Nevel

ДОКУМЕНТАЦИЯ ДЛЯ КОМПАНИЙ

Документация Nevel

Platform · Guardrail · API

48 документов

https://docs.nevel.sh

1. Как устроен Nevel и где решать свою задачу
2. Первый вход и первый рабочий результат
3. Чат: запросы, история и совместная работа
4. Модели и режимы работы
5. Ассистенты для повторяющихся задач
6. Файлы: загрузка, папки, доступ и корзина
7. Пространства, участники и общий контекст
8. Документы и DOCX
9. Таблицы и XLSX
10. Презентации и PPTX
11. Изображения и видео
12. Поиск в интернете и детальное исследование
13. Интеграции и MCP
14. Автоматизации по расписанию
15. Создание и управление API-ключами
16. Аналитика использования и расходов
17. Настройки аккаунта и администрирование команды
18. Популярные модели
19. Безопасность Nevel
20. Архитектура и границы доверия
21. Потоки данных и маршруты обработки
22. Guardrail: маскирование и маршрутизация
23. Шифрование и управление ключами
24. Идентификация, SSO и права

25. Контроль доступа и аудит
26. Хранение, сроки и удаление данных
27. Защита файлов и баз знаний
28. Сетевая защита и свой AI firewall
29. Модель угроз и негативные проверки
30. Меры защиты и требования РФ
31. Мониторинг и интеграция с SIEM
32. Реагирование на инциденты
33. Эксплуатация, резервирование и восстановление
34. Корпоративная среда и российские средства ИБ
35. Защита от prompt injection
36. Облако, on-prem и маршруты моделей
37. Облачное размещение
38. On-prem в инфраструктуре компании
39. Первый запрос к API
40. API-ключи и лимиты
41. Ошибки API и повторные запросы
42. Как устроена стоимость Nevel
43. Облачные планы Pro и Advanced
44. Стоимость установки в контуре
45. Использование, кредиты и бюджеты
46. Оплата, продление и документы
47. 152-ФЗ: согласование обработки данных
48. Документы для закупки и согласования

НАЧАЛО РАБОТЫ

Как устроен Nevel и где решать свою задачу

Карта платформы: чаты, модели, ассистенты, файлы, пространства, документы, интеграции, автоматизации и API.

Облако / В контуре

https://docs.nevel.sh/start/overview/

Nevel Platform — рабочий интерфейс для задач с AI: задавать вопросы, анализировать материалы, готовить документы, обращаться к подключённым сервисам и запускать повторяющиеся задачи. Начать можно с обычного чата. Когда задача становится постоянной, её контекст переносится в пространство или ассистента, а регулярный запуск — в автоматизацию.

Выберите инструкцию по своей задаче: работа в чате, администрирование команды, подключение API или размещение платформы.

Найдите нужную возможность

Что нужно получить	Куда перейти
Ответ, анализ или черновик по задаче	Чат: запросы, история и ветки
Подходящий стиль работы модели	Модели и режимы
Сравнение популярных моделей и базовых цен	Топ-10 моделей
Повторяемый сценарий с постоянной инструкцией	Ассистенты
Анализ исходного PDF, таблицы или другого файла	Файлы
Контекст проекта и доступ коллег	Пространства
DOCX, текстовый документ или экспорт	Документы
XLSX с данными и формулами	Таблицы
PPTX с планом и отдельными слайдами	Презентации
Иллюстрация или короткий видеоролик	Изображения и видео
Актуальные сведения и отчёт по источникам	Поиск и исследование
Работа с корпоративным сервисом из чата	Интеграции и MCP
Сводка или проверка по расписанию	Автоматизации
Дирекции, права и массовые назначения	Администрирование команды

Если вы впервые открыли платформу, начните с входа и первого результата. Там описана последовательность от приглашения до проверки созданного файла.

Ориентируйтесь в навигации

В узкой левой панели сверху расположен выбор рабочего пространства команды. Ниже находятся иконки внутренних пространств, объединяющих контекст отдельных проектов. В нижней части — аватар с меню настроек, темы, обновлений и выхода.

Основная боковая панель содержит **«Новый чат»**, **«Поиск»**, **«Автоматизации»**, **«Ассистенты»**, **«Файлы»**, **«Управление»** и историю разговоров. Видимость отдельных пунктов зависит от текущего пространства, прав и включённых возможностей. На небольшом экране панель открывается кнопкой меню.

«Поиск» в боковой панели находит сообщения в истории. **«Поиск в Интернете»** в поле ввода обращается к внешним источникам. Это разные действия: для возвращения к старому обсуждению не нужен интернет-режим.

«Управление» относится к выбранному внутреннему пространству: инструкции, доступ и интеграции. **«Настройки»** в меню аватара открывают профиль и доступные административные разделы всей команды.

Различайте модель, режим и ассистента

Модель или профиль модели определяет, как обрабатывается запрос. В селекторе доступны варианты **«Авто»**, **«Быстрый»**, **«Думающий»**, **«Пишущий»** и разрешённые в установке модели. «Авто» подходит для первого запроса; другие варианты выбирают под характер задачи и доступные условия.

Режим задачи задаёт нужный тип работы: документы, таблицы, презентации, изображения, видео, поиск или исследование. Выбор режима перед отправкой помогает получить нужный инструментальный результат. Название «Документы» не означает, что ответ обязательно должен стать длиннее: вы по-прежнему задаёте объём и формат.

Ассистент хранит инструкцию для повторяющейся роли и связанные материалы. Например, он может каждый раз разбирать предложение поставщика по одному шаблону. **Пространство** объединяет контекст проекта и участников. Их можно использовать вместе: проект «Закупки» и ассистент «Проверка коммерческого предложения» решают разные задачи.

Пройдите путь от исходников до готового результата

Сначала выберите правильное пространство и исходные файлы. Затем сформулируйте задачу с проверяемым выходом: «таблица различий», «документ из пяти разделов», «презентация на восемь слайдов». Если требований много, попросите модель сначала показать план.

После первого результата вносите конкретные изменения в том же чате: называйте раздел, лист, слайд или версию. Не отправляйте без пояснения новый файл с тем же названием — модель должна понимать, какой материал является актуальным.

В завершение скачайте результат и откройте в целевом приложении. Предпросмотр удобен для быстрой проверки, но отличается от полноценного редактора: DOCX просматривается в интерфейсе, XLSX показывает табличные данные, а слайды — изображения страниц. Для проверки формул, печати и финального оформления откройте сам файл. Подробные ограничения приведены в статьях соответствующих форматов.

Что делает администратор

Администратор помогает сотрудникам войти, назначает роли и группы, организует общий контекст и следит за использованием. Начните с настроек команды, затем проверьте доступ к одному рабочему пространству и практический сценарий каждого подразделения.

Для программных интеграций выдаются API-ключи. Отдельные ключи для приложений и сред упрощают ограничение расхода и замену. Аналитика показывает расходы API, обращения, токены и активность участников. Эти показатели помогают понять использование, а качество и пользу результата оценивают на реальных задачах.

Pro и Advanced дают одинаковые возможности и инструменты; основная разница — лимиты использования. Условия и порядок выбора плана описаны в тарифах. Если функция скрыта, проверьте политику установки и права пользователя: наличие статьи не заменяет конфигурацию конкретной организации.

Platform и Guardrail/API

Platform предназначена для работы человека в интерфейсе. Guardrail/API подключает модели и защитную обработку к корпоративным приложениям и агентам. Разработчик использует адрес API, ключ, модель и формат запроса, описанные в быстром старте API.

Если модель из чата читает данные вашей системы, это интеграция Platform. Если ваша система вызывает модель через Nevel, это программный доступ. Для архитектуры полезно сначала определить направление взаимодействия, затем источник данных и учётные данные. Состав защитной обработки и границы данных описаны в безопасности.

Размещение платформы и маршруты обработки

Облако — Platform в инфраструктуре Nevel. **Установка в контуре** — Platform в инфраструктуре заказчика. Это выбор места размещения приложения, его данных и компонентов; место выполнения моделей задаётся отдельно.

Доступные маршруты моделей включают собственные модели заказчика, российских провайдеров — например Яндекс и Сбер — и глобальные модели. В облаке Nevel также доступна **self-hosted DeepSeek на серверах в РФ**. Размещение модели у Nevel и её размещение внутри инфраструктуры заказчика — разные варианты, даже если оба используют российские серверы.

Platform в контуре может обращаться к разрешённым внешним провайдерам по политике организации. **Изолированный профиль** — частный случай, где модели, инструменты и источники ограничены внутренним периметром. Для него внешний веб-поиск и публичные генераторы не используются. Подробнее — модели и маршруты.

На страницах возможностей оба варианта разобраны отдельно. Для выбора и подготовки установки используйте сравнение размещений, а для повседневной работы — соответствующую операционную статью.

НАЧАЛО РАБОТЫ

Первый вход и первый рабочий результат

Принятие приглашения, вход, выбор пространства, анализ файла, создание документа и проверка доступа коллег.

Облако / В контуре

https://docs.nevel.sh/start/launch/

Первый рабочий сценарий должен подтвердить весь путь: вы вошли в нужную команду, выбрали пространство, передали материал, получили полезный результат и смогли его открыть. Ниже — последовательность для нового сотрудника и короткая проверка для администратора.

Получите адрес и приглашение

Используйте адрес Platform, который выдала организация. Для облака это её точка входа в Nevel, для установки в контуре — адрес развёрнутой у заказчика платформы. Адрес этой документации не является формой входа в рабочую платформу.

Если вы получили приглашение, откройте ссылку из письма. На странице указано, под каким email нужно войти. Используйте именно этот адрес: авторизация другим аккаунтом не даёт права принять чужое приглашение. После входа нажмите **«Присоединиться к проекту»**. При успешном принятии появится сообщение **«Приглашение успешно принято»** и переход **«Открыть проект»**.

Слово «проект» в приглашении может относиться к рабочему пространству команды. После принятия отдельно убедитесь, что видите нужное внутреннее пространство и материалы. Если ссылка недействительна, администратор может отправить приглашение повторно из списка участников.

Войдите по электронной почте

На странице **«Войти на платформу»** введите email и нажмите **«Продолжить с электронной почтой»**. Если включена проверка при входе, завершите её. После сообщения **«Проверьте почту»** откройте новое письмо.

Для входа используются временная ссылка и/или одноразовый код, в зависимости от конфигурации. Можно перейти по ссылке либо нажать **«Ввести код вручную»**, ввести шестизначный код из актуального письма и выбрать **«Продолжить с кодом»**. Не используйте код из документации или старого письма.

Если письмо не пришло, сначала проверьте написание адреса и папку нежелательной почты. Затем воспользуйтесь **«Отправить письмо ещё раз»**. При истёкшем коде нужен новый запрос на вход. Не открывайте много параллельных форм: это усложняет понимание, к какому запросу относится очередное письмо.

Если доступен **«Войти с помощью Telegram»**, завершите предложенную проверку и подтверждение в приложении. При блокировке всплывающего окна разрешите его для этой страницы и повторите попытку. Используйте способ входа, настроенный для вашей организации; внешний Telegram не является обязательным компонентом установки в контуре и не используется в изолированном профиле без внешних подключений.

Проверьте команду и профиль

После входа нажмите логотип команды сверху слева и посмотрите выбранное рабочее пространство. Если у вас несколько команд, выберите корпоративную. Затем нажмите иконку нужного внутреннего пространства.

Откройте меню аватара → **«Настройки»** → **«Профиль»** и проверьте имя и привязанный email. Это помогает избежать ситуации, когда приглашение отправлено на рабочий адрес, а вы вошли в личный аккаунт. Для возврата нажмите **«Вернуться в чат»**.

Если ожидаемого пространства нет, передайте администратору адрес аккаунта и название нужного пространства. Администратор проверит членство в команде и доступ к конкретному контексту. Создание одноимённого пустого пространства не вернёт доступ к исходным материалам.

Начните с короткой текстовой задачи

Нажмите **«Новый чат»**, оставьте выбор модели **«Авто»** и отправьте задачу без конфиденциальных данных:

> Составь структуру внутреннего отчёта о запуске проекта: цель, сделанная работа, риски, решения на следующую неделю. Для каждого раздела напиши, какие исходные сведения нужны. Не придумывай выполненные работы и показатели.

На компьютере Enter отправляет запрос, Shift+Enter добавляет перенос строки. Дождитесь завершения ответа. Затем напишите уточнение: «Сократи структуру до одной страницы и добавь таблицу решений». Так вы проверите не только первый ответ, но и продолжение диалога.

Название чата можно изменить через меню. Назовите его по работе, например «Шаблон еженедельного отчёта». Это пригодится для поиска и дальнейших уточнений.

Передайте первый файл

Выберите небольшой разрешённый для обработки документ: описание проекта, учебную таблицу или утверждённый шаблон. Прикрепите его к запросу и дождитесь окончания загрузки. В одном сообщении поддерживается до десяти вложений; отдельный файл ограничен 50 МБ. Поддерживаемые типы перечислены в работе с файлами.

Напишите задачу с конкретной проверкой:

> По приложенному описанию составь таблицу: цель проекта, ожидаемый результат, срок, ответственный. Для каждого значения укажи раздел исходного файла. Если сведений нет, напиши «не указано».

Сверьте несколько строк с исходником. Если не распознана таблица или нужный раздел отсутствует, не переходите сразу к финальному отчёту. Укажите пропущенную часть, загрузите более читаемую версию или передайте нужный фрагмент текстом. Наличие файла в сообщении не гарантирует, что плохой скан превратился в достоверный источник.

Создайте и доработайте результат

Выберите **«Документы»** и попросите: «На основе проверенной таблицы создай DOCX с отчётом. Сохрани неизвестные значения как “не указано”. Не добавляй новые факты». Дождитесь карточки файла и откройте предпросмотр.

Просмотрите заголовки, факты и таблицы. Для изменения продолжите тот же чат: «В последнем документе добавь раздел “Открытые вопросы” из трёх пунктов; остальные разделы сохрани». Скачайте итоговый DOCX и откройте его в приложении, в котором будете использовать результат.

Для другого результата выберите «Таблицы» или «Презентации». Эти инструменты доступны и в Pro, и в Advanced; планы отличаются лимитами использования. Не ожидайте, что обычная текстовая таблица в сообщении уже является XLSX: явно назовите нужный формат.

Найдите и передайте результат

Откройте **«Поиск»** в боковой панели или нажмите Ctrl/Cmd+K. Найдите характерную фразу из вашего запроса и перейдите к нужному сообщению. Убедитесь, что можете вернуться к работе без повторной генерации.

Для передачи коллегам используйте доступные действия файла или чата. У чата есть **«Поделиться»**, копирование ссылки и скачивание Markdown. Перед отправкой ссылки проверьте её режим доступа и какие сообщения входят в публикуемый материал. Для рабочего документа зачастую удобнее передать проверенный файл, чем весь разговор с исходниками.

Проверка администратора

После подключения нового сотрудника выполните вместе короткий сценарий:

1. Сотрудник входит под приглашённым адресом и видит нужную команду.
2. Открывает внутреннее пространство и один разрешённый файл.
3. Получает ответ с указанием источника и создаёт документ.
4. Скачивает и открывает результат.
5. Находит чат повторно и понимает, где получить помощь.

Если сотрудник будет работать с сервисами или расписаниями, отдельно проведите проверку интеграции и тестовый запуск автоматизации. Успешный чат не подтверждает авторизацию в корпоративной системе.

Размещение платформы и маршруты обработки

В облаке проверьте разрешённый маршрут модели и правила передачи материалов. Кроме разрешённых российских и глобальных провайдеров, доступна self-hosted DeepSeek Nevel на серверах в РФ. Если используется модель заказчика, заранее должно быть настроено подключение к её точке выполнения. Первую задачу удобно выполнить на небольшом разрешённом файле.

