 Недостаточно прав`. Сохраните действие, время и идентификатор обращения и передайте их роли `portal-admin`.

Завершите работу

Не меняйте `tenant_id` в адресе, запросе или хранилище браузера. После подтверждения контекста выберите нужный раздел на странице «Разделы пользовательского портала».

Выполнение программы

Выберите результат

- Найти готовый сервис или получить доступ: «Выбрать один проверяемый пример» и «Запросить доступ к сервису».
- Создать модель машинного обучения: «Выбрать источник модели машинного обучения».
- Создать большую языковую модель: «Выбрать среду выполнения большой языковой модели».
- Создать собственный сетевой сервис: «Развернуть собственный сетевой сервис».
- Получить ответ: «Выбрать способ запроса к сервису».
- Создать сценарий большой языковой модели: «Создать сценарий большой языковой модели».
- Создать конвейер: «Создать конвейер из готового шаблона» или «Создать конвейер с пустого графа».
- Прочитать показатели: «Прочитать качество и дрейф модели» или «Прочитать показатели большой языковой модели».
- Очистить временный объект: «Завершить работу с временным объектом».

Промежуточные состояния `Разворачивается`, `Выполняется` и `На рассмотрении` не подтверждают успех или ошибку. Конечное состояние указано на странице выбранного действия.

Разделы пользовательского портала

Назначение

Открывайте раздел через левое меню. На узком экране меню закрывается после перехода; кнопка в левом верхнем углу имеет подпись `Развернуть левое меню` или `Свернуть левое меню`. Название SmartMLOps в шапке возвращает в `Сервисы` на вкладку `Каталог`.

Разделы

- `Каталог` - опубликованные сервисы текущей организации.
- `Мои решения` - сервисы, которыми может управлять текущая учетная запись.
- `Мои заявки` - запросы доступа и их состояния.
- `Мой доступ` - активные ключи, потребление и ограничения за 30 дней.
- `Мониторинг машинного обучения` - качество, дрейф и события моделей.
- `Панели управления` - доступные ролевые панели и аудит без прямого перехода во внешнюю систему.
- `Сценарии больших языковых моделей` - создание, изменение и тестовый запуск сценария.
- `Мониторинг больших языковых моделей (LLM)` - трассировки, токены и задержка.
- `Конструктор конвейеров (pipeline)` - шаблоны, граф, сохранение и запуски.
- Меню учетной записи - имя текущей учетной записи и команда `Выйти`.

Разделы мониторинга, сценариев и конвейеров находятся в группе `Платформа`.

Границы

Пункт меню подтверждает наличие страницы, но не данных и не права на действие. Отдельный пункт уведомлений и переключатель организации для роли `user` не реализованы. Неизвестный адрес портал перенаправляет в `Сервисы`, поэтому прямой адрес не заменяет путь по меню.

Связанные действия

Для первого результата откройте «Выбрать один проверяемый пример». Если пункт приглушен и подсказка сообщает `Войдите, чтобы продолжить работу`, сначала выполните вход.

Выбрать один проверяемый пример

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Откройте `Сервисы`. Тестовые данные не должны содержать персональные данные, коммерческую тайну или секреты.
- *Подготовьте.* Выберите один безопасный ввод: вымышленную табличную запись по контракту сервиса, короткую вымышленную фразу для классификации текста или синтетическое изображение `PNG` либо `JPEG` для распознавания текста.

Выполните

1. На вкладке `Каталог` введите в поле `Поиск по названию и описанию` термин, связанный с подготовленным вводом.
2. Откройте одну подходящую карточку.
3. Прочитайте `Тип данных`, `Режим доступа` и `Готовность`.
4. Найдите раздел `API` и сверяйте поля подготовленного ввода только с контрактом этой карточки.
5. До запроса запишите ожидаемый вид результата. Для табличных данных это числовой ответ, класс или оценка. Для текста - метка и объявленная контрактом уверенность. Для распознавания - текст, структурированный `JSON` или скачиваемый файл.

Ожидаемый результат: открыта карточка одного сервиса, а вход и выход определены до отправки запроса.

Проверьте результат

Карточка должна показывать доступный сервис или понятный порядок получения доступа. Подготовленный ввод соответствует ее полям, а результат можно проверить по ответу, состоянию операции или скачиваемому файлу.

Названия `tabular`, классификация текста и `OCR` обозначают группы примеров. Они не подтверждают, что организация уже опубликовала отдельный готовый сервис. Готовность зависит от карточки, совместимого профиля выполнения и успешной серверной проверки.

Классификация текста требует доступной модели семейства `BERT`, совместимого профиля выполнения и прав. Распознавание текста требует опубликованного сервиса, проверенного пакета модели или проверенного образа контейнера.

Если не получилось

- `Ничего не найдено`. Сбросьте фильтры и измените поисковую фразу. Если каталог пуст, запросите название доступного сервиса у роли `portal-admin`.
- Готовность указана как `экспериментально`. Не используйте карточку как доказательство промышленной готовности.
- Раздела `API` нет. Не придумывайте адрес или тело запроса для этого сервиса.
- Сервер отклоняет тестовые данные. Сверьте поля со страницей `Вызов`; поля другого примера не добавляйте.

Завершите работу

Если карточка найдена, откройте «Запросить доступ к сервису». Если подходящего сервиса нет, выберите одну инструкцию развертывания по типу подготовленного источника.

Запросить доступ к сервису

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Получите название или назначение сервиса. Не создавайте новую заявку, если действующая уже есть.
- *Подготовьте.* Поисковую фразу, причину и цель без персональных данных, ключей и содержимого рабочих запросов. Используйте безопасные формулировки `Проверка интеграции в учебном контуре` и `Отправка синтетических запросов для проверки формата ответа`.

Выполните

1. В разделе `Сервисы` откройте вкладку `Каталог`.
2. Введите фразу в поле `Поиск по названию и описанию` и нажмите `Открыть` в строке нужного сервиса.
3. Прочитайте режим доступа и готовность. Надпись `Одобряется автоматически` означает, что открытый сервис может одобрить заявку без решения владельца.
4. Нажмите `Запросить доступ`.
5. В поля `Причина запроса` и `Цель использования` введите подготовленные значения.
6. Нажмите `Отправить заявку`.

Ожидаемый результат: портал показывает `Заявка отправлена` и открывает список `Мои заявки`.

Проверьте результат

Найдите заявку по названию сервиса. Значение каждого состояния указано в списке.

- `на рассмотрении` - требуется решение владельца;
- `одобрено` - доступ активен;
- `отклонено` - доступ не выдан;
- `отменено` - пользователь отменил заявку;
- `отозвано` - ранее выданный доступ прекращен.

Успех этого действия подтверждают первые два состояния.

Если не получилось

- `Ничего не найдено`. Нажмите `Сбросить фильтры` и измените фразу.
- `У вас уже есть активная заявка`. Нажмите `Перейти к заявке`; дубль не создавайте.
- `Не удалось отправить заявку`. Нажмите `Повторить отправку` один раз. При повторе сохраните идентификатор обращения.
- `Недостаточно прав`. Сохраните сервис, время и идентификатор обращения и передайте их роли `portal-admin`.
- `Сервис не найден`. Карточка могла быть снята с публикации. Вернитесь в каталог и повторите поиск.

Завершите работу

Пока заявка имеет статус на рассмотрении, ее можно открыть в Мои заявки, нажать Отменить заявку и подтвердить действие в окне Отмена заявки. После статуса одобрено откройте «Выполнить запрос к сервису через портал».

Выбрать источник модели машинного обучения

Выберите один путь

- Файлы модели, каталог или небольшой архив ZIP: «Развернуть модель из загружаемых файлов».
- Разрешенная ссылка вида https://huggingface.co/<владелец>/<модель>: «Развернуть модель по разрешенной ссылке».

Оба пути используют серверное описание и проверку. Предварительный показ в браузере не подтверждает формат и возможность развертывания. Готовый пакет модели в эту книгу не входит.

Развернуть модель из загружаемых файлов

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Учетной записи разрешено создавать решения. Источник не содержит обучающие данные, персональные данные и секреты.
- *Подготовьте.* Один набор файлов модели, каталог модели или небольшой архив ZIP. Состав должен отвечать профилю выбранной модели. Сервер определяет итоговый формат и возможность развертывания.

Выполните

1. В разделе Сервисы на вкладке Каталог нажмите Развернуть новый сервис и выберите Модель машинного обучения.
2. В блоке Откуда взять модель? выберите Загрузить архив.
3. Нажмите Выберите файлы модели либо Или выберите каталог модели целиком и выберите подготовленный источник.
4. Дождитесь блока Описание с сервера; прочитайте возвращенные формат и модальность.
5. Выберите предложенную модальность и нажмите Подтвердить модальность.
6. В поле Какой шаблон модели использовать? выберите совместимый шаблон.
7. Выберите размещение на центральном или графическом процессоре и дождитесь блока Рекомендованный сервером профиль.
8. При необходимости измените название решения. После сообщения Все проверки пройдены нажмите Развернуть проверенную модель.

Ожидаемый результат: портал открывает вкладку Развертывания созданного решения.

Проверьте результат

На вкладке Развертывания нажмите Обновить состояние. Успех подтверждают состояние Готов, заголовок <название> готов к работе, непустой идентификатор операции и точка вызова. Состояние Разворачивается промежуточное; сохраните идентификатор операции и дождитесь конечного состояния.

Если не получилось

- Модель еще проверяется на сервере. Дождитесь проверки; развертывание повторно не отправляйте.

- `Не подтвержден тип данных модели`. Выберите модальность из ответа сервера и подтвердите ее.
- `Требуется решение модератора`. Сохраните идентификатор запроса и дождитесь решения роли `portal-admin`.
- `Для этого пакета не найдена совместимая среда выполнения`. Выберите предложенный профиль или другой источник. Если вариантов нет, передайте формат и идентификатор модели роли `portal-admin`.
- Состояние `Сбой`. Сохраните идентификатор операции, код блокирующего условия и время. Нажимайте `Повторить развертывание` только после устранения названного условия.

Завершите работу

После теста поместите временное решение в архив по странице «Архивировать временный сервис». Для проверки готового решения откройте «Выполнить запрос к сервису через портал».

Развернуть модель по разрешенной ссылке

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Учетной записи разрешено создавать решения. Правила организации и лицензия разрешают источник.
- *Подготовьте.* Разрешенную ссылку вида `https://huggingface.co/<владелец>/<модель>`. Не добавляйте токен или другой секрет к ссылке.

Выполните

1. В разделе `Сервисы` на вкладке `Каталог` нажмите `Развернуть новый сервис` и выберите `Модель машинного обучения`.
2. В блоке `Откуда взять модель?` выберите `Ссылка на Hugging Face`.
3. В поле `Ссылка на репозиторий модели` введите подготовленную ссылку.
4. Дождитесь блока `Описание с сервера`; прочитайте возвращенные формат и модальность.
5. Выберите предложенную модальность и нажмите `Подтвердить модальность`.
6. В поле `Какой шаблон модели использовать?` выберите совместимый шаблон.
7. Выберите размещение на центральном или графическом процессоре и дождитесь блока `Рекомендованный сервером профиль`.
8. При необходимости измените название решения. После сообщения `Все проверки пройдены` нажмите `Развернуть проверенную модель`.

Ожидаемый результат: портал открывает вкладку `Развертывания` созданного решения.

Проверьте результат

На вкладке `Развертывания` нажмите `Обновить состояние`. Успех подтверждают состояние `Готов`, заголовок `<название> готов к работе`, непустой идентификатор операции и точка вызова. Состояние `Разворачивается` промежуточное; сохраните идентификатор и дождитесь конечного состояния.

Если не получилось

- `Модель еще проверяется на сервере`. Дождитесь проверки; развертывание повторно не отправляйте.
- `Не подтвержден тип данных модели`. Выберите модальность из ответа сервера и подтвердите ее.

- `Требуется решение модератора`. Сохраните идентификатор запроса и дождитесь решения роли `portal-admin`.
- `Для этого пакета не найдена совместимая среда выполнения`. Выберите предложенный профиль или другую модель. Если вариантов нет, передайте формат и идентификатор модели роли `portal-admin`.
- `Активный арендатор не выбран` или состояние `Сбой`. Сохраните время, идентификатор операции и код условия и передайте их роли `portal-admin`.

Завершите работу

После теста поместите временное решение в архив по странице «Архивировать временный сервис». Для проверки готового решения откройте «Выполнить запрос к сервису через портал».

Выбрать среду выполнения большой языковой модели

Выберите один путь

- Источник содержит файл `.gguf`: Развернуть модель в этой среде.
- Источник совместим с `vLLM`: Развернуть модель в этой среде.

Текущий экран разворачивает новую модель по ссылке. Отдельного пути `Создать маршрут к готовой большой языковой модели` в форме нет. Форма не принимает учетные данные для закрытого источника; доступ к нему заранее готовит роль `portal-admin`.

Параметры выполнения поступают из активного каталога. Если блок пуст, имена параметров нельзя вводить по догадке. Выбор выделенного графического процессора создает заявку; повторный запуск допустим после одобрения роли `portal-admin`. Инструкции ниже используют совместно используемый графический процессор.

Развернуть модель в среде `llama.cpp`

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Учетной записи разрешено создавать решения. Лицензия и правила организации разрешают модель.
- *Подготовьте.* Название, разрешенную ссылку вида `https://huggingface.co/<владелец>/<модель>` и точное имя файла с расширением `.gguf`. Токен к ссылке не добавляйте.

Выполните

1. В разделе `Сервисы` на вкладке `Каталог` нажмите `Развернуть новый сервис` и выберите `Большая языковая модель`.
2. Заполните `Название сервиса` и `Ссылка на модель на платформе Hugging Face`.
3. В поле `Размещение графического процессора (GPU)` оставьте `Совместно используемый графический процессор (GPU)`.
4. Нажмите `Расширенные настройки` и выберите `llama.cpp (KServe)` в поле `Среда выполнения`.
5. В поле `Файл модели в формате GGUF` введите точное имя файла.
6. Нажмите `Рассчитать потребление памяти`.
7. Прочитайте `Память видеокарты`, `Политика арендатора` и сообщение о размещении.
8. Только при результатах `достаточно` и `допускает запуск` установите флажок `Я принимаю текущую оценку`.
9. Нажмите `Развернуть сервис`.

Ожидаемый результат: портал создает решение и открывает вкладку `Развертывания`.

Проверьте результат

Нажмите Обновить состояние. Успех подтверждают состояние Готов, заголовок <название> готов к работе, идентификатор операции и точка вызова. При состоянии Разворачивается сохраните идентификатор и дождитесь конечного состояния.

Если не получилось

- Укажите корректное имя файла модели в формате GGUF. Введите имя, которое оканчивается на .gguf.
- Память видеокарты: недостаточно. Выберите меньшую модель или уменьшите параметры контекста в Параметры выполнения.
- Политика арендатора: запуск не разрешен. Сохраните требуемую память и название модели и передайте их роли portal-admin.
- Свободной карты нет или Загрузка видеокарт платформе сейчас не видна. Дождитесь ресурса; второе сообщение не подтверждает готовность размещения.
- Состояние Сбой. Сохраните идентификатор операции, код блокирующего условия и время и передайте их роли portal-admin.

Завершите работу

После теста поместите временное решение в архив по странице «Архивировать временный сервис». К запросу переходите только после состояния Готов.

Развернуть модель в среде vLLM

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Учетной записи разрешено создавать решения. Источник совместим с vLLM; лицензия и правила организации разрешают модель.
- *Подготовьте.* Название и разрешенную ссылку вида https://huggingface.co/<владелец>/<модель>. Токен к ссылке не добавляйте.

Выполните

1. В разделе Сервисы на вкладке Каталог нажмите Развернуть новый сервис и выберите Большая языковая модель.
2. Заполните Название сервиса и Ссылка на модель на платформе Hugging Face.
3. В поле Размещение графического процессора (GPU) оставьте Совместно используемый графический процессор (GPU).
4. Нажмите Расширенные настройки и выберите vLLM (KServe) в поле Среда выполнения.
5. Нажмите Рассчитать потребление памяти.
6. Прочитайте Память видеокарты, Политика арендатора и сообщение о размещении.
7. Только при результатах достаточно и допускает запуск установите флажок Я принимаю текущую оценку.
8. Нажмите Развернуть сервис.

Ожидаемый результат: портал создает решение и открывает вкладку Развертывания.

Проверьте результат

Нажмите Обновить состояние. Успех подтверждают состояние Готов, заголовок <название> готов к работе, идентификатор операции и точка вызова. При состоянии Разворачивается сохраните идентификатор и дождитесь конечного состояния.

Если не получилось

- Память видеокарты: недостаточно. Выберите меньшую модель или уменьшите параметры контекста в Параметры выполнения.
- Политика арендатора: запуск не разрешен. Сохраните требуемую память и название модели и передайте их роли portal-admin.
- Свободной карты нет. Дождитесь освобождения ресурса или уменьшите потребление памяти.
- Загрузка видеокарт платформе сейчас не видна. Оценка получена, но размещение не проверено; не считайте это готовностью.
- Состояние Сбой. Сохраните идентификатор операции, код блокирующего условия и время и передайте их роли portal-admin.

Завершите работу

После теста поместите временное решение в архив по странице «Архивировать временный сервис». К запросу переходите только после состояния Готов.

Развернуть собственный сетевой сервис

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. Учетной записи разрешено создавать решения. Образ находится в разрешенном реестре, прошел проверку и зафиксирован контрольной суммой SHA-256. Сервис слушает известный порт и отвечает на путь готовности.
- *Подготовьте.* Название, ссылку вида registry.example.invalid/team/demo@sha256:<64-шестнадцатеричных-символа>, порт и путь вида /healthz. Не используйте только метку :latest. Загрузку образа и команды реестра выполняет роль developer, а не пользователь этой инструкции.