При установке Platform в контуре сначала обеспечьте доступ к адресу платформы, настроенной авторизации, хранилищу и выбранной модели. Модель может работать у заказчика, у российского провайдера, например Яндекса или Сбера, либо у глобального провайдера — по политике организации. Текстовый ответ и генерация документа проверяют разные компоненты: успешный чат не доказывает работоспособность сервиса файлов.

Для изолированного профиля выберите внутреннюю модель и внутренние инструменты, а тест выполняйте без интернет-поиска. Если требуется подтверждение полностью локальной обработки, одной успешной загрузки внутреннего сайта недостаточно: администратор проверяет весь маршрут модели и инструментов.

При ошибке сохраните время, текст сообщения и ID запроса, если он показан. Подробные решения типичных проблем находятся в статьях чата, файлов и настроек.

РАБОТА С ПЛАТФОРМОЙ

Чат: запросы, история и совместная работа

Первый запрос, вложения и голос, продолжение диалога, поиск сообщений, ветки и экспорт чата.

Облако / В контуре

https://docs.nevel.sh/platform/chat/

Чат — место, где вы ставите задачу, передаёте исходные материалы и уточняете результат. В одном диалоге можно анализировать файл, подготовить документ, попросить изменить его и скачать результат. Для новой независимой задачи создавайте новый чат; для уточнения уже полученного результата продолжайте существующий.

Начните с правильного пространства

Выберите команду и нужное пространство в боковой панели. Пространство определяет общий контекст: системные инструкции, материалы и доступные подключения. Затем нажмите **«Новый чат»**. Если вы переключили пространство с уже прикреплёнными, но не отправленными файлами, вложения очищаются: прикрепите их заново в правильном месте.

В поле **«Спроси о чем угодно...»** напишите, что нужно получить. Укажите исходный материал, требуемое действие, формат результата и существенные ограничения. Вместо «посмотри договор» полезнее написать:

> Сравни два приложенных проекта договора. Составь таблицу: пункт, изменение, практическое последствие. Не делай юридический вывод; если исходный пункт неоднозначен, приведи его текст. Заверши списком вопросов, которые надо обсудить с контрагентом.

Такой запрос задаёт проверяемый результат. Если задача большая, сначала попросите план или перечень найденных источников. После этого уточните объём и дайте команду выполнить работу. Не требуется заранее знать названия внутренних инструментов модели.

Отправка, перенос строки и остановка

На компьютере **Enter** отправляет сообщение, **Shift+Enter** добавляет перенос строки. На сенсорном устройстве используйте кнопку отправки: нажатие Enter в экранной клавиатуре может оставаться переносом строки. Если открыт список `@`-упоминаний, Enter выбирает его пункт, а не отправляет запрос.

После отправки поле ввода очищается, а в диалоге появляется выполняемый запрос. Во время ответа используйте кнопку остановки, если направление задачи выбрано неверно. Уже полученная часть ответа может остаться в истории. Не считайте остановленный документ готовым: откройте артефакт и посмотрите, какая часть работы действительно завершилась.

Файлы и голосовой ввод

Нажмите **«+» → «Загрузить файлы»** или вставьте файл из буфера обмена. Дождитесь завершения загрузки: отправка блокируется, пока вложения загружаются. В одном сообщении допускается до 10 файлов, включая файлы, выбранные через упоминания. Текущий лимит одного загружаемого файла — 50 МБ; ограничения конкретной модели могут быть строже.

Для уже сохранённого материала введите `@` и выберите файл в списке. Это помогает однозначно указать, с каким документом работать. Просто написанное название файла и выбранное упоминание — разные вещи: во втором случае интерфейс связывает запрос с конкретным доступным материалом. Подробнее — в руководстве по файлам.

Кнопка микрофона записывает речь и преобразует её в текст поля ввода. Разрешите браузеру доступ к микрофону, завершите запись, затем проверьте распознанные имена, числа и знаки препинания до отправки. Очень короткая или практически бесшумная запись может не дать текста. Голосовой ввод не означает непрерывный голосовой разговор с моделью.

Продолжайте и исправляйте результат

В следующем сообщении ссылайтесь на уже полученный результат: «Сократи второй раздел», «Добавь формулы на листе Итоги», «Оставь структуру, но перепиши вывод». Для точного редактирования откройте нужный документ или упомяните его через `@`. Если в чате несколько похожих файлов, укажите название и нужный фрагмент.

У сообщения доступны действия редактирования и повторного получения ответа. Редактирование исходного запроса полезно, когда ошибка возникла именно в постановке задачи. Перегенерация помогает получить другой вариант ответа, но не исправляет отсутствующий файл, запрещённую модель или отключённую интеграцию. При наличии нескольких версий используйте переключатели **«Предыдущая версия»** и **«Следующая версия»**; счётчик показывает текущую версию.

Создайте отдельную ветку

Нажмите **«Создать ветку»** у сообщения, до которого хотите сохранить контекст. Nevel создаст отдельный чат с историей до выбранной точки; в названии появится префикс **«Ветка -»**. Продолжайте в нём альтернативный сценарий: например, один вариант предложения для технической аудитории, другой — для закупок.

Ветка удобна, когда надо сравнить два направления, не смешивая дальнейшие обсуждения. Это отдельный диалог, а не переключатель роли пользователя. Доступ к использованным файлам и сервисам по-прежнему определяется их собственными настройками.

Найдите и организуйте историю

Нажмите **«Поиск»** или **⌘K / Ctrl+K** и введите запомнившийся фрагмент сообщения. Результаты показывают название чата, дату и совпадение. Нажатие открывает диалог у найденного сообщения. Этот поиск по сообщениям отличается от поиска файлов и от **«Поиска в Интернете»**.

В меню чата доступны **«Переименовать»**, **«Закрепить чат»**, **«Поделиться»** и **«Удалить чат»**, с учётом контекста диалога. Закрепите регулярно используемый диалог, а для названия выберите предмет задачи: «Сравнение предложений — сентябрь» легче найти, чем «Новый чат». Удаление не заменяет экспорт нужного результата. У ранее созданного проектного чата может отображаться команда исключения из проекта; она не является настройкой доступа к пространству.

Поделиться и скачать

Откройте **«Поделиться»**. **«Скопировать ссылку»** создаёт ссылку для публикации чата; **«Скачать в Markdown»** сохраняет текстовый экспорт. До передачи ссылки проверьте содержимое переписки и её назначение. Публичная ссылка на материал и выдача человеку доступа к пространству — разные способы распространения.

Если работа остановилась

При превышении контекста сократите последнее сообщение, уберите часть вложений или начните новый чат с кратким резюме. Если доступна кнопка **«Сжать и продолжить»**, она предназначена для сокращения истории. Ошибка о слишком больших инструкциях или наборе инструментов требует участия администратора: повтор той же отправки её не устраняет.

При **«Подключение прервалось»** используйте **«Повторить»**. При ошибке интеграции сначала проверьте подключение. Сохраните показанные **ID запроса** и **ID ответа** для поддержки.

Platform может работать в облаке Nevel или в контуре заказчика. Модель для чата выбирается отдельно: собственная модель заказчика, российский провайдер, глобальная модель либо доступная в облаке Nevel self-hosted DeepSeek на серверах в РФ. Используйте только разрешённый организации маршрут. Для изолированного профиля доступны внутренние модели и источники; обычная установка в контуре может обращаться и к разрешённым внешним сервисам. Текст запроса не меняет сетевые правила.

РАБОТА С ПЛАТФОРМОЙ

Модели и режимы работы

Как Авто выбирает модель, чем отличается от режима инструмента, как работает резервный выбор и где выполняется обработка.

Облако / В контуре

https://docs.nevel.sh/platform/models/

В Nevel отдельно выбираются модель обработки и режим задачи. Модель задаёт способ получения ответа; режим помогает начать конкретный сценарий — поиск, исследование, работу с документами, таблицами, презентациями или медиа. Для обычного запроса начните с **«Авто»**, а инструментальный режим включайте под требуемый результат.

Выбор модели

Рядом с полем ввода находится кнопка с названием выбранной модели. Нажмите её: откроется меню **«Модель»**. Выберите доступный пункт; галочка показывает текущий выбор. Названия конкретных провайдерских моделей берутся из каталога установки, поэтому не ориентируйтесь на старый список из презентации.

Конкретные семейства моделей и порядок обновления — в каталоге моделей.

Вариант	Подсказка интерфейса	Когда попробовать
Авто	Баланс качества и скорости, подходит для большинства задач	Первый запуск сценария, смешанные задачи
Быстрый	Отвечает быстро, подходит для простых вопросов	Короткий ответ, небольшой пересказ, простая правка
Думающий	Думает дольше для лучших ответов	Сложное сравнение, анализ условий, последовательное решение
Пишущий	Для создания текстов	Черновик письма, статья, редактура формулировок

Эти варианты помогают выбрать характер работы. Они не обещают одинаковые затраты, точность или время ответа для любых исходных данных. Длинный файл остаётся длинной задачей и в быстром режиме. Сложность результата зависит также от необходимых инструментов: создание файла включает дополнительные операции помимо написания текста.

Как работает «Авто»

«Авто» поручает Nevel выбрать модель для текущего обращения. Пользователю не нужно заранее сопоставлять каждую задачу с названием провайдера. В автоматическом сценарии работает отдельный маршрутизатор — Router. Он анализирует задачу, выбирает один вариант из доступного набора и передаёт выполнение выбранной модели.

Выбор происходит перед основным ответом. Он не означает одновременную отправку одного вопроса всем моделям или голосование между их ответами. Router получает текущее обращение и ограниченный настройкой фрагмент последних сообщений пользователя и ассистента. Поэтому существенные требования полезно повторить в самой задаче: тему, ожидаемый результат, необходимость краткости или глубокого разбора. Нельзя рассчитывать, что условие из начала длинного диалога обязательно попадёт именно в контекст выбора.

Порядок обработки:

1. Platform определяет доступные для выбора модели и действующую конфигурацию автоматического сценария.
2. Router получает недавний контекст и описание вариантов. Он оценивает область задачи, её сложность и специальные требования.
3. Router возвращает выбранную модель и краткое обоснование решения.

4. Platform проверяет, что выбранный идентификатор входит в предложенный набор. Придуманное название не становится новым подключением.
5. Основная модель выполняет задачу. Если ей нужны разрешённые инструменты, начинается их отдельная работа.

СХЕМА 1

Автоматический выбор модели



В отдельных конфигурациях «Авто» может быть связано с заранее заданной моделью без отдельного вызова Router. Поэтому само название в переключателе ещё не доказывает, что в каждом запросе выполнялся динамический выбор. Для технического разбора нужны сведения о фактически выбранной модели и конфигурации сценария.

По каким признакам выбирается модель

В первую очередь учитывается **область задачи**, затем сложность и дополнительные требования. В настройке маршрутизатора сопоставлены направления: программирование, математика и логика, общие задачи разной сложности. Один и тот же выбранный провайдер не закрепляется за человеком навсегда: следующее обращение может требовать другого способа обработки.

Признак запроса	Как влияет на выбор
Простое объяснение или короткая правка	Подходит вариант для несложных задач
Сравнение нескольких условий, последовательный разбор	Учитывается необходимость более глубокого анализа
Программирование	Различаются обычные изменения и сложные архитектурные задачи
Математика или логическая задача	Используется назначение для соответствующего направления
Просьба ответить быстро или кратко	Учитывается предпочтение скорости
Просьба подробно обосновать вывод	Учитывается требование глубины ответа

Фраза «ответь быстро» не отменяет сложность исходного материала и не гарантирует определённое время. Обоснование выбора — объяснение Router, а не доказательство, что модель объективно лучшая для любой такой задачи. Если результат важен, проверьте выводы по источникам и повторите задачу с явно выбранной моделью для сравнения.

Router, модель ответа и инструменты

У сложной задачи могут быть разные обработчики на разных этапах. Например, поиск получает материалы, основная модель связывает сведения в ответ, а генератор документа создаёт файл. Для внутренних инструментов могут быть настроены собственные модели. Выбор «Авто» относится к автоматическому сценарию ответа; он не означает, что все эти операции выполняет одна и та же модель.

Различайте три решения:

- **Подбор модели задачи:** какой вариант подходит запросу по области и сложности.
- **Маршрут по политике:** кому разрешено получить эти данные и где допустима обработка.
- **Модель инструмента:** какой обработчик выполняет поиск, создание изображения или другую специализированную операцию.

Предпочтение Router не расширяет права на данные или разрешённые направления передачи. Для корпоративного запроса проверяется также путь самого Router, поскольку он получает содержание, достаточное для выбора. Разрешение основной модели не делает автоматически разрешёнными маршрутизатор, поиск или генератор файла. Правила получателей описаны в архитектуре и границах доверия.

У маршрутизатора есть собственный вызов модели, время работы и учёт потребления. Поэтому автоматический выбор не следует считать бесплатной вычислительной операцией или гарантией уменьшения общей стоимости. Для оценки эффективности сравнивают завершённые задачи одинакового качества с учётом выбора модели, инструментов и повторов. Сам факт использования разных моделей на этапах не является измерением экономии.

Резервный выбор и диагностика

Если Router не может выбрать модель — например, недоступен, не успел ответить или вернул вариант вне доступного набора, — используется настроенный резервный выбор, **fallback**. Это возврат к предусмотренной модели, а не поиск произвольного провайдера. Резервный маршрут подчиняется ограничениям организации на получателей и данные.

Для разбора запроса администратору нужны выбранный пользователем режим, фактическая модель ответа, время обращения, признак резервного выбора и причина. В обработке автоматического выбора также учитываются модель самого Router и её потребление. Краткое обоснование решения и сведения о fallback относятся к технической диагностике; их наличие не означает отдельную кнопку просмотра в каждом пользовательском интерфейсе.

Если выбор кажется неудачным, действуйте последовательно:

1. Сохраните формулировку запроса и отметьте, что именно не подошло: глубина ответа, формат, работа с файлом или доступность инструмента.
2. Проверьте, был ли выбран «Авто» и какой результат фактически получен. Ошибка генерации файла отличается от ошибки выбора модели.

3. В новой ветке повторите запрос с конкретной разрешённой моделью и теми же материалами.
4. Для обращения в поддержку приложите время и идентификатор запроса, если он показан. Содержание закрытых документов передавайте только по принятому организацией каналу.

Неудачный ответ сам по себе не доказывает, что сработал fallback. И наоборот, резервный выбор не означает потерю диалога: это изменение обработчика текущего обращения. При изменении доступного каталога или правил администратор повторяет проверку типичных задач и резервного пути.

Изменение модели в диалоге

Можно выбрать другую модель для следующего сообщения. Уже написанный ответ от этого не изменится. Чтобы получить сравнимый вариант, повторите задачу с теми же источниками или создайте ветку чата до спорного ответа. Иначе вы сравните не только модели, но и разный накопленный контекст.

Удобный контрольный запрос для сравнения:

> По приложенному регламенту перечисли пять условий возврата. Для каждого укажи раздел документа. Если условие не найдено, напиши «в источнике не найдено». Не дополняй ответ общими знаниями.

Оцените полноту, точность ссылок и наличие неподтверждённых деталей. Не выбирайте модель только по уверенности тона: уверенная формулировка не является доказательством правильности ответа.

Режим задачи

Нажмите **«+»** у поля ввода. Кроме **«Загрузить файлы»**, доступны режимы:

- **«Поиск в Интернете»** — запрос с поиском внешних сведений.
- **«Детальное исследование»** — многоэтапное исследование с подготовкой отчёта.
- **«Документы»** — создание и изменение текстовых документов.
- **«Таблицы»** — работа с табличным файлом.
- **«Презентации»** — подготовка и изменение слайдов.
- **«Изображения»** и **«Видео»** — генерация соответствующих материалов.