Выполните

1. В разделе Сервисы на вкладке Каталог нажмите Развернуть новый сервис и выберите Свой HTTP-сервис.
2. На шагах Основные сведения и Тип сервиса введите название и выберите http.
3. На шаге Образ и загрузка вставьте ссылку с контрольной суммой в поле Образ контейнера.
4. На шаге Контракт среды выполнения заполните Порт и Проверка готовности.
5. На шагах Ресурсы и масштабирование и Сетевой доступ оставьте исходные значения.
6. На шаге Связи и привязки (bindings) оставьте список пустым.
7. На шаге Предварительная проверка безопасности прочитайте требования и установите флажок только для проверенного образа.
8. На шаге Подача заявки на выпуск сверьте название и ссылку, затем нажмите Развернуть сервис.

Ожидаемый результат: портал открывает вкладку Развертывания созданного решения.

Проверьте результат

Нажмите Обновить состояние. Успех подтверждают состояние Готов, идентификатор операции и точка вызова. После этого откройте вкладку Вызов: поле Адрес программного интерфейса (API) должно содержать адрес от сервера.

В запрос создания входят только название, ссылка на образ, порт и путь готовности. Описание, тип, ресурсы, автоматическое масштабирование, сетевой доступ и привязки показаны в мастере, но портал не отправляет их как примененную конфигурацию.

На шаге образа показаны действия для разрешения загрузки и запроса версии компонента. Команды загрузки контейнера в эту пользовательскую инструкцию не входят.

Если не получилось

- `Укажите образ контейнера`. Введите полную ссылку с контрольной суммой.
- Сервер отклоняет ссылку только с меткой. Получите у роли `developer` ссылку с `@sha256:`.
- `Укажите порт`. Введите числовой порт, который слушает контейнер.
- `Укажите путь проверки готовности`. Введите путь, который начинается с `/`.
- Состояние `Заблокировано` или `Сбой`. Сохраните код условия, идентификатор операции и время. Причину доступа передайте роли `portal-admin`, а причину запуска контейнера или пути готовности - роли `developer`.

Завершите работу

После теста поместите временный сервис в архив по странице «Архивировать временный сервис». Для вызова откройте «Выполнить запрос к сервису через портал».

Выбрать способ запроса к сервису

Выберите один результат

- Получить ответ опубликованного сервиса по одобренной заявке: «Выполнить запрос к опубликованному сервису».
- Получить ответ собственного решения в состоянии `Готов`: «Выполнить запрос к собственному сервису».
- Получить пример внешнего вызова из фактического адреса и контракта: «Получить пример вызова программного интерфейса».

Не переносите тело запроса между сервисами с разными контрактами. Ключ нельзя сохранять в исходном коде, снимке экрана или истории команд общего устройства.

Выполнить запрос к опубликованному сервису

Перед началом

- *Проверьте.* Вход выполнен, заявка имеет состояние `одобрено`, а сервис доступен. Решите, где хранить ключ.
- *Подготовьте.* Безопасный `JSON` или разрешенный файл по контракту сервиса. Ключ храните только в разрешенном хранилище секретов.

Выполните

1. Откройте `Мои заявки` и нужную одобренную заявку.
2. Нажмите `Перейти к подключению`.
3. В блоке ключа нажмите `Показать ключ доступа к программному интерфейсу (API)` и сразу сохраните ключ в разрешенном хранилище.
4. В области `Отправить запрос` заполните поля или измените `JSON запроса`.
5. Если контракт требует файл, прикрепите подготовленный файл.
6. Нажмите `Отправить` и дождитесь конечного состояния рядом с кнопкой.

Ожидаемый результат: показано состояние `completed` или его русская подпись, а блок `Ответ` содержит результат. Для асинхронной операции портал опрашивает состояние на этой странице не более одной минуты. Для потокового ответа большой языковой модели сначала показана потоковая передача, затем итоговый ответ.

Проверьте результат

В блоке `Исходный ответ` должен быть ответ сервиса или управляющего контура. Ошибка выполнения имеет состояние ошибки, а не пустой успешный ответ. Если сервис вернул файл, раздел `Файлы в ответе` показывает кнопку `Скачать`. Ссылка действует пять минут; каждое нажатие `Скачать` получает новую ссылку.

Если не получилось

- `Ключ показывается один раз` или ответ `410`. Старый секрет повторно не выдается. Создайте новую заявку по правилам организации или передайте идентификатор ключа роли `portal-admin`.
- `JSON запроса содержит ошибку` или кнопка `Отправить` заблокирована. Исправьте названное поле, дождитесь загрузки файла или вставьте ключ.
- `Неподдерживаемый протокол`. Сохраните идентификатор решения и передайте его роли `developer`.
- Ответ `401` или `403`. Для `401` обновите сеанс и проверьте, что используете ключ, а не пароль. Для `403` сохраните идентификатор обращения и передайте его роли `portal-admin`.
- Ответ `429` или сообщение `Выполнение запроса завершилось ошибкой`. Для `429` дождитесь времени из `Retry-After`. При ошибке сохраните идентификатор запроса, время и исходный ответ без чувствительного содержимого.

Завершите работу

Удалите локальный тестовый файл, если этого требуют правила организации. Действия `Обновить ключ` и `Отозвать ключ` для опубликованного доступа в разделе `Мой доступ` отключены. Для временного ключа передайте роли `portal-admin` только его идентификатор.

Выполнить запрос к собственному сервису

Перед началом

- *Проверьте.* Решение находится в `Мои решения`, а его развертывание имеет состояние `Готов` и непустую точку вызова.
- *Подготовьте.* Безопасный `JSON` или разрешенный файл по контракту сервиса. Если точка вызова требует ключ, получите ключ этого решения через разрешенный раздел `Ключи` и сохраните его в хранилище секретов.

Выполните

1. Откройте `Сервисы`, вкладку `Мои решения` и нужное решение.
2. Откройте вкладку `Вызов`.
3. Заполните поля или измените `JSON запроса` по контракту на этой вкладке.
4. Если контракт требует файл или ключ, добавьте подготовленное значение.
5. Нажмите `Отправить` и дождитесь конечного состояния рядом с кнопкой.

Ожидаемый результат: показано состояние `completed` или его русская подпись, а блок `Ответ` содержит результат. Для асинхронной операции портал опрашивает состояние на этой странице не более одной минуты.

Проверьте результат

В блоке `Исходный ответ` должен быть ответ сервиса или управляющего контура. При ошибке показано состояние ошибки. Если сервис вернул файл, раздел `Файлы в ответе` показывает кнопку `Скачать`; ссылка действует пять минут и обновляется при следующем нажатии.

Если не получилось

- Вкладка `Вызов` недоступна. Сверьте, что развертывание имеет состояние `Готов`.
- `JSON запроса содержит ошибку`. Исправьте синтаксис и повторите отправку.
- Кнопка `Отправить` заблокирована. Прочитайте причину рядом с кнопкой и заполните названное поле.
- `Неподдерживаемый протокол`. Сохраните идентификатор решения и передайте его роли `developer`.
- Ответ `401`, `403` или `429`. Обновите сеанс для `401`, передайте идентификатор обращения роли `portal-admin` для `403`, либо дождитесь времени из `Retry-After` для `429`.

Завершите работу

Удалите локальный тестовый файл по правилам организации. Если решение было временным, откройте «Архивировать временный сервис».

Получить пример вызова программного интерфейса

Перед началом

- *Проверьте.* Откройте страницу запроса к конкретному сервису. Его точка вызова и контракт должны быть опубликованы.
- *Подготовьте.* Разрешенное место для примера. Оставьте показанные порталом заполнители до настройки разрешенного клиентского контура.

Выполните

1. На странице запроса откройте `Примеры вызова`.
2. Выберите вариант `curl`, `Python requests` или `Node fetch`; для совместимого протокола также доступен `OpenAI Python`.
3. Оставьте заполнитель `<API_KEY>` в тексте примера.
4. Скопируйте пример в разрешенный файл или средство разработки.

Проверьте результат

Скопированный вариант должен содержать фактический адрес, тело и заголовки, которые портал выбрал по описанию сервиса. Для управляющего контура пример может включать создание обращения и чтение его состояния. Сформированный порталом текст - единственный источник точной последовательности; не заменяйте его общей командой и не переносите тело между сервисами.

Если не получилось

- Раздела `Примеры вызова` нет. Пример для этого сервиса не опубликован; адрес и тело не придумывайте.
- Адрес пуст. Проверьте готовность развертывания или одобрение доступа.
- Пример содержит другой идентификатор решения. Вернитесь на страницу нужного сервиса и сформируйте пример заново.
- Тело не соответствует полям вкладки `Вызов`. Сохраните идентификатор решения и передайте расхождение роли `developer`.