На стартовом экране часть режимов также показана отдельными кнопками; оставшиеся находятся под **«Ещё»**. Выбранный режим отображается рядом с вводом. Уберите его нажатием на элемент режима, если задача изменилась. Одновременно выбирается один такой режим.

После отправки выбор режима сбрасывается. Для следующего сообщения проверьте поле ввода: если снова нужен явный инструментальный режим, выберите его повторно. При этом историю и уже созданные файлы диалог сохраняет. Сброс кнопки не означает исчезновение предыдущего документа.

Модель не заменяет постановку задачи

Выбор **«Пишущий»** сам по себе не задаёт формат DOCX, аудиторию или длину текста. Напишите это в запросе. Аналогично **«Думающий»** не включает автоматически доступ к закрытой базе, а **«Таблицы»** не определяет, какой лист нужно изменить.

Сформулируйте задачу четырьмя частями: источник, операция, результат и ограничения. Например: «По @План.xlsx на листе Продажи посчитай квартальные итоги. Добавь лист Сводка с формулами, исходный лист не меняй». Для презентации укажите аудиторию, число слайдов как желаемый объём и основную мысль; для поиска — период и тип источников.

Доступность моделей и инструментов

Значок замка означает недоступный выбор. Если коллега видит другой набор, сравните выбранную команду, роль и конфигурацию доступа. Модель, инструмент и источник данных проверяются отдельно: наличие модели для ответа не даёт автоматически доступ к интеграции или чужому файлу.

Текст запроса не выдаёт новые права. Разрешения сотрудника и его дирекции настраиваются через администрирование команды и действующую матрицу доступа. Если нужен конкретный инструмент, сначала проверьте его доступность, а затем сравнивайте модели, которые могут с ним работать.

Размещение платформы и маршруты моделей

Сначала определите, где работает Platform: в облаке Nevel или в инфраструктуре заказчика. Затем отдельно выберите, куда отправляется запрос модели. Внутренний адрес платформы не означает, что все модели локальные; облачный интерфейс не означает обязательное использование глобального провайдера.

Маршрут модели	Где выполняется обработка	Что учитывать
Собственные модели заказчика	В инфраструктуре заказчика	Адрес модели, доступность из Platform, поддерживаемые операции и вычислительную мощность
Российские провайдеры, например Яндекс и Сбер	В инфраструктуре выбранного российского провайдера	Разрешённый сервис, договорные условия и данные, передаваемые за пределы платформы
Глобальные модели	У выбранного глобального провайдера	Политику компании, допустимые данные и фактический маршрут подключения
Self-hosted DeepSeek Nevel в облаке	На серверах Nevel в РФ	Это модель, размещённая Nevel; она не становится собственной моделью заказчика

И облачная Platform, и Platform в контуре могут использовать разрешённые для них маршруты. Для подключения модели заказчика из облака нужен настроенный канал к её точке доступа; для внешнего провайдера из контура — разрешённое исходящее соединение. Выбор в интерфейсе не открывает сетевой доступ автоматически.

Название модели и её маршрут — разные сведения. Например, DeepSeek может означать модель в инфраструктуре заказчика, self-hosted модель Nevel или предложение провайдера. При выборе корпоративного сценария уточните именно точку выполнения и условия обработки, а не только семейство модели.

Изолированный профиль

Полностью закрытая установка — частный случай on-prem. В этом профиле модели, инструменты и источники работают внутри разрешённого периметра; внешние обращения отключены. Для веб-поиска, генерации медиа и интеграций нужны внутренние средства либо эти операции будут недоступны в изолированном сценарии.

При обычной установке в контуре российские и глобальные провайдеры могут использоваться, если их разрешает политика организации. Размещение платформы у заказчика не означает, что запрос целиком обработан у заказчика: для этого все компоненты маршрута должны работать в его инфраструктуре.

Проверьте модель на своём сценарии

Не все модели одинаково поддерживают изображения, инструменты и структурированный вывод. Если текстовый запрос работает, а запрос с файлом — нет, уменьшите пример до одного вложения и проверьте возможности выбранной модели. Смена провайдера может изменить формат ответа, доступный контекст и поддерживаемые операции.

После подключения или замены модели выполните короткий текстовый запрос, запрос по файлу и нужное инструментальное действие. Проверьте не только текст ответа, но и фактически созданный документ или выполненную операцию. Размер поддерживаемого контекста и возможности обработки определяются выбранной моделью и конфигурацией сценария.

Типичные затруднения

«Модель не выбрана» — дождитесь загрузки каталога и выберите доступный вариант. Если список не появился, обновите страницу и проверьте доступ к установке. **«Модель не найдена»** — выбранное имя могло измениться; используйте текущий каталог. При длинном ожидании посмотрите, идёт ли создание файла или работа инструмента: это отдельный этап. Если видна ошибка, сначала исправьте её причину и только затем сравнивайте модели.

Ассистенты для повторяющихся задач

Создание ассистента с инструкциями и файлами, проверка в чате, изменение и предоставление доступа.

Облако / В контуре

https://docs.nevel.sh/platform/assistants/

Ассистент сохраняет повторяемую постановку задачи: название, описание, инструкции и файлы. Вместо копирования длинного промпта в каждый новый чат вы выбираете нужного ассистента и передаёте конкретный материал. Это подходит для разбора обращений, редакторской проверки, подготовки предложений и ответов по ограниченному набору документов.

Откройте каталог

Нажмите **«Ассистенты»** в боковой панели. В списке используются группы **«Ваши ассистенты»** и **«Общие ассистенты»**. Поле **«Поиск ассистентов»** помогает найти нужное название. У карточки есть меню с действиями **«Начать чат»**, **«Редактировать»**, **«Поделиться»** и **«Удалить ассистента»**, когда соответствующее действие разрешено.

Если нужен разовый анализ одного документа, начните обычный чат. Создавайте ассистента, когда инструкция должна повторяться на разных входных данных или использоваться несколькими людьми. Отдельное пространство полезнее для общего набора проектов и участников; ассистент полезнее для одинакового способа работы. Эти механизмы можно использовать вместе.

Создайте ассистента

В редакторе на вкладке **«Форма»** заполните **«Название»**, **«Описание»** и **«Инструкция»**. В блоке **«Файлы»** добавьте материалы, которыми ассистент должен пользоваться. Затем нажмите **«Создать»**. При редактировании существующего ассистента кнопка называется **«Обновить»**; без изменения формы она может быть неактивной.

Название должно обозначать работу: «Проверка коммерческого предложения» полезнее, чем «Помощник отдела». В описании объясните, что человек получит и что ему надо предоставить. Инструкция предназначена для поведения модели: последовательность анализа, формат ответа, правила обращения с неполными сведениями.

Пример инструкции

Для ассистента, который проверяет обращения в поддержку, можно начать с такого текста:

```
Разбирай одно обращение клиента за раз.
Сначала выдели продукт, симптом, ожидаемое поведение и уже выполненные действия.
Если важного поля нет, пометь его «не указано».
Найди в приложенном справочнике подходящий порядок диагностики.
Верни: краткое резюме, классификацию проблемы, следующие шаги и вопросы клиенту.
Указывай название раздела справочника для каждой рекомендации.
Не придумывай номер заказа, версию продукта, срок устранения или выполненные проверки.
Не отправляй ответ клиенту; подготовь текст для проверки сотрудником.
```

Такое описание задаёт воспроизводимую процедуру. Команда «будь экспертом» не сообщает, как разбирать неполные сведения и что считать готовым ответом. Если ассистент должен менять данные в подключённой системе, отдельно определите разрешённое действие и порядок подтверждения в рабочем процессе.

Подготовьте файлы знаний

Прикрепляйте документы, которые объясняют задачу: актуальный справочник, структуру отчёта, правила оформления. Не складывайте в ассистента все доступные файлы «на всякий случай»: лишние источники затрудняют понимание, на какое правило опирается ответ.

Давайте файлам различимые названия и убирайте заменённые версии. Если два документа содержат разные условия, явно укажите в инструкции применимость каждого. Удаление файла из ассистента изменяет его контекст; проверьте после этого новый чат. Не предполагается, что уже написанные ответы автоматически пересчитаются по обновлённому справочнику.

Проверьте во вкладке «Чат»

Перейдите на вкладку **«Чат»** редактора или выберите **«Начать чат»** из меню карточки. Подайте три разных примера: обычный, неполный и находящийся вне компетенции. Для приведённого ассистента это может быть описанная ошибка входа, обращение без версии продукта и запрос, которого нет в справочнике.

Проверьте, что формат сохраняется, недостающие данные не выдуманы, а ссылки действительно ведут к нужным разделам источника. Затем измените одну инструкцию и повторите тест в новом диалоге. Так видно влияние инструкции без накопленной истории предыдущего эксперимента.

В обычном чате ассистента также можно выбрать через `@`. В списке упоминаний есть отдельная группа **«Ассистенты»**. После выбора появляется элемент ассистента; его удаление снимает выбор. Для одного запроса используйте одного выбранного ассистента, не рассчитывая, что несколько упоминаний автоматически запускают команду агентов.

Настройте доступ

Откройте **«Поделиться»** → **«Настройки доступа»**. Доступны варианты **«Только я»**, **«Общий доступ»**, **«Все, у кого есть ссылка»**, а также **«Скопировать ссылку»**. Выберите вариант и нажмите **«Сохранить»**; сообщение **«Настройки доступа обновлены»** подтверждает сохранение настройки.

До выбора доступа по ссылке просмотрите инструкции и вложения: они могут содержать внутренние правила или сведения команды. После изменения откройте ассистента под учётной записью предполагаемого получателя. Копирование URL само по себе не заменяет настройку доступности и проверку того, какие материалы сможет использовать получатель.

Изменение и удаление

Используйте **«Редактировать»**, когда изменились процесс или источники. Сохраняйте описание назначения ассистента понятным для коллег. После **«Удалить ассистента»** он перестаёт быть доступным элементом каталога; заранее перенесите нужные инструкции и файлы, если их следует сохранить. Отсутствие карточки не является доказательством удаления всех исторических ответов из созданных ранее чатов.

Ограничения и неполадки

Ошибка **«Не удалось загрузить данные ассистента»** требует проверки доступа и повторной загрузки страницы. При ошибке удаления файла повторите действие и убедитесь, что состав вложений действительно изменился. Если ответы стали хуже после расширения инструкции, уберите дублирование и противоречащие требования; слишком длинная инструкция занимает контекст модели.

Место хранения ассистента определяется размещением Platform: облако Nevel или контур заказчика. Модель и инструменты выбираются по политике организации отдельно от этого решения: можно использовать модели заказчика, российских провайдеров и глобальные модели; в облаке доступна также self-hosted DeepSeek Nevel на серверах в РФ. Файлы, передаваемые выбранной модели, следуют этому маршруту обработки.

В изолированном профиле используются только внутренние модели и подключения. Обычная установка в контуре может обращаться к разрешённым внешним системам. Инструкция ассистента не выдаёт сетевые разрешения или доступ к чужому материалу. Для настройки используйте интеграции и пространства.

РАБОТА С ПЛАТФОРМОЙ

Файлы: загрузка, папки, доступ и корзина

Работа с каталогом файлов, упоминаниями в чате, предпросмотром, перемещением, публикацией и восстановлением.

Облако / В контуре

https://docs.nevel.sh/platform/files/

Раздел **«Файлы»** помогает находить загруженные материалы и результаты работы ИИ, раскладывать их по папкам, открывать и использовать в новых задачах. Это отдельный каталог: наличие текста о файле в сообщении ещё не означает, что нужный файл открыт или прикреплён к следующему запросу.

Откройте нужный набор файлов

Сначала выберите пространство, затем нажмите **«Файлы»**. Используйте **«Все файлы»** и хлебные крошки для перехода между корнем и папками. В табличном виде доступны колонки **«Название»**, **«Владелец»**, **«Добавлено»**, **«Действия»**. Переключатели **«Табличный вид»** и **«Карточный вид»** меняют представление, сохраняя набор материалов.

Поле **«Поиск файлов»** предназначено для поиска в каталоге. Фильтры **«Все»**, **«Только мои»**, **«Общие»** помогают сузить набор. В фильтре **«Тип»** есть текстовые документы, презентации, таблицы, картинки и видео, а также **«Другое»**. Если файл пропал из списка, сначала очистите поиск и фильтры, проверьте выбранное пространство и папку.

Загрузка и поддерживаемые форматы

Нажмите **«Загрузить файлы»** или перетащите файлы в предусмотренную область. В списке загрузки видны состояния **«Ожидает загрузки»**, **«Файл загружается»**, процент прогресса, **«Загружен»** или **«Ошибка»**. При сбое используйте **«Повторить загрузку»**; **«Убрать из списка»** удаляет строку неудачной попытки из очереди.

В пользовательской загрузке предусмотрены PDF, DOC/DOCX, XLSX, PPTX, Pages, Numbers, Keynote, TXT, RTF, JSON, CSV, Markdown и изображения GIF, PNG, JPEG, WebP. Возможность приложить формат не означает, что каждый его элемент поддерживается при редактировании. Например, сложную презентацию с макетами нужно проверить после обработки в целевом редакторе.

Текущий общий предел загрузки — **50 МБ на файл**. В одно сообщение чата можно включить до **10 файлов**, считая загруженные и выбранные через упоминания. Ограничение модели на объём контекста и изображения проверяется отдельно. Несколько файлов, успешно загруженных в каталог, могут не поместиться в один запрос к модели.

Папки и названия

Нажмите **«Новая папка»**, введите название и выберите **«Создать папку»**. Для изменения используйте **«Переименовать папку»**. Если появляется **«Папка с таким названием уже есть»**, выберите другое имя или откройте существующую папку.

Файл переименовывается через **«Переименовать файл»**. Название помогает находить материал через каталог и `@`-упоминание. Практичный вариант — «Отчёт-продажи-2026-09.xlsx», если в пространстве есть несколько периодов. Не используйте одинаковое «Финал» для разных документов: это затрудняет выбор источника даже при технически допустимом расположении.

Для переноса откройте **«Переместить»**, выберите папку назначения и нажмите **«Переместить сюда»**. Поддерживается также перетаскивание. При конфликте имени появится объяснение: переименуйте элемент или выберите другую папку. Перемещение в папку организует каталог, но не является самостоятельной командой выдать доступ новой аудитории.

Используйте материал в чате

Введите `@` и часть названия файла, выберите результат в группе **«Файлы»**. После выбора добавьте действие: «По @Отчёт.xlsx найди три позиции с наибольшим отклонением» или «Сравни @Регламент-старый.docx и @Регламент-новый.docx». Так запрос связан с конкретными материалами.

Открытый документ также может передавать модели контекст предпросмотра. Для неоднозначных задач всё равно указывайте название, лист, слайд или выделенный фрагмент. Если модель отвечает по другому файлу, не продолжайте цепочку исправлений на неверном источнике: выберите правильный материал и повторите исходную постановку.

Предпросмотр и скачивание

Нажмите **«Открыть файл»**. Для поддерживаемого формата откроется панель предпросмотра. Команда **«Скачать файл»** сохраняет исходный или текущий файл на устройство. В документной панели доступны формат, название, обновление и скачивание.

Предпросмотр не всегда показывает весь материал. Для XLSX выводится до 1000 строк и 50 столбцов на лист; при сокращении появляется пояснение. Для полного анализа скачайте файл или попросите ИИ обработать конкретный диапазон. Ошибка визуального предпросмотра сама по себе не означает, что файл пуст или повреждён: попробуйте скачивание и откройте его в соответствующем редакторе.

Общий доступ

Команда **«Поделиться»** у файла предлагает сделать его доступным участникам проекта. Подтверждение объясняет результат: участники смогут найти и использовать файл в каталоге, `@`-упоминаниях и файловых инструментах. Это доступ внутри совместной работы; отдельно существует публикация ссылки для поддерживаемых документов.

Проверьте состав пространства до публикации общего материала. После предоставления доступа попросите получателя найти файл через собственную учётную запись. Если он видит карточку, но не может открыть содержимое, сохраните название и текст ошибки для администратора.

Корзина и восстановление

После **«Удалить»** файл перемещается в **«Корзину»**. Откройте её и нажмите **«Вернуть»**, чтобы восстановить материал. Если имя уже занято, переименуйте конфликтующий файл и повторите восстановление. В текущем интерфейсе корзина показывает срок 30 дней с последующим автоматическим удалением; это правило корзины файлов, а не общий срок хранения журналов или истории.

«Очистить корзину» окончательно удаляет её содержимое после подтверждения. Перед удалением папки она должна быть пустой.

В облаке каталог размещён в инфраструктуре Nevel; при установке в контуре — в инфраструктуре заказчика. Место анализа содержимого определяется выбранным маршрутом модели, а доступ к внешним ссылкам и источникам — политикой подключений. Контур может использовать разрешённые внешние модели и сервисы. Только изолированный профиль ограничивается внутренними моделями и источниками. Разрешение «использовать только этот файл» задаёт источник ответа, но само по себе не гарантирует локальное выполнение модели.

Для DOCX, XLSX и PPTX продолжите с документами, таблицами и презентациями.

РАБОТА С ПЛАТФОРМОЙ

Пространства, участники и общий контекст

Два уровня организации работы, создание пространства, системные инструкции, файлы и доступ.

Облако / В контуре

https://docs.nevel.sh/platform/workspaces/

В Nevel есть два уровня организации работы: **рабочее пространство команды** и **пространства внутри него**. Первый задаёт состав команды, административные роли, тариф и API-ключи. Второй объединяет контекст конкретной работы: чаты, системные инструкции, материалы и подключения. В интерфейсе слово «пространство» встречается для обоих уровней, поэтому ориентируйтесь также на место настройки.

Различайте два уровня

Что вы настраиваете	Где это находится	На что влияет
Команда, её название и общие настройки	Логотип команды сверху; «Настройки» → «Пространство»	Рабочее пространство организации
Приглашения и административные роли	«Настройки» → «Участники»	Членство в команде
Проектный контекст, инструкции, материалы и доступ	Иконка пространства слева → «Управление»	Выбранное пространство внутри команды
Собственный профиль и способы входа	Меню аватара → «Настройки» → «Профиль»	Ваш аккаунт

Например, у компании может быть одно рабочее пространство «Компания» и внутри — «Закупки», «Маркетинг», «Проект Каталог». Для сотрудников доступны пространства, в которые их добавили. Папка с файлами помогает упорядочить материалы, но сама по себе не заменяет состав участников пространства.

Чтобы отделить контекст одного проекта от другого, начните с внутреннего пространства. Создание ещё одного рабочего пространства команды — более крупное изменение: у него отдельные участники, настройки и использование. Кнопка **«Добавить пространство»** в меню выбора команды относится к этому верхнему уровню и может зависеть от тарифа.

Выберите пространство перед работой

Нажмите логотип команды в верхней части узкой левой панели и выберите нужное рабочее пространство. Ниже расположены иконки внутренних пространств; наведение показывает название. Нажмите нужную иконку перед созданием чата или загрузкой файла.

При переключении пространство становится текущим контекстом. Если вы находились внутри конкретного чата, интерфейс может вернуть вас к началу чата нового пространства. Проверьте выбранное название перед отправкой материалов. Вложения, которые ещё не были отправлены, при смене пространства очищаются.

Иконки пространств можно менять местами перетаскиванием. Это порядок навигации, а не изменение прав доступа или приоритета инструкций. Если список не загрузился, воспользуйтесь кнопкой повторной загрузки; не создавайте копии пространств из-за временно пустой панели.

Создайте новое пространство

1. Нажмите **«Создать пространство»** — кнопку с плюсом в левой панели.
2. Введите **«Название пространства»**, описывающее работу: «Подготовка тендера» или «Каталог продукции».
3. Выберите **«Иконка»**, чтобы быстро отличать пространство в панели.
4. При необходимости добавьте участников через поле **«Имя или email»**.

5. Нажмите **«Создать»** и дождитесь появления пространства в списке.
6. Откройте **«Управление»**, чтобы добавить инструкции и материалы.

Предусмотрен предел числа создаваемых пространств, а интерфейс сообщает, когда он достигнут. Не путайте его с лимитом рабочих пространств верхнего уровня. Перед созданием очередного пространства посмотрите, нет ли существующего контекста с тем же назначением и составом участников.

При добавлении человека по email возможны разные результаты: существующий участник команды получает доступ к пространству; существующий пользователь Nevel может быть добавлен в команду и пространство; для нового адреса создаётся приглашение. Проверяйте статус человека в доступе после создания.

Задайте системные инструкции

Откройте **«Управление» → «Кастомизация»**. Здесь доступны название, иконка и **«Системные инструкции»**. Инструкции используются чатами внутри пространства как общий контекст. Запишите устойчивые правила проекта, а разовую задачу оставляйте в сообщении чата.

> Мы готовим материалы для внутренней команды закупок. Отвечай по-русски. В сравнительных таблицах сохраняй исходные названия поставщиков и единицы измерения. Для чисел указывай файл и раздел-источник. Если сведений нет, ставь «нет данных». Не подменяй фактическое содержание документов предположениями.

Хорошая инструкция объясняет назначение и формат работы. Она не должна содержать пароли, длинную переписку или весь архив проекта. Большой постоянный контекст расходует место, необходимое запросу и ответу. Подробную справочную информацию удобнее хранить в файлах и явно выбирать для задачи.

Нажмите **«Сохранить»** после изменений. Если кнопка недоступна, проверьте, изменили ли вы значения и есть ли у вас право управления. Создайте новый чат, задайте короткий вопрос и оцените применение правил. Уже готовые сообщения не переписываются при изменении инструкции.

Добавьте материалы пространства

В блоке **«Файлы пространства»** нажмите **«Добавить файлы»**. Добавляйте документы, нужные большинству задач этого контекста: требования, справочник, утверждённый шаблон, описание продукта. Назовите файлы так, чтобы сотрудник мог отличить актуальную редакцию от черновика.

Файлы пространства и отдельные вложения чата дополняют друг друга. Общий справочник полезно оставить в пространстве; два варианта документа для разового сравнения можно прикрепить только к запросу. Для важного факта просите модель указывать конкретный источник. Присутствие файла в списке не доказывает, что каждый его фрагмент был прочитан при каждом ответе.

Список **«Файлы»** в основной навигации предоставляет поиск, папки, фильтры и управление материалами. Инструкция по загрузке, ограничениям форматов, общему доступу и корзине находится в разделе файлов.

Управляйте доступом

На вкладке **«Доступ»** видны пользователи с доступом и ожидающие приглашения. Внутри пространства есть роли **владелец** и **участник**. Владелец управляет названием, системными инструкциями, составом и удалением пространства. Обычный участник не получает эти полномочия автоматически.

Чтобы добавить коллегу, найдите его по имени или email, выберите запись и нажмите **«Пригласить»**. Для удаления используйте **«Удалить из пространства»** рядом с участником. Удаление из одного пространства и удаление из команды — разные операции. Общие административные роли OWNER / ADMIN / MEMBER настраиваются отдельно в участниках команды.

После выдачи доступа попросите коллегу открыть пространство и найти один контрольный файл. Если он не видит материалы, проверьте адрес приглашения, выбранную команду, принятие приглашения и состав конкретного пространства. Не решайте проблему доступности включением общей ссылки на весь архив.

Подключения и повторяемые сценарии

На вкладке **«Интеграции»** подключаются доступные сервисы и собственный MCP. Начните с задачи чтения, чтобы проверить аккаунт и права источника. Подробные шаги приведены в интеграциях.

Для специализированного сценария, например проверки коммерческого предложения по шаблону, создайте ассистента. Пространство задаёт общий контекст проекта, а ассистент — конкретную роль и последовательность работы. Для регулярного запуска по времени используйте автоматизации.

Удаление пространства

Удалить пространство может его владелец; пространство по умолчанию удалить нельзя. Диалог предупреждает об удалении пространства и связанных данных. Это отдельная операция, её нельзя считать обычным переносом файла в корзину.

До удаления сохраните нужные результаты и убедитесь, что коллеги больше не используют этот контекст. Если требуется лишь сменить название или состав участников, выполните соответствующее изменение. Не удаляйте пространство как способ устранить ошибку загрузки или подключения.

Размещение платформы и маршруты обработки

В облаке эти действия выполняются в организации, размещённой в Nevel. Пространства помогают организовать контекст и доступ внутри неё; они не являются отдельным вариантом физического размещения.

При установке Platform в контуре пространства, их настройки и материалы размещаются в инфраструктуре заказчика. Маршруты моделей задаются отдельно: собственные модели, российские провайдеры или глобальные модели могут использоваться по политике организации. Создание нового пространства не включает внешний сервис автоматически и не меняет разрешения на передачу материалов.

Изолированный профиль — отдельная конфигурация, в которой обработка и инструменты ограничены внутренним периметром. В облаке Nevel также можно использовать self-hosted DeepSeek на серверах в РФ; это размещение модели у Nevel, а не внутри инфраструктуры заказчика. Подробнее — модели и маршруты и варианты размещения.

РАБОТА С ПЛАТФОРМОЙ

Документы и DOCX

Создание и правка текстов, работа с существующим Word-файлом, выделением, Markdown-редактором и экспортом.

Облако / В контуре

https://docs.nevel.sh/platform/documents/

Nevel помогает создавать документы с нуля и изменять существующие DOCX. Рабочий процесс проходит в чате: вы описываете результат, открываете созданный файл, просматриваете его и задаёте следующую правку. Для текстовых материалов также доступен Markdown-документ с режимом ручного редактирования и экспортом.

Выберите исходную задачу

Для нового документа нажмите **«+» → «Документы»** и укажите назначение, аудиторию, структуру и формат. Для изменения Word-файла сначала загрузите DOCX или выберите его через `@`. Если нужен именно редактируемый Word-файл, напишите «Верни результат в DOCX»: абзацы в сообщении чата сами по себе не являются документом.

Пример создания:

> Подготовь DOCX-инструкцию для сотрудника по приёмке оборудования. Аудитория — новый офис-менеджер. Разделы: подготовка, проверка комплекта, фиксация повреждений, завершение. Добавь таблицу для модели, серийного номера, состояния и комментария. Используй только сведения из приложенного регламента; неизвестные условия вынеси в вопросы.

Пример изменения:

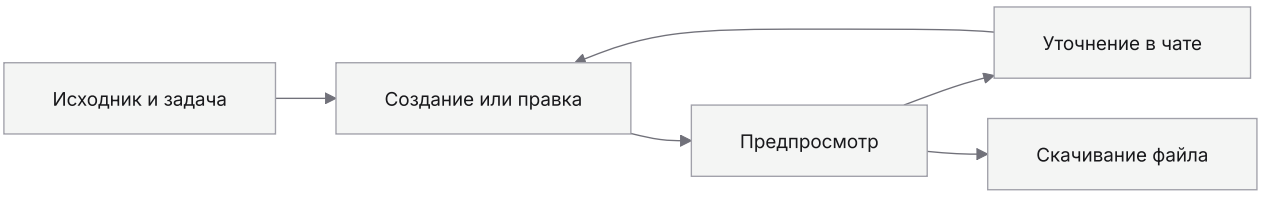
> В @Инструкция.docx перепиши раздел «Передача оборудования» понятным языком. Сохрани заголовки и таблицу учёта. Не меняй сроки и ответственных в остальных разделах. После правки перечисли, какие фрагменты изменены, и покажи обновлённый DOCX.

Что происходит во время работы

Модель читает доступный документ, выполняет необходимые операции и возвращает карточку результата. До завершения может показываться **«Создаю документ...»**. В карточке есть название и приглашение **«Нажмите, чтобы открыть»**. Нажмите карточку, чтобы перейти от сообщения об операции к самому файлу.

СХЕМА 1

Создание и правка документа



Смотрите не только на сообщение модели «готово», но и на открывшийся результат. Если документ пуст, содержит лишь часть разделов или всё ещё показывает загрузку, работа требует продолжения. Для длинного текста полезно сначала получить структуру и один образец раздела, затем развить остальные по согласованному формату.

Предпросмотр DOCX

В панели отображаются название, отметка **DOCX**, действия скачивания, обновления и закрытия. **«Оглавление»** строится по заголовкам документа. Если выводится **«В документе нет заголовков»**, это может означать, что визуально крупный текст в исходнике не размечен стилями заголовков. Попросите применить настоящие заголовки нужных уровней.

Предпросмотр Word-файла предназначен для чтения, навигации и выделения. Ввод текста прямо в эту панель заблокирован; правки выполняются через чат. Поэтому отсутствие курсора для набора не является ошибкой редактирования. Для ручной правки в Word скачайте DOCX, измените его в редакторе и загрузите обновлённый файл как новый источник.

Правка выделенного фрагмента

Откройте документ и выделите нужный текст. Контекст открытого документа и выделения помогает передать задачу модели. Напишите конкретное изменение: «Сократи выделенный абзац до двух предложений, сохрани числа» или «Преврати этот перечень в таблицу с колонками действие и ответственный».

Если выделение неоднозначно или не попало в запрос, укажите дополнительно название раздела и начало абзаца. Для повторяющихся заголовков добавьте соседний фрагмент. Формулировка «измени это» без выбранного документа может относиться к последнему сообщению, а не к нужному месту файла.

После операции нажмите обновление в панели, если виден старый вариант. Просмотрите изменённый участок и ближайшие заголовки, таблицы, списки. При изменении структуры попросите модель перечислить затронутые элементы: это облегчает проверку, но не заменяет чтение файла.

Markdown-документ и ручное редактирование

Для заметки, отчёта или черновика можно попросить создать текстовый документ. В Markdown-панели доступны **«Просмотр»** и **«Редактирование»**, если у вас есть право изменять файл. Внесите правку, нажмите **«Сохранить изменения»** и вернитесь к просмотру. **«Отменить»** отменяет текущую ручную правку.

Для изменения через модель используйте **«Редактировать с AI»**. Это удобно для преобразования текста: сменить тон, перестроить разделы, выделить выводы. Для небольшой опечатки ручной редактор позволяет обойтись без нового запроса.

При **«Не удалось сохранить изменения»** сначала сохраните свой текст вне поля, чтобы не потерять его, затем обновите документ и повторите правку. Сообщение о слишком большом файле относится к возможности сохранения в редакторе; уменьшите документ или разбейте его на самостоятельные части.

Скачивание и публикация

У DOCX нажмите **«Скачать»**. У Markdown используется меню **«Скачать как»** с MD, HTML, DOCX и, когда доступен экспорт для этого материала, PDF. После экспорта откройте скачанный файл: особенно важно проверить таблицы, длинные заголовки, нумерацию и переносы страниц.

Для поддерживаемого текстового документа есть **«Опубликовать ссылку»**. Диалог **«Публичная ссылка готова»** прямо сообщает, что опубликованный снимок доступен всем, у кого есть ссылка. Это снимок материала; не предполагается, что последующие правки автоматически изменят уже переданный экспорт или опубликованную версию.

Как получить стабильный результат

Разделяйте содержательные и оформительские изменения. Сначала проверьте факты и структуру, затем шрифт, отступы и таблицы. Укажите элементы, которые должны остаться неизменными. Для сложного шаблона загрузите образец, а не описывайте фирменную вёрстку только словами.