- Ключ оказался в файле или записи экрана. Прекратите использование ключа и отзовите его поддерживаемым действием.

Завершите работу

Действие Подставить мой ключ показывает действующий секрет на экране. Не используйте его во время демонстрации или записи. Храните ключ отдельно от примера.

Создать сценарий большой языковой модели

Перед началом

- *Проверьте.* Вход и рабочий контекст подтверждены. В списке есть хотя бы одна доступная готовая большая языковая модель. При сообщении Нет доступных больших языковых моделей сначала разверните модель или запросите доступ.
- *Подготовьте.* Название, описание, системную инструкцию, название шага, пользовательскую инструкцию и короткое имя результата, например summary. Выберите значение Максимум токенов ответа в границе, которую показывает портал.

Выполните

1. В левом меню откройте Сценарии больших языковых моделей и нажмите + Создать сценарий.
2. Заполните Название, выберите готовую Модель и доступное Подразделение либо оставьте - не выбрано -.
3. Заполните Максимум токенов ответа и Описание.
4. Введите правило поведения модели в поле Системная инструкция.
5. В первом шаге заполните Название шага, имя результата и Пользовательская инструкция.
6. Нажмите Создать.
7. На странице сценария проверьте поля и нажмите Сохранить.

Ожидаемый результат: сохраненная страница сценария показывает выбранную модель и один шаг.

Проверьте результат

Вернитесь к списку Сценарии больших языковых моделей, найдите сценарий по названию и снова откройте его. Название модели, системная инструкция и шаг должны совпадать с введенными значениями.

Если не получилось

- Добавьте хотя бы один шаг сценария перед сохранением. Заполните первый шаг.
- Максимум токенов ответа должен быть от 1 до .... Введите значение внутри показанной границы.
- Текущая модель сценария недоступна вашей учетной записи. Выберите доступную модель.
- Поле не сохраняется. Исправьте сообщение рядом с полем и повторите Сохранить один раз.

Завершите работу

Удаление сценария через портал не реализовано. Для второго шага откройте «Добавить шаг в сценарий». Для проверки одного шага откройте «Выполнить тестовый запуск сценария».

Добавить шаг в сценарий

Перед началом

- *Проверьте.* Сценарий сохранен и содержит предыдущий шаг с уникальным именем результата.
- *Подготовьте.* Название нового шага, его инструкцию и новое уникальное имя результата. Секреты в имена и инструкции не добавляйте.

Выполните

1. Откройте сценарий и нажмите `+ Добавить шаг`.
2. Заполните `Название шага` и имя результата.
3. В поле `Пользовательская инструкция` откройте доступную палитру и вставьте результат предыдущего шага, например `{{summary}}`.
4. Допишите инструкцию нового шага и нажмите `Сохранить`.

Ожидаемый результат: новый шаг остается на странице после сохранения, а ссылка на предыдущий результат показана в его инструкции.

Проверьте результат

Обновите страницу сценария. Убедитесь, что имя результата нового шага уникально, порядок шагов сохранен, а ссылка `{{summary}}` указывает на существующий предыдущий результат.

Если не получилось

- Палитра не показывает имя результата. Сохраните предыдущий шаг и откройте страницу заново.
- Имя результата отклонено. Используйте короткое уникальное имя без секрета.
- Ссылка помечена как неизвестная. Выберите имя из палитры, а не вводите его по памяти.
- Сохранение отклонено. Исправьте названный шаг и повторите `Сохранить` один раз.

Завершите работу

Для проверки цепочки откройте «Выполнить тестовый запуск сценария». Удалять весь сценарий через скрытый программный интерфейс нельзя.

Выполнить тестовый запуск сценария

Перед началом

- *Проверьте.* Сценарий сохранен, выбранная модель готова, все шаги заполнены. Получите ключ решения модели из разрешенного раздела `Интеграция/Ключи`.
- *Подготовьте.* Синтетические значения для всех переменных. Пример: `Составь резюме вымышленного текста о планировании учебного проекта.` Не используйте персональные данные и секреты.

Выполните

1. Откройте сценарий и в блоке `Тестовый запуск` нажмите `Открыть тестовый запуск`.
2. Вставьте ключ в поле `Ключ доступа к программному интерфейсу (API) решения`.
3. Заполните каждую переменную подготовленным значением.
4. Нажмите `Запустить тест` и дождитесь конечного состояния каждого шага.

Ожидаемый результат: страница показывает непустой идентификатор запуска и блок `Итоговый ответ`.

Проверьте результат

Для каждого шага должно быть состояние `Готово`; этап `Завершение` в технических событиях имеет значение `Успешно`. Кнопка `Скачать результат (JSON)` должна быть доступна.

Тестовый запуск на этой странице не сохраняется в истории сценария и не влияет на его показатели. Если результат нужен как доказательство, скачайте `JSON` до ухода со страницы.

Если не получилось

- `Заполните обязательные поля шаблона`. Заполните перечисленные переменные.
- `Для запуска нужно выпустить или вставить ключ решения`. Используйте ключ выбранного решения, а не другого сервиса.
- `Текущая модель сценария недоступна вашей учетной записи`. Выберите доступную модель и сохраните сценарий.
- Шаг имеет состояние `Ошибка`. Сохраните идентификатор запуска и имя шага; полный запрос в обращение не включайте.
- `Предупреждение шлюза больших языковых моделей`. Сохраните идентификатор и текст предупреждения без запроса и передайте их роли `portal-admin`.

Завершите работу

Закройте страницу после сохранения разрешенного результата. Удалите скачанный файл и временную копию ключа по правилам организации.

Открыть внешний доступ к сценарию

Перед началом

- *Проверьте.* Сценарий сохранен, его модель доступна и готова, а состояние сценария допускает запуск. Внешний доступ не дает прямого доступа к модели и другим сценариям.
- *Подготовьте.* Дополнительные значения и файлы не требуются.

Выполните

1. Откройте сценарий и при необходимости нажмите `Готов к запуску`.
2. Найдите раздел `Доступ по API`.
3. Нажмите `Открыть доступ по API`.
4. Дождитесь состояния `Открыт`.

Ожидаемый результат: раздел показывает состояние `Открыт`.

Проверьте результат

Обновите страницу сценария. Раздел `Доступ по API` должен сохранить состояние `Открыт`. Открытие доступа само по себе не выдает клиентский ключ.

Если не получилось

- Кнопка открытия недоступна. Сохраните сценарий и переведите его в состояние, которое допускает запуск.
- Модель недоступна. Выберите доступную готовую модель и сохраните сценарий.
- Состояние не изменилось. Обновите страницу и проверьте фактическое состояние до повтора.
- Ответ `403`. Сохраните идентификатор сценария и обращения и передайте их роли `portal-admin`.

Завершите работу

Для внешнего клиента откройте «Выдать ключ сценария». Для прекращения всех внешних вызовов используйте страницу «Закрыть внешний доступ к сценарию».

Выдать ключ сценария

Перед началом

- *Проверьте.* Раздел `Доступ по API` имеет состояние `Открыт`. Ключ разрешает запуск только этого сценария.
- *Подготовьте.* Безопасную метку назначения ключа и разрешенное хранилище секретов.

Выполните

1. Откройте сценарий и раздел `Доступ по API`.
2. В поле метки укажите назначение ключа.
3. Нажмите `Выдать ключ`.
4. Скопируйте секрет из блока `Ключ показывается один раз — скопируйте сейчас` и сразу сохраните его в разрешенном хранилище.

Ожидаемый результат: таблица `Выданные ключи` содержит новую строку с подготовленной меткой.

Проверьте результат

Обновите страницу и найдите строку по метке. Ключ должен оставаться действующим, но секрет повторно не показывается.

Если не получилось

- Кнопка выдачи недоступна. Проверьте, что доступ имеет состояние `Открыт`.
- Ключ не показан после выдачи. Не создавайте несколько ключей подряд; обновите таблицу и проверьте фактическое состояние.
- Секрет закрыт до сохранения. Отзовите строку ключа и выдайте новый ключ.
- Ответ `403`. Сохраните идентификатор сценария и обращения и передайте их роли `portal-admin`.

Завершите работу

Удалите секрет из буфера обмена после помещения в хранилище. Для проверки ключа откройте «Вызвать сценарий через программный интерфейс». Для отзыва используйте страницу «Отозвать один ключ сценария».

Вызвать сценарий через программный интерфейс

Перед началом

- *Проверьте.* Доступ сценария имеет состояние `Открыт`, а выданный ключ действует. Получите публичный базовый адрес у службы поддержки организации.
- *Подготовьте.* `<PUBLIC_API_BASE>`, путь из поля `Эндпоинт`, ключ `<SCENARIO_KEY>`, показанный арендатор `<TENANT_ID>`, безопасное поле `input` и уникальное непустое значение заголовка `Idempotency-Key` для одного логического запуска.

Выполните

1. Подставьте значения в запрос:

```
curl -X POST "<PUBLIC_API_BASE><RUN_ENDPOINT>?wait=true" \
  -H "Authorization: Bearer <SCENARIO_KEY>" \
  -H "X-SmartMLOps-Tenant-Id: <TENANT_ID>" \
  -H "Idempotency-Key: <IDEMPOTENCY_KEY>" \
  -H "Content-Type: application/json" \
  -d '{"input":"синтетический тестовый запрос","variables":{}}'
```

1. Выполните запрос из разрешенного клиентского контура и дождитесь ответа. Синхронное ожидание ограничено 30 секундами.

Проверьте результат

Откройте сценарий и блок `Обращения за 7 дней`. Число прогонов должно увеличиться, а таблица - показать запуск с непустым идентификатором и успешным состоянием.

Без `?wait=true` сервер возвращает код ответа `202`, поле `run_id` и заголовок `Location`. Состояние читается запросом `GET` по адресу из `Location` с тем же ключом и заголовком арендатора. Для повтора того же логического запуска сохраняйте прежний `Idempotency-Key`.

Если не получилось

- `400 invalid_request` или `401 unauthorized`. Для `400` проверьте заголовок арендатора, идентификатор сценария и `JSON`. Для `401` проверьте формат `Authorization`.
- `403 forbidden` или `404 not_found`. Для `403` используйте ключ этого сценария. Ответ `404` не подтверждает существование чужого объекта.
- `409 not_published` или `422 model_unavailable`. Откройте доступ и подготовьте модельный ключ либо выберите доступную модель.
- `429 rate_limited`. Дождитесь времени из `Retry-After` и повторите запрос с прежним `Idempotency-Key`.
- `502 gateway_error`, `503 key_service_unavailable` или `503 gateway_unavailable`. Сохраните время, идентификатор обращения и запуска. Для `502` передайте сведения роли `portal-admin`, для `503` - роли `system-admin`.

Завершите работу

Удалите временную копию ключа и тело тестового запроса из истории клиента по правилам организации. Ключ оставьте в разрешенном хранилище до окончания интеграции.

Создать конвейер из готового шаблона

Перед началом

- *Проверьте.* Вход выполнен, а нужные сервисы доступны учетной записи.
- *Подготовьте.* Шаблон со значком `Готов к использованию`, доступные сервисы для обязательных привязок и название конвейера. Ключи в поле `Входы JSON` не вставляйте.

Выполните

1. Откройте `Конструктор конвейеров (pipeline)` и в разделе `Мои конвейеры` нажмите `Новый конвейер`.
2. В разделе `Готовые шаблоны` выберите шаблон со значком `Готов к использованию` и нажмите `Использовать шаблон`.
3. Для каждого узла `Требует привязки` откройте в панели `Сборка` вкладку `Привязка`.
4. В поле `Совместимый сервис` выберите доступный сервис и нажмите `Проверить и сохранить ссылку привязки`.
5. В верхней панели нажмите `Проверить`.

6. Если вкладка Ошибки содержит записи, исправьте каждую названную связь или обязательный параметр и снова нажмите Проверить.
7. После сообщения Граф прошел проверку нажмите Сохранить граф.

Ожидаемый результат: сохраненный граф остается в разделе Мои конвейеры и не содержит узлов со статусом обязательной привязки.

Проверьте результат

Вернитесь в Мои конвейеры, откройте граф и снова нажмите Проверить. Портал должен показать Граф прошел проверку, а выбранные привязки должны сохраниться.

Если не получилось

- Каталог узлов не загрузился. Палитра отключена. Обновите страницу один раз. При повторе сохраните время и передайте его роли system-admin.
- Программный интерфейс (API) не создал копию шаблона. Сохраните код и название шаблона и передайте их роли portal-admin.
- Нет доступа к сервису. Выберите доступную привязку или запросите доступ через каталог.
- Граф отклонен. Исправьте каждую запись на вкладке Ошибки.
- Сохранение не выполнено. Не запускайте граф. Сохраните идентификатор конвейера и код ответа.

Завершите работу

Удаление конвейера через портал не реализовано. Для выполнения сохраненного графа откройте «Запустить сохраненный конвейер». Для пустого графа используйте отдельную страницу «Создать конвейер с пустого графа».

Создать конвейер с пустого графа

Перед началом

- *Проверьте.* Вход выполнен, а нужные сервисы и узлы доступны учетной записи.
- *Подготовьте.* Название конвейера, схему входа, безопасные испытательные значения и список нужных узлов. Ключи в Входы JSON не добавляйте.

Выполните

1. Откройте Конструктор конвейеров (pipeline), нажмите Новый конвейер и выберите Создать с нуля.
2. Оставьте ручной триггер и задайте имя и схему входа графа.
3. Добавьте нужные узлы из палитры.
4. Соедините совместимые выходные и входные порты.
5. Для каждого узла Требует привязки выберите Совместимый сервис и нажмите Проверить и сохранить ссылку привязки.
6. Нажмите Проверить и исправьте каждую запись на вкладке Ошибки.
7. После сообщения Граф прошел проверку нажмите Сохранить граф.

Ожидаемый результат: сохраненный граф находится в Мои конвейеры и проходит повторную проверку.

Проверьте результат

Обновите страницу, откройте сохраненный граф и нажмите Проверить. Имена узлов, связи, схема входа и привязки должны сохраниться.

Если не получилось

- Порт помечен как `Несовместимо`. Не соединяйте объектный выход напрямую с текстовым входом; добавьте разрешенный преобразователь.
- Узел имеет состояние `Требует привязки`. Выберите доступный совместимый сервис.
- `Граф отклонен`. Исправьте каждую связь или обязательный параметр на вкладке `Ошибки`.
- `Сохранение не выполнено`. Сохраните идентификатор конвейера и код ответа; запуск не начинайте.
- Палитра не загрузилась. Обновите страницу один раз, затем передайте время роли `system-admin`.

Завершите работу

Удаление конвейера через портал не реализовано. Для проверки результата откройте «Запустить сохраненный конвейер».

Запустить сохраненный конвейер

Перед началом

- *Проверьте.* Граф сохранен, прошел проверку, а привязанные сервисы доступны.
- *Подготовьте.* Безопасные значения для каждого обязательного поля. Верхнее значение `Входы JSON` должно быть объектом. Секреты передаются только через разрешенную привязку.

Выполните

1. Откройте конвейер в разделе `Мои конвейеры`.
2. Нажмите `Запустить` в верхней панели.
3. В разделе `Эксплуатация` заполните обязательные поля формы запуска.
4. Нажмите `Запустить` в форме и дождитесь конечного состояния.

Ожидаемый результат: запуск имеет состояние `Успешно`, а выполненные узлы - `Успешно` или `Пропущено` для ветки, которую условие не выбрало.

Проверьте результат

Откройте `Отладка запуска`. Проверьте непустой идентификатор, конечное состояние, вход и выход каждого шага и ссылки на файлы результата, если сервис их вернул. Вкладка `Эксплуатация` должна содержать завершенный запуск. Состояния и результаты вычисляет сервер, а не браузер.

Если не получилось

- `Нет доступа к сервису`. Получите доступ или измените привязку и снова сохраните граф.
- `Введите JSON-объект`. Исправьте ввод так, чтобы верхним значением был объект `{...}`.
- `Ошибка запуска`. Сохраните идентификатор запуска и имя первого ошибочного узла.
- События не загружаются. Портал переходит к опросу состояния; дождитесь конечного состояния до повторной отправки.
- `Сохранение не выполнено`. Вернитесь в редактор и сохраните граф; не запускайте граф до сохранения.

Завершите работу

Для выполняющегося запуска используйте «Отменить выполняющийся запуск конвейера». Кнопка `Повторить` у завершенного или ошибочного запуска создает новый запуск с исходным вводом. Удалить весь конвейер через портал нельзя.

Прочитать качество и дрейф модели

Перед началом

- *Проверьте.* Вход выполнен, у сервиса подключен профиль мониторинга, а для графика накоплены наблюдения. Качество требует правильных ответов или другого разрешенного источника меток.
- *Подготовьте.* Название сервиса и один период: 7, 30 или 90 дней.

Выполните

1. Откройте `Мониторинг машинного обучения`, выберите период и нажмите `Обновить`.
2. В блоке `Состояние сервисов` откройте строку нужного сервиса.
3. На вкладке `Обзор` прочитайте `Текущий показатель качества`, `Дрейф входных данных`, `Дрейф выходных данных` и `Открытых событий`.
4. Раскройте `Технические детали профиля` и прочитайте `Последний расчет`, временное окно и число наблюдений.
5. При необходимости откройте `Профиль и сегменты`, `Конфиденциальность и метки` или `Сравнение развертываний`, не меняя состояние сервиса.

Ожидаемый результат: карточка показывает значения или точную причину их отсутствия и время последнего расчета.