Если модель переписала слишком много, восстановите исходный файл как источник и задайте узкую правку. Если результат ссылается на несуществующие положения, попросите показать использованные фрагменты исходника и убрать неподтверждённые выводы. Сохранение сложной вёрстки, комментариев и отслеживаемых исправлений проверяйте на

фактическом DOCX; наличие предпросмотра не является гарантией полной совместимости всех функций Word.

Документы доступны и в Pro, и в Advanced: планы различаются лимитами использования, а не наличием этого инструмента. При остановке по квоте проверьте конкретный лимит и административную политику.

Platform и средства работы с файлами могут быть размещены в облаке Nevel или в контуре заказчика. Модель для чтения и подготовки текста выбирается отдельно из разрешённых маршрутов: модели заказчика, российские провайдеры или глобальные модели; в облаке Nevel доступна также self-hosted DeepSeek на серверах в РФ. При отправке содержимого внешней модели оно обрабатывается за пределами инфраструктуры самой платформы.

Для изолированного профиля используются внутренние модели, документные инструменты и источники. Фраза «Используй только приложенные документы» ограничивает источники ответа; для полностью локальной обработки дополнительно нужен соответствующий маршрут модели.

РАБОТА С ПЛАТФОРМОЙ

Таблицы и XLSX

Анализ и изменение Excel-файлов: листы, диапазоны, формулы, проверка итогов и ограничения предпросмотра.

Облако / В контуре

https://docs.nevel.sh/platform/spreadsheets/

Режим **«Таблицы»** позволяет читать и изменять табличные файлы, строить расчёты и возвращать результат в XLSX. Вы задаёте операции в чате, а результат открываете в панели предпросмотра. Для точной работы указывайте файл, лист, диапазон и правило расчёта.

Подготовьте исходную таблицу

Загрузите XLSX через **«+» → «Загрузить файлы»** или выберите сохранённый файл через `@`. Затем включите **«Таблицы»**. Если данных пока нет, опишите структуру новой книги: названия листов, колонки, типы данных и желаемые формулы.

Перед сложным анализом попросите перечислить листы, заголовки и количество записей. Это помогает сразу заметить, что модель выбрала другой лист, не распознала строку заголовков или приняла промежуточные итоги за обычные записи. Для нескольких таблиц уточните ключ связывания: код товара, номер заявки или другой устойчивый идентификатор.

Не полагайтесь только на цвет ячеек для передачи смысла. Напишите, что означает цвет или пометка, либо добавьте отдельную колонку со статусом. Даты, суммы с валютами и значения с разделителями тысяч лучше описать явно, если файл собран из разных источников.

Пример анализа без изменения исходника

> В @Продажи.xlsx используй лист «Операции». Строка 1 содержит заголовки. Сначала перечисли колонки и уточни, где находятся дата, регион и выручка. Затем посчитай выручку по регионам за второй квартал. Не включай строки с отменённым статусом. Покажи итоговую таблицу и количество учтённых строк. Исходную книгу не изменяй.

В хорошем результате видно не только итоговое число, но и выборка, по которой оно получено. Если данных недостаточно, модель должна обозначить недостающее поле. Уточняйте правило обработки пустых значений: ноль, пропуск и неизвестное значение могут давать разные результаты.

Добавьте расчёт в книгу

Для изменения файла опишите конкретную операцию:

> В @План-факт.xlsx добавь лист «Сводка». Сгруппируй данные по отделам. Добавь колонки План, Факт, Отклонение и Выполнение. Отклонение считай как Факт минус План, выполнение — как Факт делённый на План. При нулевом плане оставляй выполнение пустым. Используй формулы, сохрани лист «Исходные данные» без изменений и верни XLSX.

Если нужны формулы, скажите об этом прямо. Ответ с числами в сообщении и файл с формулами — разные результаты. Также укажите, должны ли новые строки автоматически входить в расчёт, или книга представляет фиксированный отчёт за период.

Для диапазона используйте привычную запись: «лист Бюджет, B2:F40». При переименовании листа предупредите о формулах, которые могут на него ссылаться. Сначала полезно запросить план изменений, если книга включает много взаимосвязанных листов.

Предпросмотр результата

Откройте карточку **«Обновлённая таблица»**. В панели отображается **XLSX**, название, обновление и скачивание. Если в книге несколько листов, переключайтесь между вкладками с их именами. Отсутствие вкладки может означать, что в книге один лист или метаданные ещё загружаются.

Предпросмотр показывает первые **1000 строк и 50 столбцов** листа. При сокращении есть сообщение **«В предпросмотре показаны первые … Скачайте файл для полных данных»**. Это ограничение отображения, а не доказательство удаления остальных данных из книги. Для итоговой проверки скачайте файл и откройте его в Excel или другом совместимом редакторе.

Табличная панель предназначена для просмотра; операции над ячейками задаются через чат. Она не является полноценной заменой интерфейса Excel. Если требуется ручное редактирование большого диапазона, скачайте книгу, выполните его в редакторе и загрузите новую версию.

Проверьте расчёты

Выберите несколько исходных строк и пересчитайте их вручную. Затем сравните общий итог, число записей и обработку исключений. Для процентов проверьте деление на ноль; для средних — исключение пустых ячеек; для дат — границы периода. Если данные объединялись из двух источников, проверьте дубликаты ключей и строки, для которых пары не нашлось.

Попросите модель добавить отдельный лист с методикой: использованные листы, фильтры, правила очистки и формулы. Такой лист полезен для передачи результата коллеге. Он должен описывать фактически выполненный расчёт, а не общую идею анализа.

После скачивания проверьте формулы, значения ошибок и формат чисел. Убедитесь, что денежные величины не превратились в текст, даты не сдвинулись из-за локали, а процент не отображается как обычное число. Для сложных функций, макросов, внешних ссылок и подключений необходима отдельная проверка совместимости; загрузка XLSX не обещает выполнение любого макроса.

Исправление типичных проблем

«Эта таблица пуста» — проверьте выбранный лист и скачанный файл. Рабочие данные могут находиться на другом листе. **«Не удалось загрузить предпросмотр таблицы»** — используйте повтор запроса или скачивание. **Итоги не сходятся** — уточните фильтр, диапазон, типы значений и исключения; не просите просто «подогнать итог».

Если модель изменила исходные данные вместо создания сводки, повторите задачу с исходной копией и явно отделите листы для чтения и изменения. При слишком большом контексте попросите обработать определённые листы или диапазоны. Если у файла есть доступная программная обработка, не требуется вставлять все строки текстом в сообщение.

Размещение платформы и маршруты обработки

Таблицы XLSX доступны и в Pro, и в Advanced; основная разница планов — лимиты использования. Ограничения конкретного файла, модели и административной политики действуют в обоих планах.

В облаке Platform размещается у Nevel, при установке в контуре — у заказчика. Модель, обрабатывающая данные книги, может быть локальной у заказчика, российской провайдерской или глобальной по разрешённому маршруту; в облаке доступна и self-hosted DeepSeek Nevel на серверах в РФ. Табличные инструменты и доступ к источникам настраиваются для установки отдельно.

Только изолированный профиль ограничивает выполнение и источники внутренним периметром. В таком профиле приложите нужную выгрузку либо используйте внутреннюю интеграцию. В остальных вариантах внешняя ссылка книги работает лишь при разрешённом и настроенном подключении.

Следующий рабочий шаг — сохранить и предоставить доступ к файлу или включить его в презентацию. Перед передачей укажите дату источника и период отчёта, чтобы получатель не принял старую выгрузку за текущие данные.

Презентации и PPTX

Подготовка структуры слайдов, создание презентации, точечная правка существующего файла и проверка скачанного PPTX.

Облако / В контуре

https://docs.nevel.sh/platform/presentations/

В режиме **«Презентации»** можно подготовить слайды по заданию и исходным материалам, затем изменить конкретный слайд через чат. Результат открывается как набор превью и скачивается в PPTX. Начинайте с задачи аудитории и структуры сообщения: оформление полезно после того, как понятен смысл каждого слайда.

Создание новой презентации

Нажмите **«+» → «Презентации»**. На стартовом экране этот режим также находится в меню **«Ещё»**. Укажите тему, аудиторию, язык, желаемое число слайдов и исходные материалы. Если нужны актуальные сведения, отдельно попросите использовать разрешённый поиск и привести источники фактов.

Пример:

> Создай презентацию на русском языке для руководителей подразделений о результатах квартального пилота. Используй @Итоги-пилота.xlsx и @Обратная-связь.docx. Нужны 8 слайдов: цель, исходная проблема, участники, выполненные сценарии, измеренные результаты, ограничения, план следующего этапа и решение, которое требуется принять. Тон деловой. Не придумывай проценты и экономию; отсутствующие показатели вынеси в список вопросов.

Для первого прохода достаточно 6–10 слайдов. Текущий инструмент подготовки новой структуры принимает от 2 до 12 слайдов за один вызов. Если нужен большой материал, лучше разделить его на смысловые блоки и работать с ними последовательно, сохраняя единый стиль.

Сначала структура, затем слайды

Попросите сначала показать заголовок и главный тезис каждого слайда. На этом этапе проще убрать повторения, переставить аргументы и сократить лишние детали. Фраза «Слайд 4: результаты» слишком общая; полезнее «Слайд 4: три сценария, которые сотрудники смогли повторить самостоятельно».

СХЕМА 1

Создание презентации



После одобрения структуры попросите собрать презентацию. Разделяйте данные, вывод и предложение: аудитории должно быть понятно, что измерено, что интерпретировано и какое действие предлагается. Если исходные документы расходятся, сначала разрешите расхождение, а не прячьте его за обобщённой формулировкой.

Использование исходных файлов

Прикрепите материалы через загрузку или `@`. Для таблицы укажите нужный лист и период; для документа — разделы, из которых брать тезисы. Если приложен образец презентации, объясните, что требуется сохранить: структуру, композицию, цвета или только тон текста. Образец оформления и источник фактов могут быть разными файлами.

Не рассчитывайте, что модель сама отличит окончательные данные от черновика. Назовите приоритетный источник и запретите заменять отсутствующие значения правдоподобными числами. Для иллюстраций укажите назначение: схема процесса, фото продукта, диаграмма по таблице или декоративная картинка. Сгенерированная иллюстрация не подтверждает существование показанного объекта.

Просмотр презентации

Откройте карточку **«Обновленная презентация»** или созданный PPTX. В панели видны название и отметка **PPTX**. Слайды представлены изображениями с нумерацией. Нажатие на превью открывает увеличенный просмотр с переходом между изображениями.

Просмотрите каждый слайд, а не только первый. Проверьте читаемость мелкого текста, подписи графиков, обрезанные заголовки, изображения и соответствие тезиса структуре. Превью показывает визуальный результат, но не заменяет проверку редактируемости элементов в скачанном файле.

Если панель показывает прежнюю версию, используйте кнопку обновления. После изменения файл может иметь новую версию при том же названии. Поэтому сравнивайте конкретный изменённый слайд, а не только имя карточки.

Точечная правка

Откройте нужную презентацию и напишите номер слайда и действие:

> На слайде 3 замени длинный абзац тремя пунктами. Сохрани заголовок и иллюстрацию. На слайде 5 покажи данные из столбцов План и Факт как сравнительную диаграмму. Остальные слайды не меняй.

Для удаления скажите, какой слайд убрать и почему. Для добавления — укажите место и тезис нового слайда. Поддерживаются операции создания, изменения и удаления слайдов; результат возвращается в обновлённый артефакт. Номер слайда задавайте так, как он показан в интерфейсе, начиная с первого.

Если речь идёт о загруженном внешнем PPTX, сначала попросите перечислить слайды и их содержание. Это позволит убедиться, что файл доступен и выбран правильно. Сложные сторонние шаблоны, анимации и встроенные объекты требуют проверки после изменения: не предполагается безусловное сохранение всех возможностей PowerPoint.

Скачивание и проверка

Нажмите **«Скачать»** в верхней части панели. Откройте PPTX в редакторе, которым будет пользоваться получатель. Проверьте количество и порядок слайдов, шрифты, таблицы, диаграммы и изображения. Если в презентации есть цифры из XLSX, выборочно сопоставьте их с исходными ячейками.

Для передачи материала приложите краткое описание периода и источников. При необходимости попросите добавить слайд с методикой или источниками. Внешние ссылки должны вести на реальные использованные материалы, а внутренние сведения не должны превращаться в публичные ссылки без разрешения.

Неполадки и ограничения

Если получен только план, попросите завершить создание PPTX по этому плану. Если генерация остановилась, продолжайте с недостающего этапа и назовите уже готовый материал. При лимите генерации или редактирования используйте доступные условия обслуживания; бесконечный повтор не снимает квоту.

Создание и изменение презентаций доступно и в Pro, и в Advanced. Для выбора плана оценивайте нужные лимиты работы, а не наличие PPTX: набор инструментов у этих планов одинаков.

Размещение Platform и маршруты подготовки презентации задаются отдельно. И облако, и установка в контуре могут использовать разрешённые модели заказчика, российских провайдеров или глобальные модели; в облаке Nevel доступна также self-hosted DeepSeek на серверах в РФ. Сервисы изображений и поиска выбираются по политике организации.

Для изолированного профиля все компоненты сценария должны находиться внутри разрешённого периметра. Если внешний поиск отключён, создавайте презентацию по приложенным источникам. Для анализа чисел используйте Таблицы, для самостоятельного текста — Документы.

Изображения и видео

Создание изображений по описанию, правки по референсам и короткие видео с начальным и конечным кадром.

Облако / В контуре

https://docs.nevel.sh/platform/media/

Изображения и видео создаются в чате и появляются в нём как отдельные результаты. Вы описываете задачу обычным языком, выбираете режим и после генерации можете продолжить работу с полученным материалом. Для редактирования изображения передайте исходник; для видео можно использовать описание, начальный кадр, конечный кадр или пару кадров.

Создайте изображение

1. Откройте **«Новый чат»** в нужном пространстве.
2. В инструментах поля ввода выберите **«Изображения»**. Если режим скрыт, откройте **«Ещё»**.
3. Опишите назначение, содержание, композицию и формат результата. Например: «Создай горизонтальную иллюстрацию для внутренней презентации: инженер рассматривает цифровую модель оборудования. Светлый фон, спокойные синие акценты, свободное место справа под заголовок, без надписей и логотипов».
4. Отправьте запрос и дождитесь изображения в диалоге.
5. Откройте результат, оцените детали и скачайте нужную версию.

Лучше обозначить реальную задачу, чем перечислять десятки несвязанных художественных эпитетов. «Иллюстрация на титульный слайд, правая треть пустая» помогает композиции; «красиво и современно» оставляет её на усмотрение модели. Соотношение сторон тоже можно указать словами: квадрат для карточки, горизонтальный формат для презентации, вертикальный — для обложки.

Один запрос может включать несколько требований, но противоречащие указания затрудняют результат. Если нужен строго определённый текст на картинке, задайте его дословно и затем проверьте каждую букву. Не используйте сгенерированное изображение таблицы как источник чисел: для расчётов и редактируемых данных предназначены таблицы.

Измените исходную картинку

Прикрепите изображение к сообщению или выберите доступный файл через `@`. Дождитесь завершения загрузки, затем укажите, что изменить и что сохранить:

> На приложенном изображении замени фон на светлый офис. Сохрани сам предмет, его пропорции и направление света. Не добавляй текст. Подготовь горизонтальный вариант для слайда.

Можно использовать несколько изображений как исходные материалы. Назовите назначение каждого: «первое — предмет, второе — пример цветовой палитры». Если материалы похожи, ссылайтесь на имена файлов. Ограничения загрузки и число вложений описаны в работе с файлами.