Проверьте результат

- `Качество не рассчитано` не означает нулевое качество. Причинами могут быть `за окно не поступило размеченных ответов` или `наблюдений в окне слишком мало`.
- Дрейф равен `0`, если распределения совпали; верхнего предела у индекса нет.
- Входной дрейф относится к признакам, выходной - к ответам модели. Сравнение идет с предыдущим окном той же длины.
- `Нет данных` не означает `В норме`. Сообщение `Данных мало` означает, что линия появится после накопления точек.

Запись проверки должна содержать сервис, период, время расчета, число наблюдений и показанные значения.

Если не получилось

- `У этого арендатора нет активных сервисов с подключенным мониторингом`. Переключателя контекста для роли `user` нет; передайте ожидаемый сервис роли `portal-admin`.
- `Часть данных не загрузилась`. Используйте только показанные значения и сохраните время обновления.
- `Динамика сейчас недоступна` или `Качество недоступно`. Проверьте период и наличие наблюдений; отсутствующие точки не заменяйте нулями.
- `Не удалось загрузить данные мониторинга`. Нажмите `Повторить` один раз. При повторе сохраните идентификатор корреляции и передайте его роли `system-admin`.
- Состояние `Критично`. Сохраните сервис, период, время расчета и идентификатор события и передайте их роли `portal-admin`.

Завершите работу

Экран просмотра не меняет данные. Для разрешенной загрузки правильных ответов откройте «Загрузить правильные ответы модели». Для большой языковой модели используйте отдельную страницу ее мониторинга.

Загрузить правильные ответы модели

Перед началом

- *Проверьте.* Решение существует, вкладка мониторинга доступна, а строка `Загрузить метки` не имеет значения `недоступно`.
- *Подготовьте.* Файл `CSV` без персональных данных. Он содержит `request_id` или `external_correlation_id` и столбец `value`. Необязательны `correct` со значением `0` или `1` и `observed_at`.

Выполните

1. Откройте `Сервисы`, вкладку `Мои решения` и нужное решение.
2. Откройте вкладку `Мониторинг`.
3. В блоке `Эталонный набор` нажмите поле `Файл CSV с правильными ответами` и выберите подготовленный файл.
4. Нажмите `Загрузить эталонный набор`.
5. Прочитайте значения `Принято строк` и `Совпало с наблюдениями`.

Ожидаемый результат: портал показывает непустой идентификатор загрузки и ненулевое число совпадений.

Проверьте результат

Значение `Принято строк` при нуле в поле `Совпало с наблюдениями` не подтверждает расчет качества. После ненулевого сопоставления вернитесь в `Мониторинг машинного обучения`, выберите нужное окно и прочитайте время следующего расчета.

Если не получилось

- `Загрузить метки` имеет значение `недоступно`. Остановите действие и запросите владельца решения.
- `Ни одна строка не сопоставлена с наблюдением`. Сверьте идентификаторы вызовов и окно наблюдения.
- Файл отклонен. Проверьте обязательные столбцы и значения `correct`.
- `Не удалось загрузить данные мониторинга`. Нажмите `Повторить` один раз; при повторе сохраните идентификатор корреляции и передайте его роли `system-admin`.
- Качество остается недоступным после сопоставления. Сохраните идентификатор загрузки, окно и время и передайте их роли `portal-admin`.

Завершите работу

Удалите локальный `CSV` по правилам организации. Удалить загруженный пакет через портал нельзя; прямой вызов скрытого программного интерфейса не используйте.

Прочитать показатели большой языковой модели

Перед началом

- *Проверьте.* Вход выполнен, а решение большой языковой модели или сценарий существует. Для агрегатов нужен хотя бы один запрос; состояние трассировки появляется только при фактических данных контура сбора показателей.
- *Подготовьте.* Название модели или сценария, примерное время последнего запуска и идентификатор запуска для разбора отдельного случая.

Выполните

1. Откройте `Мониторинг больших языковых моделей (LLM)` и нажмите `Обновить`.

2. Прочитайте плитки Готовые большие языковые модели, Токены за окно, Средняя задержка и Сохраненные трассировки.
3. В таблице Большие языковые модели и готовность среды выполнения найдите модель.
4. В ее строке прочитайте готовность среды, число токенов, задержку и состояние трассировки.
5. Открывайте Трассировку Langfuse только при наличии разрешенной ссылки.
6. В блоке Токены, стоимость и сигналы качества прочитайте ошибки, отказы и запасные переходы.
7. В блоке Последние сценарии больших языковых моделей найдите сценарий и время его последнего запуска.

Ожидаемый результат: по модели показаны значения или точные состояния отсутствия данных, а по сценарию - время запуска либо Не запускался.

Проверьте результат

- Среда выполнения не готова не позволяет считать модель готовой по наличию строки.
- Ноль отличается от отсутствующей метрики; знак – означает, что значение не опубликовано.
- Трассировка сохранена в Langfuse подтверждает трассировку. Трассировка заблокирована политикой конфиденциальности не означает ошибку модели.
- Сводка мониторинга недоступна не подтверждает отсутствие запросов.
- Оценка стоимости не является счетом. Этот экран не показывает дрейф модели машинного обучения.

Запись проверки содержит модель, готовность среды, токены, задержку, состояние трассировки и время.

Если не получилось

- Большие языковые модели пока не найдены. Проверьте, что решение создано и связано с моделью.
- Пока нет трассировок за окно. Выполните разрешенный тестовый запрос и обновите страницу.
- Метрики больших языковых моделей сейчас недоступны. Не считайте это отсутствием трафика; сохраните время и модель.
- Трассировка устарела. Сравните время запуска и последней трассировки.
- Не удалось загрузить мониторинг больших языковых моделей или рост счетчика ошибок. Нажмите Повторить один раз. При повторе сохраните время и идентификатор запуска и передайте их роли system-admin или portal-admin по наблюдаемой причине.

Завершите работу

Экран не удаляет трассировки и агрегаты. Не очищайте их прямым вызовом. Чтобы остановить новые обращения, отзовите ключ сервиса или закройте внешний доступ к сценарию.

Завершить работу с временным объектом

Выберите один результат

- Убрать временное решение без немедленного удаления: «Архивировать временный сервис».
- Вернуть архивное решение в черновик: «Восстановить сервис из архива».
- Необратимо удалить архив: «Удалить архивный сервис навсегда».

- Остановить ожидающую заявку: «Отменить ожидающую заявку».
- Прекратить работу ключа собственного сервиса: «Отозвать ключ собственного сервиса».
- Прекратить работу одного ключа сценария: «Отозвать один ключ сценария».
- Прекратить все внешние вызовы сценария: «Закрыть внешний доступ к сценарию».
- Остановить выполняющийся конвейер: «Отменить выполняющийся запуск конвейера».
- Удалить локальные материалы и выйти: «Завершить пользовательский сеанс».

Границы текущей версии

Удалить сценарий большой языковой модели, конвейер или загруженный пакет правильных ответов через портал нельзя. Для конвейера серверный контракт удаления существует, но прямой вызов без утвержденной инструкции запрещен. Сценарий можно вернуть в состояние `Черновик` и закрыть его внешний доступ. Для окончательной очистки конвейера передайте роли `portal-admin` только его идентификатор.

Тестовый запуск сценария не сохраняется в его истории, но скачанный `JSON` остается на устройстве. Архивирование сервиса не равно немедленному окончательному удалению.

Архивировать временный сервис

Перед началом

- *Проверьте.* Сохраните разрешенные результаты, запишите идентификатор сервиса и убедитесь, что его не использует другой человек или интеграция.
- *Подготовьте.* Точное название временного сервиса. Архивирование прекращает новые вызовы, но не удаляет архив немедленно.

Выполните

1. Откройте `Сервисы`, вкладку `Мои решения` и временный сервис.
2. На вкладке `Обзор` найдите строку `Архив / удаление`.
3. Нажмите `Архивировать сервис`.
4. В окне подтверждения сверьте название и подтвердите действие.

Ожидаемый результат: страница показывает `Сервис в архиве`, дату архивации и `Удаление после`, если сервер вернул эту дату.

Проверьте результат

Обновите страницу сервиса. Вкладки изменения и новые вызовы должны оставаться недоступными, а состояние - архивным.

Если не получилось

- Кнопки нет. Учетная запись не управляет сервисом или действие недоступно; передайте идентификатор роли `portal-admin`.
- `Не удалось выполнить действие`. Обновите страницу и проверьте фактическое состояние до повтора.
- Сервис уже в архиве. Повторное действие не выполняйте.
- Название в подтверждении отличается. Отмените окно и откройте нужный сервис.

Завершите работу

Запишите идентификатор, время и архивное состояние без ключей. Для возврата откройте «Восстановить сервис из архива». Для необратимого удаления используйте отдельную страницу.

Восстановить сервис из архива

Перед началом

- *Проверьте.* Сервис имеет состояние `Сервис в архиве` и еще не удален навсегда.
- *Подготовьте.* Идентификатор и точное название сервиса. Восстановление не возвращает действующую точку вызова.