После первой правки продолжайте тот же чат и явно называйте версию: «В последнем варианте сделай фон светлее; предмет не меняй». Новая генерация может затронуть детали, которые вы не просили менять, поэтому сравнивайте результат с исходником. Для точной фирменной графики отдельно проверьте форму знака, цвета и написание названия.

Инструменты работы с изображениями также позволяют запрашивать увеличение разрешения. Сформулируйте это отдельной задачей: «Увеличь разрешение последнего изображения, сохрани композицию». Увеличение размера файла не восстанавливает достоверно детали, отсутствовавшие в оригинале; для технических схем исходный векторный материал надёжнее.

Создайте видео по описанию

Выберите **«Видео»** в меню режимов. Опишите сцену, движение объекта и камеры, порядок событий. Лучше начать с одного действия: «Камера медленно приближается к изделию на нейтральном фоне; изделие плавно поворачивается». Сложная история из многих сцен удобнее как набор отдельных роликов.

Видео генерируется отдельным инструментом и появляется в чате после завершения работы. Доступная длительность, формат и возможности звука зависят от подключённого генератора; обычный выбор текстовой модели не задаёт эти параметры. Если для задачи необходима определённая длительность или звуковая дорожка, укажите требование в запросе и проверьте фактический результат, прежде чем включать ролик в готовый материал.

Не запускайте повторную генерацию только потому, что текстовый ответ появился раньше видеокарточки. Дождитесь завершения операции или сообщения об ошибке. Повторный запрос — новая работа, которая может учитываться в расходе и лимите генераций.

Видео из кадров

Для управляемого начала прикрепите картинку и напишите: «Используй приложенное изображение как первый кадр. Добавь плавное движение камеры вправо». Для управляемого завершения можно передать конечный кадр. Если нужны оба, загрузите две картинки и однозначно укажите их роли:

> «Начало.png» — первый кадр, «Финал.png» — последний. Создай плавный переход от закрытой коробки к раскрытой. Сохрани положение камеры и фон.

Начальный и конечный кадры — опорные изображения, а промежуточные кадры создаёт модель. Значительные различия перспективы или положения предметов могут привести к заметным переходам и изменению формы. Для устойчивого результата подбирайте кадры одной сцены с похожим ракурсом.

Этот сценарий использует изображения. Он не означает, что обычная форма загрузки принимает произвольные исходные видео для монтажа. Ориентируйтесь на поддерживаемые форматы в окне загрузки и раздел «Файлы».

Скачивание и контроль результата

Откройте созданный материал в просмотрщике и используйте скачивание. Перед отправкой коллегам оцените весь результат: мелкий текст, руки и лица, геометрию предмета, резкие переходы между кадрами. Для иллюстрации продукта сравните ключевые детали с достоверным изображением. Скачанный файл можно включить в презентацию или передать редактору.

Если нужна ещё одна версия, перечислите конкретные изменения одним сообщением. Не смешивайте исправление композиции с новым сюжетом: при новой задаче проще создать отдельный чат. Сохраняйте понятные названия удачных вариантов, чтобы не выбирать финал по порядку миниатюр.

Размещение платформы и маршруты обработки

В облаке генерация использует настроенные для организации сервисы изображений и видео. Доступ к инструменту, расход и квоты определяются конфигурацией и тарифом. Запрос с референсом передаёт его в маршрут генерации, поэтому используйте материалы, разрешённые для этого варианта обработки.

При установке Platform в контуре генераторы изображений и видео могут работать внутри инфраструктуры заказчика либо быть разрешёнными внешними сервисами. Выбор текстовой модели и выбор генератора — разные настройки: DeepSeek или другая текстовая модель не заменяет сервис создания медиа.

В изолированном профиле используются только внутренние генераторы. Если нужный сервис отсутствует или недоступен, администратор проверяет его развёртывание и маршрут; переключение текстового режима на «Думающий» это не исправит. Набор инструментов Pro и Advanced одинаков, а квоты генерации зависят от плана и политики организации.

Если не получилось

Ситуация	Что сделать
Модель описала картинку, но не создала файл	Выберите режим «Изображения» и прямо попросите создать изображение
Изменился не тот исходник	Прикрепите нужную версию и назовите её в запросе
Видео не использует нужный кадр	Укажите имя файла и роль: первый или последний кадр
Получено сообщение о лимите	Посмотрите доступный лимит; повторение того же запроса не сбрасывает квоту
Просмотрщик не загрузился	Повторно откройте результат; если доступно скачивание, проверьте файл локально
Постоянная ошибка генератора	Сохраните текст ошибки и ID запроса, передайте администратору без секретов и лишних исходных материалов

Поиск в интернете и детальное исследование

Как получать актуальные сведения, ставить исследовательскую задачу, проверять источники и превращать отчёт в документ.

Облако / В контуре

https://docs.nevel.sh/platform/web-research/

Для актуальной информации в чате есть **«Поиск в Интернете»** и **«Детальное исследование»**. Поиск подходит для конкретного факта или небольшой подборки источников. Детальное исследование разбивает тему на вопросы, собирает материалы и готовит связный отчёт. Эти режимы обращаются к внешним источникам и отличаются от кнопки **«Поиск»** в боковой панели, которая находит сообщения в истории чатов.

Выберите подходящий режим

Задача	Режим и формулировка
Узнать актуальные условия услуги	Поиск: «Найди действующие условия на официальном сайте и дату публикации»
Сравнить несколько продуктов	Поиск для короткой таблицы; исследование для подробного сравнения с критериями
Подготовить обзор рынка	Исследование: цель, география, период, типы источников и структура отчёта
Найти решение в старом разговоре	«Поиск» в боковой панели или Ctrl/Cmd+K
Проанализировать только загруженный договор	Обычный чат с файлом, без необходимости внешнего поиска

Для справки по одному вопросу не обязательно запускать большое исследование. Если задача предполагает несколько направлений и вывод, сначала сформулируйте критерии сравнения. Без них длинный отчёт может собрать много фактов, но не ответить на деловой вопрос.

Найдите информацию в интернете

Откройте чат в нужном пространстве. Выберите **«Поиск в Интернете»** под полем ввода или в **«Ещё»**. Введите запрос и отправьте его. Укажите период, страну или рынок и предпочтительные источники:

> Найди информацию о трёх изменениях продукта за последний квартал на его официальном сайте. Для каждого укажи дату изменения, что изменилось, ссылку на первоисточник и кому это полезно. Если публикация не содержит даты, отметь это отдельно.

Фразы «последние новости» и «актуальные условия» полезно дополнить конкретным периодом. Для воспроизводимого отчёта используйте абсолютные даты: тогда другой сотрудник сможет повторить поиск с той же рамкой. Отделяйте дату публикации от даты описываемого события: старый материал может рассказывать о будущем изменении, а свежий — пересказывать прошлое.

Ответ должен содержать источники, достаточные для проверки. Откройте ссылки на ключевые факты, особенно на цены, доступность функций и числа. Если ссылка ведёт на обзор, попросите найти первоисточник. Недоступная страница или краткий фрагмент поисковой выдачи дают меньше оснований для уверенного вывода, чем полный документ.

Запустите детальное исследование

1. Выберите **«Детальное исследование»** в меню **«Ещё»**.
2. Опишите решение, которое должен поддержать отчёт, а не только тему.
3. Задайте границы: география, период, перечень продуктов или сегментов, исключения.
4. Назовите критерии и структуру результата.
5. Отправьте задачу и дождитесь итогового отчёта.

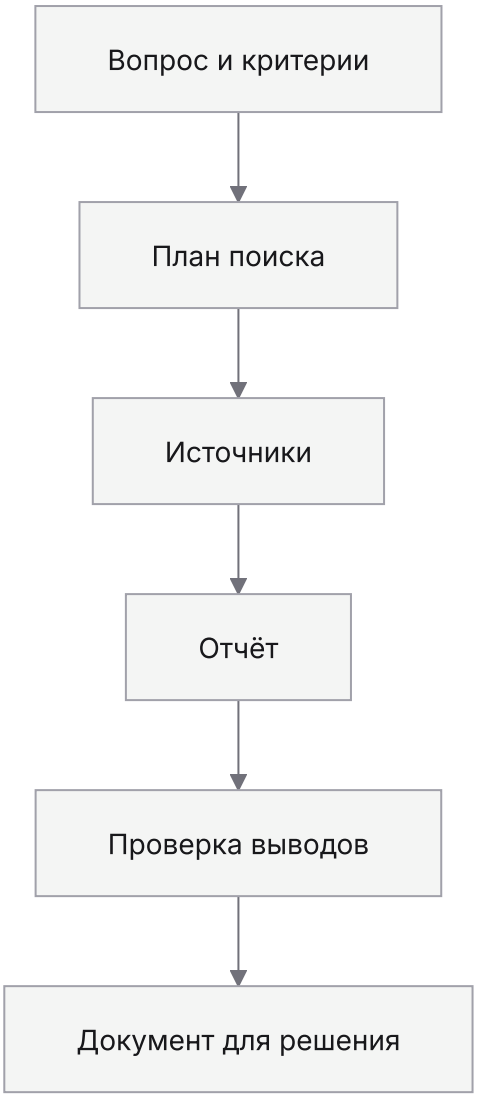
Пример для закупки программного обеспечения:

> Подготовь исследование трёх вариантов системы управления знаниями для компании на 200 сотрудников. Рассмотри корпоративный поиск, разграничение доступа, способы размещения и экспорт данных. Используй официальную документацию и публичные условия поставщиков. Для каждого критерия укажи источник; отсутствие информации отметь как «не найдено». Заверши таблицей различий и вопросами для демонстрации. Бюджет не оценивай по догадке.

Во время работы интерфейс может показывать планирование, дополнительные запросы, сбор и чтение источников, подготовку отчёта. Это этапы одного исследования. Не отправляйте подряд одинаковые задачи: вместо ускорения вы создадите дополнительные работы. Если часть источников пропущена из-за ошибки, итоговый отчёт нужно оценивать с учётом этого пробела.

СХЕМА 1

Работа над исследованием



Используйте материалы компании как контекст

Можно приложить требования, список вопросов или краткое описание проекта. В запросе объясните, какую роль играет файл: «Используй таблицу как критерии; значения в ней не публикуй и не ищи по ним». Для исследования полезнее небольшой подготовленный контекст, чем десятки необъяснённых вложений.

Инструмент исследования получает сформулированную задачу и переданный в неё контекст. Не считайте, что он автоматически читает все файлы пространства или знает все предыдущие обсуждения. Попросите модель сначала перечислить критерии из вложения и исправьте их, если что-то пропущено. Затем запускайте исследование с получившимся перечнем.

Не включайте в поисковые формулировки персональные данные, секретные названия проектов и содержание закрытых документов без разрешённого для них маршрута. Для анализа только внутренних материалов используйте пространство и файлы; внешнее исследование нужно тогда, когда требуется информация снаружи.

Доработайте и оформите отчёт

Продолжайте чат с точными запросами: «Проверь вывод в третьем разделе по официальной документации», «Добавь столбец с датой источника», «Убери рекомендации, которые не следуют из критериев». Если меняется существенное условие, явно назовите его: сравнение решения с разрешёнными внешними провайдерами и изолированного профиля требует разных критериев и

выводов.

После проверки попросите создать DOCX, таблицу сравнения или презентацию. Укажите, какие ссылки и оговорки должны сохраниться. Краткие слайды не должны превращать неопределённый вывод в категоричное утверждение. Для архива можно также экспортировать сам чат в Markdown.

Размещение платформы и маршруты обработки

В облаке интернет-поиск и исследование используют подключённые поисковые сервисы. Доступность и объём использования определяются настройками организации и тарифом. Для исследования могут понадобиться несколько обращений к источникам, поэтому одна деловая задача не обязательно равна одному поисковому вызову.

При установке Platform в контуре интернет-поиск и исследование могут обращаться к разрешённым внешним источникам через настроенный маршрут. Модель для анализа выбирается отдельно: собственная модель заказчика, российский провайдер или глобальная модель. Нужно учитывать и запрос к поисковому сервису, и данные, передаваемые модели.

В изолированном профиле без внешних соединений сайты в интернете недоступны. Аналогичную аналитическую задачу выполняют по локальным документам, внутренним базам и поисковым сервисам. Изоляция — отдельная политика установки, а не обязательное свойство любого on-prem. Кнопка режима не меняет сетевые правила. Поиск и исследование доступны в Pro и Advanced с лимитами соответствующего плана и административной политики.

Разберите неполадки

Если нет ссылок, прямо попросите источники для каждого существенного факта и не используйте неподтверждённые числа как основание решения. Если результаты нерелевантны, сузьте географию и период, добавьте точные названия и исключения. Если сайт закрыт авторизацией, обычный поиск не заменяет подключение к сервису.

При сообщении о лимите остановите повторные попытки и используйте уже собранные материалы либо дождитесь доступного лимита. При частичном результате сохраните отчёт и список нераскрытых вопросов, затем повторите конкретный пробел. Для постоянной ошибки передайте администратору ID запроса, режим и короткое описание задачи; полное содержимое конфиденциальных материалов для первичной диагностики не требуется.

РАБОТА С ПЛАТФОРМОЙ

Интеграции и MCP

Подключение корпоративных сервисов, авторизация, настройка MCP, использование в чате и устранение ошибок.

Облако / В контуре

https://docs.nevel.sh/platform/integrations/

Интеграции дают агенту инструменты подключённых сервисов: искать сведения, читать доступные документы, создавать и изменять объекты. Набор действий зависит от конкретного подключения и прав учётной записи. Интеграция с корпоративным сервисом и API-ключ Nevel работают в противоположных направлениях: интеграция позволяет агенту обратиться к сервису, а API-ключ позволяет вашему приложению обратиться к Nevel.

Откройте интеграции нужного пространства

Выберите пространство в боковой панели, откройте **«Управление»** и вкладку **«Интеграции»**. Используйте поиск по названию. В зависимости от конфигурации каталога вы увидите подключённые сервисы и доступные варианты подключения, категории или фильтр **«Мои интеграции»**. Отсутствующий раздел или сообщение «Интеграции недоступны для этого рабочего пространства» означает, что возможность не включена для текущего рабочего пространства.

Перед подключением убедитесь, что выбрано нужное пространство и правильная учётная запись внешнего сервиса. Например, личный аккаунт в системе задач и корпоративный аккаунт с доступом к проекту могут иметь одинаковое отображаемое имя, но разные данные. Подключение в одном контексте не является основанием считать его доступным в остальных пространствах.

Подключите сервис из каталога

1. Найдите карточку сервиса и нажмите **«Подключить»** или откройте его настройки.
2. Пройдите предложенный способ авторизации. Для OAuth откроется страница сервиса; для других подключений форма запросит ключ или учётные данные.
3. На странице внешнего сервиса выберите нужную учётную запись и прочитайте запрашиваемые права.
4. Завершите авторизацию и вернитесь в Nevel.
5. Убедитесь, что появилась отметка **«Подключено»** или уведомление об успешном подключении.
6. Выполните в чате короткую задачу чтения: «Найди задачу с таким-то названием и покажи её статус и ссылку».

Каталог может включать корпоративную почту, календари, системы задач, базы знаний и другие сервисы. Наличие названия в описании продукта не заменяет карточку в вашем интерфейсе: конкретный список зависит от установки и подключённых провайдеров. Для нужного сервиса используйте доступный в каталоге вариант или собственный MCP-сервер.

Проверочная задача чтения помогает отделить три проблемы: авторизация завершилась, инструмент доступен агенту, учётная запись действительно видит нужный объект. Сообщение «Подключено» подтверждает соединение, но не даёт аккаунту дополнительных прав в исходной системе.

Подключите собственный MCP-сервер

MCP — способ предоставить агенту именованные инструменты вашего сервиса. Для настройки нужен адрес сервера и предусмотренный его владельцем способ авторизации. Адрес обычной веб-страницы, REST-метода или главной страницы сервиса не становится MCP-адресом автоматически.

Откройте **«Создать свою»**. В диалоге **«MCP-интеграция»** заполните **«Имя интеграции»**, понятное **«Описание интеграции»** и URL сервера. Название «Каталог оборудования» полезнее, чем «MCP 2»: агент и пользователи смогут отличить назначение подключения.

Выберите способ передачи секрета:

Тип	Что ввести
Bearer	Ключ, который сервер ожидает как Bearer-токен
Header	Имя и значение заголовка, предусмотренные сервером
Несколько заголовков	Добавьте необходимые пары через «Добавить заголовок»

Не придумывайте названия заголовков по аналогии с другим сервисом. Ключ, выданный для Nevel API, не заменяет ключ корпоративного MCP. Секреты вводятся в поля подключения; их не нужно помещать в описание интеграции, системные инструкции или обычное сообщение чата.

Прочитайте предупреждение в форме. Отметка **«Я доверяю этому MCP-серверу»** относится к выбранному вами серверу и его инструментам. Затем нажмите **«Проверить соединение»**. Успешный ответ показывает число найденных инструментов. Ноль инструментов или ошибка означают, что полезное подключение ещё не подтверждено. Сохраните интеграцию после успешной настройки.

Если владелец сервера изменил набор инструментов, откройте подключение и используйте **«Обновить тулы»**. Это обновляет доступные описания действий; переименование карточки само по себе не меняет возможности сервера.

Поставьте агенту задачу

Укажите сервис, объект, действие и ожидаемый результат. Например:

> В подключённой системе задач найди открытые задачи проекта «Обновление каталога», изменённые за последнюю неделю. Составь таблицу: задача, ответственный, статус, ссылка. Ничего не меняй в исходной системе.

Для изменения объектов задайте точные границы: «В задаче с указанной ссылкой замени срок на 15 сентября; остальные поля не меняй». Если сначала нужен просмотр результата, прямо попросите подготовить черновик и дождаться вашего сообщения перед записью. Интеграции могут содержать действия создания, отправки и удаления, поэтому не рассчитывайте, что любая задача автоматически ограничена чтением.

После выполнения откройте исходный сервис и проверьте объект по ссылке. Ответ агента полезен как отчёт о работе, но для важной операции окончательным результатом является состояние в системе, где выполнялось действие. При ошибке записи сначала проверьте это состояние: повторение запроса может создать второй объект, если первая попытка частично завершилась.

Повторная авторизация и отключение

Если истёк доступ, используйте **«Повторная авторизация»** в настройках подключения. При смене корпоративного аккаунта завершите новую авторизацию под нужной учётной записью и повторите задачу чтения. Для сервиса из блока подключений доступны **«Управлять»** и **«Отключить»**; отключение лишает агента соответствующих инструментов.

Отключение в Nevel не удаляет созданные письма, задачи или документы во внешней системе. Если требуется отозвать разрешение самого провайдера, используйте также настройки подключённых приложений в этом сервисе. Перед отключением проверьте автоматизации, использующие данное соединение: их следующий запуск может завершиться ошибкой.

Размещение платформы и маршруты обработки

В облаке сервис должен быть доступен из маршрута выполнения Nevel. Адрес, который открывается только с ноутбука сотрудника в VPN, не обязательно доступен серверу интеграций. Для внутреннего сервиса используйте предусмотренный организацией канал подключения.

При установке Platform в контуре подключение может вести к внутреннему MCP-серверу и системе заказчика либо к разрешённому внешнему сервису. Для каждого подключения важны адрес, сетевой маршрут, способ авторизации, права аккаунта и передаваемые данные. Размещение самой Platform не запрещает внешнюю интеграцию и не разрешает её автоматически.

В изолированном профиле агент использует только внутренние сервисы и внутреннюю авторизацию. Внешние OAuth-провайдеры и публичные каталоги в этот профиль не входят. Маршрут выбранной модели рассматривается отдельно от маршрута интеграции: прочитанные агентом данные могут попасть в контекст разрешённой модели. Pro и Advanced предоставляют одинаковый набор инструментов интеграции; лимиты и административные разрешения проверяются отдельно.

Если подключение не работает

Симптом	Следующий шаг
«Сессия подключения истекла»	Начните подключение заново из карточки, не используйте старую вкладку авторизации
Ошибка проверки MCP	Проверьте точный MCP URL, тип авторизации и доступность сервера из среды Nevel
Инструменты есть, нужный объект не найден	Проверьте аккаунт, права в исходном сервисе и точность имени или ссылки
После изменения сервера нет нового действия	Выполните «Обновить тулы» и повторите задачу
Действие вернуло ошибку после отправки	Сначала посмотрите состояние объекта в исходном сервисе, затем решайте, нужен ли повтор
Карточки не загрузились	Обновите страницу; при повторении передайте администратору время и текст ошибки без токена

Автоматизации по расписанию

Создание повторяющейся задачи, выбор времени и интеграций, тестовый запуск, просмотр результата и управление лимитом.

Облако / В контуре

https://docs.nevel.sh/platform/automations/

Автоматизация — задача для AI, которая запускается по заданному расписанию. Она содержит название, инструкцию, время запуска и выбранные интеграции. Результаты связаны с чатом задачи. Такой сценарий подходит для регулярного дайджеста, сводки изменений или подготовки повторяющегося отчёта.

Сначала получите правильный результат вручную

В обычном чате выполните будущую задачу один раз. Убедитесь, что доступны источники, понятен период отчёта и результат имеет нужную структуру. Если модель при ручном запуске регулярно спрашивает, какой проект выбрать, ту же неоднозначность нужно убрать из инструкции автоматизации.

Хорошая инструкция называет источник и критерий отбора, объясняет действие и задаёт результат. Например:

> Найди в подключённой системе задач все открытые задачи проекта «Каталог», изменённые за предыдущие семь календарных дней. Подготовь сводку: завершённые задачи, изменения сроков, блокеры. Для каждого пункта дай ссылку. Если изменений нет, напиши это одной фразой. Не меняй задачи и не отправляй сообщения во внешние сервисы.

Слова «сделай как раньше» в регулярной задаче хуже самостоятельной инструкции. Расписание должно быть понятно и после изменения соседних разговоров. Если источников нет или они недоступны, заранее задайте поведение: сообщить о недоступности, не заменять данные догадками и не выдавать прошлый отчёт за новый.

Создайте задачу

1. В боковой панели откройте **«Автоматизации»**.
2. Нажмите **«Новая задача»**. При пустом списке доступна кнопка создания первой задачи.
3. Введите **«Название»**, по которому задача будет понятна в списке, например «Недельная сводка проекта Каталог».
4. Заполните **«Инструкции»** полным текстом задачи.
5. Настройте **«Расписание»** и часовой пояс.
6. В блоке **«Интеграции»** добавьте нужные подключения.
7. Нажмите **«Создать»** и убедитесь, что задача появилась в списке.

Название и инструкция обязательны. Если кнопка создания недоступна, сначала заполните их и проверьте выбранное расписание. Само создание задачи не доказывает, что источники доступны: для этого нужен пробный запуск.

Настройте время

Интерфейс предлагает три вида расписания:

Периодичность	Что задать	Пример применения
«Ежедневно»	Время и часовой пояс	Утренняя сводка открытых обращений

Периодичность	Что задать	Пример применения
«Еженедельно»	День недели, время и часовой пояс	Отчёт в понедельник перед планированием
«Ежемесячно»	Число месяца, время и часовой пояс	Обзор изменений за предыдущий месяц

Смотрите на всю строку расписания: в ней показаны день, время и часовой пояс. «09:00» без пояса недостаточно для распределённой команды. Например, задача сотрудника в другом регионе может запускаться в неудобный для основной команды момент, если оставить автоматически выбранный пояс.

Период данных укажите отдельно в инструкции. Запуск каждый понедельник сам по себе не означает, что модель должна взять прошлую рабочую неделю. Напишите «с понедельника по воскресенье предыдущей недели» или другую нужную границу. При отчёте по нескольким регионам укажите также, в каком часовом поясе интерпретировать даты исходных записей.

Выберите интеграции

Нажмите **«Подключить»** и выберите сервисы, которыми должна пользоваться задача. Можно выбрать **не больше пяти интеграций**; счётчик в форме показывает выбранное количество. Подключение должно быть заранее авторизовано и давать доступ к нужным данным.

Не подключайте сервис только потому, что он есть в каталоге. Для дайджеста задач достаточно системы задач; почта потребуется, только если инструкция действительно предусматривает почтовый источник или действие. Меньший, понятный набор инструментов упрощает проверку результата и диагностику ошибок.

Для отправки письма, публикации сообщения или изменения записей укажите точные действия, получателей и ограничения. **«Тестовый запуск»** выполняет реальную задачу, поэтому он тоже может совершить разрешённые интеграцией изменения. Для первого теста удобно временно ограничить инструкцию подготовкой результата в чате, проверить его и затем добавить нужное действие.

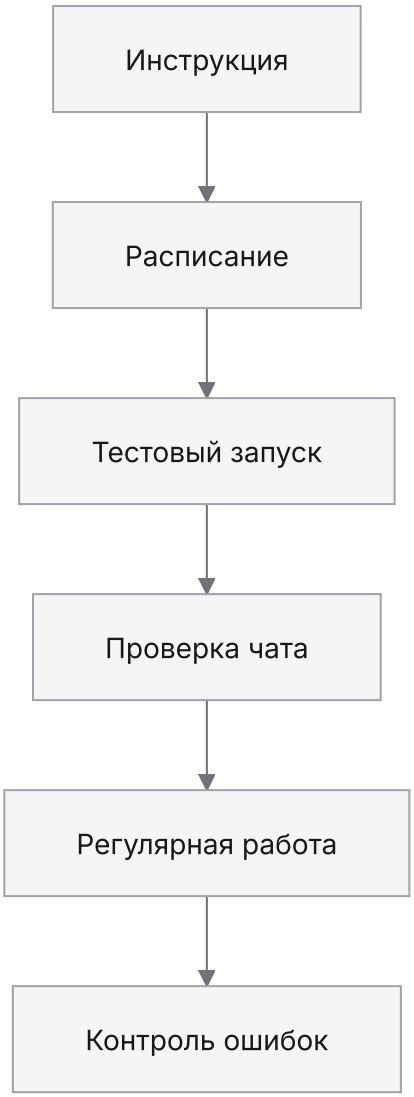
Выполните тестовый запуск

Откройте меню **«Действия задачи»** и выберите **«Тестовый запуск»**. Уведомление **«Тестовый запуск поставлен в очередь»** означает, что запрос на запуск принят; это ещё не готовый отчёт. В списке может появиться состояние **«Выполняется»**.

Откройте задачу, чтобы перейти в связанный чат. Проверьте, что использован правильный источник, выбран нужный период и результат содержит ожидаемые поля. Если автоматизация меняет внешний объект, откройте его в исходной системе. После успешного теста проверьте ближайший запуск по расписанию: он подтверждает уже не только инструкцию, но и выбранное время.

СХЕМА 1

Запуск автоматизации



Измените, приостановите или удалите задачу

В меню задачи выберите **«Редактировать»**, измените поля и нажмите **«Сохранить»**. После изменения источника, инструкции или интеграции повторите тест. Для изменения только времени дополнительно посмотрите итоговую строку расписания.

Переключатель **«Включить или выключить задачу»** управляет активностью. Выключенная задача остаётся в списке: её можно вернуть в работу, не создавая заново. Это удобно для отпусков, временных проверок и сезонных отчётов. Пауза прекращает будущие запуски по расписанию; она не является отменой уже совершённых действий во внешних системах.

Количество одновременно активных задач ограничено тарифом. В разделе показано **«Активно … из …»**. Если лимит исчерпан, поставьте ненужную задачу на паузу, затем включите нужную. При снижении доступного лимита список может показать превышение: сохранённые задачи не следует удалять только ради смены активного набора.

«Удалить» используйте для задачи, которая больше не нужна. Перед этим сохраните необходимую инструкцию и результаты. Повторное создание задачи не должно считаться продолжением прежней истории или отменой её прошлых действий.

Размещение платформы и маршруты обработки

В облаке запуск выполняет серверная инфраструктура Nevel; оставлять открытую вкладку для работы расписания не требуется. Доступность раздела и число активных задач зависят от конфигурации рабочего пространства и тарифа.

При установке Platform в контуре планировщик и платформа работают у заказчика. Автоматизация использует модели и интеграции, разрешённые её политикой: внутренние ресурсы заказчика, российские сервисы или глобальные провайдеры. Для каждого источника нужны работающий маршрут и действующая авторизация.

В изолированном профиле модели и источники находятся внутри периметра. Недоступность интернета нельзя устранить инструкцией «найди свежие новости»; для этого профиля используйте внутренние источники. Автоматизации доступны и в Pro, и в Advanced; количество активных задач и объём выполнения определяются лимитами плана и административной политикой.

Если задача не сработала

При состоянии **«Ошибка»** откройте чат и посмотрите последнюю неудачную операцию. Ошибка авторизации требует обновить подключение, ошибка доступа — проверить права учётной записи, а неверный период — изменить инструкцию. Если задача не запускалась в ожидаемое время, сначала проверьте активность, часовой пояс и вид расписания.

Если тест поставлен в очередь, но долго не начинается, не нажимайте запуск многократно. Сохраните название задачи, время запроса и видимый статус для администратора. При частично выполненной операции проверьте исходный сервис перед повтором, чтобы не создать дублирующее письмо, задачу или запись.

Создание и управление API-ключами

Выдача ключа в интерфейсе, лимит расходов и период сброса, отключение, замена и проверка использования.

Облако / В контуре

https://docs.nevel.sh/platform/api-keys/

API-ключ даёт приложению программный доступ к моделям через Nevel. Ключ создаётся в выбранном рабочем пространстве и должен использоваться с адресом API этой установки. Эта инструкция описывает выдачу и управление ключом в интерфейсе; формат запроса, выбор модели и настройка клиента приведены в первом запросе к API.

Подготовьте назначение ключа

Создавайте отдельный ключ для каждого приложения и среды, если ими нужно независимо управлять. Например, «CRM — тест» и «CRM — рабочая среда» позволяют различать расход и отключить тестовый клиент, не затрагивая рабочий. Один ключ на весь отдел делает анализ и замену сложнее.

До создания определите, кто отвечает за приложение, где хранится секрет и какой расход допустим. Название в Nevel должно помогать найти владельца и назначение, но не содержать сам секрет. Для первого подключения полезен небольшой согласованный лимит, который можно изменить после проверки запросов.

Откройте раздел ключей

Выберите нужное рабочее пространство и откройте его **«Настройки»**. Перейдите в **«API Ключи»**. Раздел доступен пользователям с разрешением управления ключами; его отсутствие у обычного участника не означает, что API не существует. В таком случае ключ выдаёт администратор с соответствующим доступом.

В таблице есть название, сокращённый ключ, создатель, дата создания, сведения об использовании, лимит расходов и статус. Поле **«Поиск по ключу, названию или создателю»** помогает найти запись. Ввод полного секрета для поиска не нужен: используйте название или отображаемый префикс.

Создайте ключ и сохраните секрет

1. Нажмите **«Добавить ключ»**.
2. В окне **«Создать API ключ»** заполните **«Название ключа»**.
3. При необходимости задайте **«Кредитный лимит, USD»**.
4. Если лимит указан, выберите **«Сброс лимита»**.
5. Нажмите **«Добавить ключ»** и дождитесь сообщения **«Ключ успешно создан»**.
6. Скопируйте показанный секрет и поместите его в хранилище секретов приложения.
7. Только после сохранения нажмите **«Готово»** и закройте окно.