Выполните

1. Откройте архивный сервис в `Мои решения`.
2. Нажмите `Восстановить из архива`.
3. В подтверждении сверьте название и подтвердите действие.

Ожидаемый результат: архивное состояние снято, а решение открывается как черновик.

Проверьте результат

Обновите страницу. Решение должно быть доступно для разрешенного изменения, но прежняя точка вызова не должна считаться действующей. Новое развертывание выполняйте отдельной инструкцией.

Если не получилось

- Кнопки нет. Сервис уже удален, не находится в архиве или учетная запись не управляет им.
- `Не удалось выполнить действие`. Обновите страницу и прочитайте фактическое состояние до повтора.
- Восстановленное решение показывает прежний адрес. Не используйте его до нового состояния `Готов`.
- Ответ `403`. Передайте идентификатор сервиса и обращения роли `portal-admin`.

Завершите работу

Запишите новое состояние. Если восстановление было ошибочным, после проверки зависимостей снова поместите сервис в архив отдельным действием.

Удалить архивный сервис навсегда

Перед началом

- *Проверьте.* Сервис имеет состояние `Сервис в архиве`. Запишите идентификатор, сохраните разрешенные результаты и подтвердите, что сервис не нужен человеку или интеграции.
- *Подготовьте.* Точное название сервиса для повторной сверки. Отката после подтверждения нет.

Выполните

1. Откройте архивный сервис и строку `Архив / удаление`.
2. Сверьте идентификатор и название с подготовленной записью.
3. *Предупреждение.* Следующее подтверждение необратимо удаляет архивный сервис. Нажмите `Удалить архив навсегда` и подтвердите действие.

Ожидаемый результат: портал возвращает к списку решений, а удаленный сервис в нем отсутствует.

Проверьте результат

Обновите `Мои решения` и выполните поиск по точному названию. Сервис не должен находиться. Не проверяйте удаление вызовом прежней точки доступа.

Если не получилось

- Кнопки нет. Сервис не архивирован или учетная запись не управляет им.
- Название в подтверждении отличается. Отмените действие и откройте нужный сервис.
- `Не удалось выполнить действие`. Обновите список и проверьте существование объекта до повтора.
- Ответ `403`. Передайте идентификатор и обращение роли `portal-admin`.

Завершите работу

Сохраните запись с идентификатором, действием, временем и фактом отсутствия объекта. Секреты и содержимое запросов в запись не включайте.

Отменить ожидающую заявку

Перед началом

- *Проверьте.* Заявка находится в `Мои заявки` и имеет состояние `на рассмотрении`.
- *Подготовьте.* Название сервиса и идентификатор заявки. Отмена не отзывает уже одобренный доступ.

Выполните

1. Откройте `Мои заявки` и нужную заявку.
2. Нажмите `Отменить заявку`.
3. В окне `Отмена заявки` сверьте сервис и подтвердите действие.

Ожидаемый результат: состояние заявки меняется на `отменено`.

Проверьте результат

Обновите `Мои заявки` и найдите ту же запись по идентификатору. Она должна иметь состояние `отменено` и не должна возвращаться в `на рассмотрении`.

Если не получилось

- Кнопки нет. Заявка уже получила конечное состояние или принадлежит другой учетной записи.
- Состояние изменилось на `одобрено`. Для прекращения выданного доступа передайте идентификатор роли `portal-admin`.
- `Не удалось выполнить действие`. Обновите страницу и проверьте фактическое состояние до повтора.
- Ответ `403`. Передайте идентификатор заявки и обращения роли `portal-admin`.

Завершите работу

Сохраните идентификатор, время и состояние `отменено`. Новую заявку создавайте только при новой подтвержденной потребности.

Отозвать ключ собственного сервиса

Перед началом

- *Проверьте.* Получатели предупреждены, выбран ключ собственного сервиса, а не опубликованного доступа.

- *Подготовьте.* Идентификатор или метку ключа. Секрет для отзыва не нужен и не должен попадать в запись.

Выполните

1. Откройте `Сервисы`, вкладку `Мои решения` и нужный сервис.
2. Откройте вкладку `Ключи` и найдите действующий ключ по метке или идентификатору.
3. Нажмите `Отозвать` и подтвердите действие.

Ожидаемый результат: строка ключа имеет состояние `Отозван`.

Проверьте результат

Обновите таблицу `Ключи`. Затем выполните только разрешенную проверку отзыва: следующий запрос с этим ключом должен получить отказ авторизации.

Если не получилось

- Кнопки нет. Учетная запись не управляет ключом или ключ уже отозван.
- В разделе `Мой доступ` действия `Обновить ключ` и `Отозвать ключ` отключены. Для опубликованного доступа передайте роли `portal-admin` идентификатор ключа.
- `Не удалось выполнить действие`. Обновите таблицу и прочитайте фактическое состояние до повтора.
- Интеграция продолжает работать. Сверьте идентификатор ключа в журнале вызовов и передайте время отзыва роли `portal-admin`.

Завершите работу

Удалите секрет из локального хранилища интеграции по правилам организации. Сохраните только идентификатор, время и состояние отзыва.

Отозвать один ключ сценария

Перед началом

- *Проверьте.* Откройте нужный сценарий и убедитесь, что прекращается работа только одного клиента.
- *Подготовьте.* Метку или идентификатор ключа и уведомление его получателю. Секрет ключа не нужен.

Выполните

1. В разделе `Доступ по API` найдите нужную строку таблицы `Выданные ключи`.
2. Сверьте метку или идентификатор.
3. Нажмите `Отозвать` и подтвердите действие.

Ожидаемый результат: выбранная строка имеет состояние `Отозван`, а остальные действующие ключи не изменены.

Проверьте результат

Обновите страницу сценария. Сверьте состояние отозванной строки и число остальных действующих ключей. Отзыв действует немедленно и необратим.

Если не получилось

- Строка не найдена. Проверьте сценарий и метку ключа.
- Ключ уже имеет состояние `Отозван`. Повторное действие не выполняйте.
- Кнопки нет. Учетная запись не управляет сценарием или доступ уже закрыт всем.

- `Не удалось выполнить действие`. Обновите таблицу и проверьте фактическое состояние до повтора.

Завершите работу

Удалите секрет у клиента и сохраните идентификатор, время и состояние отзыва. Для прекращения всех внешних вызовов используйте отдельную страницу.

Закрыть внешний доступ к сценарию

Перед началом

- *Проверьте.* Уведомите всех получателей ключей и сохраните разрешенные результаты выполняющихся клиентов.
- *Подготовьте.* Идентификатор сценария и число действующих ключей. Закрытие влияет на всех внешних клиентов.

Выполните

1. Откройте сценарий и раздел `Доступ по API`.
2. Сверьте число действующих ключей в таблице `Выданные ключи`.
3. *Предупреждение.* Следующее действие немедленно прекращает все внешние вызовы. Нажмите `Закрыть доступ всем` и подтвердите показанное число ключей.

Ожидаемый результат: действующих ключей в таблице не остается.

Проверьте результат

Обновите страницу и снова прочитайте таблицу. Разрешенный проверочный запрос со старым ключом должен получить `409 not_published` или отказ авторизации, а не новый запуск.

Если не получилось

- Число ключей в подтверждении отличается. Отмените окно, обновите страницу и снова проверьте таблицу.
- Кнопки нет. Доступ уже закрыт или учетная запись не управляет сценарием.
- `Не удалось выполнить действие`. Обновите страницу и проверьте фактическое состояние до повтора.
- Старый ключ создает новый запуск. Сохраните идентификатор запуска и время закрытия и передайте их роли `portal-admin`.

Завершите работу

Попросите получателей удалить секреты из своих хранилищ. Сохраните только идентификатор сценария, время и число отозванных ключей.

Отменить выполняющийся запуск конвейера

Перед началом

- *Проверьте.* Запуск еще выполняется. Запишите его идентификатор и проверьте, какие шаги уже завершились.
- *Подготовьте.* Решение об отмене. Результаты, которые внешний сервис уже создал, могут сохраниться по правилам этого сервиса.

Выполните

1. Откройте конвейер и раздел `Эксплуатация`.

2. Откройте `Отладка запуска` для нужного идентификатора.
3. Нажмите `Отменить`.

Ожидаемый результат: запуск переходит в конечное состояние `Отменено`.

Проверьте результат

Обновите `Отладка запуска`. Состояние должно оставаться `Отменено`, новые шаги не должны начинаться. Отдельно проверьте уже созданные внешние результаты по правилам соответствующего сервиса.

Если не получилось

- Кнопки нет. Запуск уже завершен или учетная запись не управляет им.
- Отмена не завершается. Сохраните идентификатор и последнее состояние и передайте их роли `system-admin`.
- После обновления показано `Успешно` или `Ошибка запуска`. Операция успела завершиться; отмену повторно не отправляйте.
- Часть внешних результатов осталась. Очистите их только поддерживаемым действием соответствующего сервиса.

Завершите работу

Кнопка `Повторить` создает новый запуск с исходным вводом. Не нажимайте ее как способ продолжить отмененный запуск. Удалить весь конвейер через портал нельзя.

Завершить пользовательский сеанс

Перед началом

- *Проверьте.* Все нужные результаты сохранены в разрешенном месте, а создающие операции имеют конечное состояние.
- *Подготовьте.* Список локальных тестовых файлов, скачанных ответов и временных копий ключей. Секреты не включайте в отчет об очистке.

Выполните

1. Удалите локальные тестовые файлы и скачанные ответы по правилам организации.
2. Удалите временные копии ключей из буфера обмена, редактора и истории разрешенного клиента.
3. Откройте меню учетной записи и нажмите `Выйти`.
4. На общем устройстве закройте вкладку после появления `Гость` и `Перейти ко входу`.

Ожидаемый результат: защищенные разделы требуют нового входа, а временных локальных материалов не осталось.

Проверьте результат

Откройте `Сервисы`. Портал должен показать `Войдите, чтобы продолжить работу` или кнопку `Перейти ко входу`, а не данные предыдущего сеанса.

Если не получилось

- Команды `Выйти` нет. Обновите страницу и снова откройте меню учетной записи.
- После выхода видны защищенные данные. Закройте вкладку и передайте время и адрес роли `system-admin`.
- Локальный файл нельзя удалить. Прекратите общий доступ к устройству и обратитесь в поддержку рабочего места по правилам организации.

- Секрет попал в историю клиента. Отзовите соответствующий ключ поддерживаемым действием.

Завершите работу

Сохраните короткую запись: тип объекта, безопасный идентификатор, действие, конечное состояние и время. Ключи и содержимое запросов в нее не включайте.

Сообщения пользователю

Выберите страницу

- Если известны сообщение, код или состояние, откройте «Сообщения пользователя».
- Если безопасное восстановление не помогло или причина относится к другой роли, откройте «Кому передать пользовательскую проблему».

Сначала сохраните безопасный идентификатор и проверьте фактическое состояние объекта. Машинный код дополняет русское объяснение, но не заменяет его.

Сообщения пользователя

Как пользоваться справкой

Сохраните страницу, действие, время с часовым поясом, безопасный идентификатор объекта и доступный идентификатор обращения, операции или запуска. Сначала проверьте фактическое состояние объекта. Опасное или платное действие не повторяйте, пока неизвестно, создался ли объект.

Пароль, ключ, токен, `cookie`, полный запрос, ответ модели и персональные данные сохранять и передавать нельзя.

Записи

Вход и права

- `Войдите, чтобы продолжить работу` или `401 unauthorized`: сеанс отсутствует или истек. Выполните выход и вход. При повторе передайте время роли `portal-admin`.
- `Недостаточно прав` или `403 forbidden`: вход выполнен, но действие или объект недоступны. Запрос не меняйте; передайте объект, действие и идентификатор обращения роли `portal-admin`.
- `Активный арендатор не выбран`: в сеансе нет рабочего контекста. Повторите вход один раз. При повторе передайте время входа роли `portal-admin`.

Каталог и доступ

- `Сервис не найден` или `404 not_found`: карточка снята с публикации либо объект недоступен в текущем контексте. Вернитесь в каталог и повторите поиск. При повторе передайте безопасное название роли `portal-admin`.
- `У вас уже есть активная заявка` или `409`: откройте существующую заявку; дубль не создавайте.
- Ответ `410` при показе ключа: секрет уже показывался и удален из выдачи. Передайте роли `portal-admin` идентификатор ключа и запросите новый доступ; старый секрет не восстанавливайте.

Проверка данных

- `400 invalid_request`: обязательного поля нет или `JSON` не читается. Исправьте только поле, названное в сообщении.

- `422 model_unavailable` или другое проверяемое ограничение: объект распознан, но не может выполняться в текущем состоянии. Сохраните код и объект и передайте их роли `portal-admin`.
- `Ни одна строка не сопоставлена с наблюдением`: загрузка принята, но качество не будет рассчитано. Проверьте идентификаторы наблюдений и окно.

Ограничения и недоступность

- `429 rate_limited`: достигнуто ограничение частоты. Дождитесь времени из `Retry-After`.
- `502 gateway_error`: нижележащий сервис завершил запрос ошибкой. Сохраните идентификатор запуска и время и передайте их роли `portal-admin`.
- `503 key_service_unavailable`, `503 gateway_unavailable` или `Не удалось связаться с управляющим контуром`: компонент недоступен. Ключ и данные не меняйте; передайте время и идентификатор обращения роли `system-admin`.
- `Часть данных не загрузилась`: используйте только показанные данные; отсутствующие показатели нулями не заменяйте.

Асинхронные операции

- `Разворачивается` или `Выполняется`: операция не завершена. Обновляйте состояние существующего объекта; дубль не создавайте.
- `Заблокировано`: предварительное условие не выполнено. Сохраните код условия и устраните только названную причину.
- `Сбой` или `Ошибка запуска`: откройте события или отладку и сохраните идентификатор операции и первый ошибочный шаг.
- `Отменено`: операция закончилась без результата. Новый запуск создавайте после проверки освобождения временных ресурсов.

Границы

Машинный код не заменяет русское объяснение. Выполняйте одно безопасное действие, затем снова проверяйте исходную операцию. При повторе или выходе за полномочия пользователя выберите адресата на странице «Кому передать пользовательскую проблему».

Минимальная запись

Достаточная запись содержит время, действие, объект, фактическое состояние и хотя бы один разрешенный идентификатор. Секретов и содержимого пользовательского запроса в ней нет.

Кому передать пользовательскую проблему

Назначение

Сначала выполните безопасное действие из справочника сообщений и проверьте фактическое состояние объекта. Для передачи подготовьте страницу, действие, время с часовым поясом, безопасный идентификатор объекта, доступный идентификатор корреляции, операции или запуска, видимый код и ожидаемый результат.

Выберите одного адресата. Пароль, ключ, токен, `cookie`, персональные данные и содержимое запроса удалите из материалов.

Адресаты

Роль `portal-admin`

Передайте проблему этой роли, если причина относится к пользовательским правам, рабочему контексту, опубликованному сервису, заявке, ее одобрению, проверке общего объекта перед публикацией, ограничению арендатора или одобрению выделенного графического процессора. Эта же роль отвечает за видимость решения, сценария или конвейера, предметный ответ `403` и окончательную очистку объекта без кнопки в портале.

Роль `portal-admin` не устанавливает кластер и не восстанавливает хранилище.

Роль `system-admin`

Передайте проблему этой роли при общей недоступности портала, повторном сбое входа, недоступности управляющего контура, службы ключей или шлюза, повторном ответе `503` и операции, которая не меняет состояние после обновления. К этой роли также относятся системная сеть, хранилище, сертификат, резервное копирование, восстановление и недоступный компонент мониторинга.

Пользователь не запускает инфраструктурные команды и не получает системные секреты для диагностики.

Роль `developer`

Передайте проблему этой роли при несовместимом пакете модели, ошибке образа контейнера, пути готовности или контрольной суммы, расхождении опубликованного контракта с ответом и отсутствующих метаданных входа или выхода. Воспроизводимый дефект портала после проверки прав и доступности компонентов, новый преобразователь, профиль выполнения и исправление исходного кода также относятся к роли `developer`.

Действующий ключ не передавайте. Для воспроизведения нужны вымышленный ввод и идентификатор обращения.

Единый канал поддержки

Если причина не классифицирована, отправьте одно обращение в утвержденный канал поддержки. Укажите признак и выполненные безопасные проверки. Не назначайте техническую роль по догадке и не рассылайте дубль нескольким ролям.

Отказы при передаче

- Обращение ушло нескольким ролям: прекратите дублирование и оставьте одного адресата.
- Идентификатора нет: вернитесь на страницу объекта и скопируйте разрешенный идентификатор.
- В материал попал секрет: не отправляйте его; отзовите раскрытый ключ поддерживаемым действием.
- Получатель просит пароль или полный токен: секрет не передавайте; используйте утвержденный безопасный канал.
- Получатель вернул обращение как чужую область: сохраните ссылку на исходное обращение и передайте ее одной названной роли.

Минимальная запись

Обращение содержит ожидаемое и фактическое состояние, точное время, безопасный идентификатор объекта, доступный идентификатор корреляции, операции или запуска и одну целевую роль. До ответа не повторяйте создание, развертывание или удаление, если фактическое состояние неизвестно.

Лист регистрации изменений

Редакция	Дата	Содержание изменения
0.1	22.08.2026	Первый выпуск книги в Typst. Добавлены проверенные маршруты портала, программного интерфейса, ограничения и границы передачи проблем.
0.2	22.08.2026	Каждая практическая страница получила один результат. Отдельные входные пути, запуск, внешний вызов и очистка вынесены в самостоятельные инструкции.
0.3	23.08.2026	Обновлены версия продукта, содержание и язык пользовательских инструкций.