Полный ключ показывается один раз. При повторном открытии списка вы увидите сведения о ключе, но не сможете восстановить секрет из его сокращённого значения. Если закрыли окно, не сохранив ключ, создайте новый и удалите или отключите неиспользуемый.

Не вставляйте ключ в публичный репозиторий, браузерный код, пример документации или сообщение обычного чата. Для серверного приложения используйте переменную окружения или секрет, доступный процессу выполнения. В примерах документации переменная называется `NEVEL_API_KEY`; её значение подставляет ваша среда, а не текст страницы.

Настройте лимит расходов

Поле принимает положительную сумму в USD с точностью до шести знаков после десятичной точки. Пустое поле означает **«Без лимита»** для этого ключа. Это не отменяет других ограничений рабочего пространства и доступности моделей.

Сброс	Как работает период
«Ежедневно в 00:00 UTC»	Доступный бюджет ключа обновляется по суточной границе UTC
«Еженедельно в понедельник, 00:00 UTC»	Сброс привязан к началу недели UTC
«Ежемесячно первого числа, 00:00 UTC»	Сброс привязан к календарному месяцу UTC
«Никогда»	Лимит не получает автоматического периодического сброса

Например, тестовый бюджет 10 USD с вариантом «Никогда» задаёт суммарное ограничение ключа, а не ежедневный бюджет. Это пример настройки, а не стоимость тарифа и не обещанное число запросов. Сумма, расходуемая на запрос, зависит от модели и объёма работы.

В таблице могут отображаться остаток, **«Лимит исчерпан»**, дата будущего сброса и зарезервированная сумма. Резерв связан с ещё не завершённым учётом запросов; не сравнивайте его с финальным расходом как с отдельной оплаченной услугой. Подробные ограничения описаны в аутентификации и лимитах API.

В таблице также предусмотрена колонка **«Истекает»**. При этом текущая форма создания ключа не содержит отдельного выбора срока действия. Не планируйте автоматическое истечение по полю, которого нет в форме: для управляемой замены используйте отключение и создание нового ключа.

Проверьте подключение

Используйте выданный вашей установке `NEVEL_BASE_URL`, сохранённый `NEVEL_API_KEY` и доступный `NEVEL_MODEL` в примере API. Адрес страницы браузерной платформы и адрес API могут отличаться; копирование URL открытого чата не настраивает клиента.

Отправьте короткий запрос без рабочих данных. Убедитесь, что получен ожидаемый ответ, затем посмотрите колонку **«Использован»** и аналитику. Наличие строки ключа подтверждает создание, а успешный запрос — работоспособность адреса, секрета и выбранной модели вместе.

Если запрос не проходит, не выдавайте сразу несколько новых ключей. Сначала отличите неверную авторизацию, недоступную модель, исчерпанный лимит и сетевую ошибку. Для диагностики сохраняйте код ответа и ID запроса; передавать полный секрет в обращении не нужно.

Измените, отключите или замените ключ

В меню строки выберите **«Редактировать»**. Можно изменить название, сумму и период сброса, затем нажать **«Сохранить»**. Переименование не меняет сам секрет и не требует изменения конфигурации приложения.

«Отключить» прекращает использование ключа для новых обращений; **«Включить»** возвращает его в работу. Это удобно, когда нужно остановить конкретную интеграцию и сохранить запись для дальнейшего разбора. При подозрении на раскрытие секрета используйте отключение и замену, а не только смену названия.

Чтобы заменить ключ без лишнего перерыва, создайте новый, сохраните его в приложении, выполните проверочный запрос и затем отключите прежний. Если старый секрет уже раскрыт, сначала отключите его, даже если восстановление работы потребует времени. **«Удалить»** применяйте к ключу, который окончательно выведен из использования; убедитесь, что в приложениях не осталось его конфигурации.

Размещение платформы и маршруты обработки

В облаке ключ используется с облачным API и доступными вашей организации моделями. Пользовательский интерфейс помогает управлять его расходом; финансовые условия приведены в тарифах.

При установке Platform/API в контуре ключ выдаётся этой установкой и используется с её адресом API. Облачный ключ нельзя считать действующим в локальной установке, а локальный — в облачной. Хранилище секретов, сетевой доступ и замена ключей организуются для соответствующей инфраструктуры.

Адрес API определяет точку входа, но не место выполнения модели. За ним могут быть разрешённые модели заказчика, российские провайдеры или глобальные модели; в облачном Nevel доступна также self-hosted DeepSeek на серверах в РФ. Убедитесь, какой маршрут соответствует выбранному NEVEL_MODEL. Только изолированный профиль ограничивает выполнение внутренними ресурсами. Сам ключ не выдаёт разрешение на внешний маршрут.

Аналитика использования и расходов

Графики API по моделям, ключам и создателям, активность участников и групп, интерпретация данных и поиск отклонений.

Облако / В контуре

https://docs.nevel.sh/platform/analytics/

Раздел **«Аналитика»** помогает ответить на два разных вопроса: как расходуется API и кто использует платформу. Для этого есть вкладки **«API Ключи»** и **«Пользователи»**. Они описывают использование, а не автоматически рассчитанную экономию рабочего времени или качество ответов.

Откройте данные нужной команды

Выберите рабочее пространство, перейдите в **«Настройки»** → **«Аналитика»**. Доступ к API-аналитике и аналитике участников зависит от разрешений. Если одной из вкладок нет, это может соответствовать вашей роли: право создавать запросы в чате не означает право просматривать использование всей организации.

До сравнения показателей зафиксируйте контекст: рабочее пространство, вкладку, период и группировку. График тестового проекта за неделю и график рабочей организации за месяц нельзя сопоставлять без этих уточнений. Для обсуждения с коллегами достаточно записать выбранные параметры рядом со снимком графика.

Посмотрите расходы API

На вкладке **«API Ключи»** выберите **«Период»**: **«Неделя»**, **«Месяц»** или **«Год»**. Затем задайте **«Группировку»**:

Группировка	На какой вопрос отвечает
«Модель»	Какие модели дают основную долю расхода и запросов
«Ключ»	Какие приложения или среды создают нагрузку, если для них выданы отдельные ключи
«Создатель»	Как распределяется использование ключей между их создателями

На странице представлены три показателя: **«Расходы»**, **«Запросы»** и **«Токены»**. Расход отражается в денежной единице интерфейса, запросы — в количестве обращений, токены — в объёме модельной обработки. Запрос с большой историей или длинным результатом может отличаться по токенам и расходу от короткого вопроса, даже если оба дают по одному обращению в счётчике.

Откройте показатель в расширенном виде, чтобы рассмотреть график и таблицу. Наведение на область графика помогает увидеть отдельные значения; выбор строки или серии выделяет её. В расширенной таблице можно сортировать по колонкам **«Мин»**, **«Макс»**, **«Среднее»** и **«Сумма»**.

Эти статистики относятся к агрегированным точкам выбранного периода и показателя. Например, «Среднее» на графике расходов не следует автоматически читать как стоимость одного запроса. Для сравнения приложений сначала используйте общую сумму и число запросов за одинаковый период, затем при необходимости рассчитывайте дополнительный показатель с явно указанной формулой.

Не теряйте данные в группировке

На компактных графиках показывается ограниченное число крупнейших серий, остальные могут быть объединены как **Others**. Расширенный просмотр отображает больше серий, но также может объединять оставшиеся. Поэтому отсутствие отдельного названия на графике не доказывает нулевое использование.

Смена группировки помогает разобрать суммарный рост. Сначала посмотрите модель, затем ключ, затем создателя. В текущей таблице подпись первого столбца может оставаться **«Модель»**, даже если выбран «Ключ» или «Создатель»; смысл строк задаёт выбранная группировка. Названия самих строк помогут понять, что сейчас сравнивается.

Если в организации один общий ключ, аналитика по нему не восстановит, какое из нескольких приложений отправило запрос. Для дальнейшего разграничения создайте отдельные API-ключи с понятными именами. Исторические запросы не перераспределятся между новыми ключами задним числом.

Разберите неожиданное увеличение расхода

1. Выберите период, включающий день изменения.
2. В группировке «Модель» посмотрите, изменился ли используемый набор моделей.
3. Переключитесь на «Ключ» и найдите приложение, которое дало рост.
4. Сопоставьте «Запросы» и «Токены»: выросло количество обращений или объём каждого сценария.
5. Откройте список ключей, найдите запись по имени и проверьте лимит и последнее использование.
6. При необходимости измените лимит или отключите конкретный ключ, затем проверьте приложение.

Эта последовательность сужает поиск. Она не объясняет сама по себе причину каждого запроса: её находят в приложении, которое использует ключ. Например, повторная отправка после ошибки, слишком длинная история или массовая обработка файлов требуют разных исправлений. Не меняйте все ключи организации из-за одного отклонения, пока данные позволяют выделить конкретный источник.

Посмотрите активность участников

Перейдите на вкладку **«Пользователи»**. Поле **«Поиск по имени или почте»** фильтрует таблицу. В строке показаны пользователь, группа и **«Последняя активность»**: **«Сегодня»**, **«На этой неделе»**, **«На прошлой неделе»**, **«В этом месяце»**, **«Больше месяца назад»** или **«Никогда»**.

Активность формируется по обращениям к модели. Вход в аккаунт или открытие страницы без запроса не обязательно обновит этот показатель. «Никогда» означает отсутствие зарегистрированной активности в этих данных; это не утверждение, что человек не открывал платформу или не читал чужой результат.

Такой список помогает найти сотрудников, которым после приглашения нужен первый рабочий сценарий. Например, отфильтруйте нового участника, предложите задачу по его роли и после выполнения убедитесь, что он смог получить результат. Не превращайте количество запросов в оценку качества работы: один полезный документ и множество неудачных повторов дают разную ценность при любой статистике обращений.

Используйте группы

Если доступны группы, над таблицей пользователей отображается сводка: название группы, **«Количество участников»** и активность за выбранное окно. Можно переключать **«Активность сегодня»**, **«Активность на этой неделе»**, **«Активность в этом месяце»**. Пользователи без группы обозначаются отдельно в таблице участников.

Группы помогают сравнивать внедрение по командам, но сначала приведите их состав в порядок в настройках участников. Число активных людей нужно оценивать вместе с общим числом участников и характером их задач. Одинаковая абсолютная активность в отделах разного размера означает разную долю пользователей.

Размещение платформы и маршруты обработки

В облаке интерфейс показывает данные использования, доступные для выбранной организации. Для финансового контроля сопоставляйте их с условиями тарифа; график сам по себе не заменяет договорные или расчётные документы.

При установке Platform в контуре события использования и аналитические компоненты работают в инфраструктуре заказчика. Источниками использования могут быть как модели заказчика, так и разрешённые российские или глобальные провайдеры. Место сбора аналитики не определяет, где выполнилась модель: это отдельный маршрут.

Если компонент сбора событий недоступен, успешный ответ модели не гарантирует немедленного появления точки на графике. Для сравнения затрат отделяйте расходы на внешнего провайдера от инфраструктурной стоимости собственных моделей. Pro и Advanced сравнивайте по доступным лимитам и реальному использованию, а не по разному набору инструментов.

Если данные не отображаются

Сообщение **«Данные не доступны»** может означать отсутствие данных в выбранном периоде; оно не является гарантией нулевого расхода. Сначала проверьте рабочее пространство, период и вкладку, затем выполните один проверочный сценарий соответствующего типа: API-запрос для API-графика или обращение в платформе для активности участника.

Если сведения не появляются, сохраните время запроса, название ключа или пользователя, выбранный период и снимок экрана. Передайте администратору эти параметры для проверки сбора событий. Не прикладывайте полный API-ключ и содержимое пользовательских запросов к обычному снимку аналитики.

РАБОТА С ПЛАТФОРМОЙ

Настройки аккаунта и администрирование команды

Настройка аккаунта, массовые приглашения, дирекции и группы, импорт оргструктуры, проверка назначений и отзыв доступа.

Облако / В контуре

https://docs.nevel.sh/platform/settings/

Меню аватара в нижней части левой панели ведёт в **«Настройки»**. Здесь доступны личный профиль и, в зависимости от прав, управление рабочим пространством команды, участниками, тарифом, аналитикой и API-ключами. Для настройки конкретного проектного контекста используйте отдельный пункт **«Управление»** выбранного пространства.

Найдите нужный раздел

Раздел настроек	Основная задача
«Профиль»	Имя, аватар, способы входа и удаление личного аккаунта
«Пространство»	Название и аватар команды, доступные сведения о подписке
«Участники»	Активные пользователи, приглашения, роли и группы
«Тариф»	Доступные варианты использования и управление подпиской
«Аналитика»	Расход API и активность участников
«API Ключи»	Выдача и управление программным доступом

Пункты показываются с учётом роли и включённых функций. Отсутствие «Участников» у обычного пользователя не является ошибкой загрузки. Для возврата к работе нажмите **«Вернуться в чат»** в навигации настроек.

Измените профиль

Откройте **«Профиль»**. На странице **«Аккаунт»** доступны изображение профиля и имя. Нажмите редактирование имени, введите нужное значение и сохраните. Имя помогает коллегам находить вас в участниках и отличать владельцев материалов. Оно не меняет email, на который отправлено приглашение.

Для аватара используйте элемент редактирования изображения рядом с текущим профилем. Если загрузка отклонена, выберите поддерживаемый тип и меньший файл; документ или архив не подходят для изображения профиля. После сохранения проверьте отображение в меню аккаунта.

Разовая инструкция для модели не относится к профилю. Постоянные правила проекта задаются в системных инструкциях пространства, а сценарий определённой роли — в ассистенте.

Привяжите и измените способ входа

На странице аккаунта показаны доступные способы входа, включая **Email** и **Telegram**, если они включены в установке. Кнопка **«Привязать»** добавляет способ к текущему аккаунту; **«Отвязать»** удаляет связь.

Для email нажмите «Привязать», введите адрес и подтвердите действие. После сообщения **«Проверьте почту»** откройте письмо и перейдите по ссылке подтверждения. Вернитесь в профиль и убедитесь, что адрес появился. Отправка письма сама по себе ещё не завершает привязку.

Если ссылка истекла, начните привязку заново. Если адрес уже связан с другим аккаунтом, повторение запроса не объединит аккаунты: сначала разберитесь, какой из них содержит ваши рабочие пространства. Для Telegram завершите предложенный переход и авторизацию под нужной учётной записью.

Нельзя удалить последний способ входа: интерфейс показывает **«Нужен хотя бы один способ входа»**. Сначала добавьте и проверьте новый способ, затем отвяжите старый. Для проверки удобно открыть отдельную сессию браузера и войти новым способом, сохранив текущую сессию до завершения проверки.

Пригласите коллег в команду

Откройте **«Участники»**. На странице **«Пользователи»** есть вкладки **«Активные»** и **«Приглашенные»**; при включённой функции групп появляется **«Группы»**. Поиск работает по имени или почте.

1. Нажмите **«Пригласить пользователей»**.
2. В поле **«Электронная почта»** перечислите адреса **через запятую**.
3. Выберите **«Роль»**.
4. Исправьте адреса, отмеченные как некорректные.
5. Нажмите кнопку отправки приглашений и дождитесь результата.
6. Посмотрите вкладку «Приглашенные», а после принятия — «Активные».

Отправляйте приглашение на тот адрес, под которым коллега будет входить. Страница приглашения сопоставляет email: другой авторизованный аккаунт не сможет просто принять приглашение вместо получателя.

При частичной ошибке форма сохраняет адреса, для которых отправка не удалась. Повторите их после устранения причины. На вкладке приглашений используйте **«Отправить повторно»** для конкретной записи; устаревшее или ошибочное приглашение можно удалить. Сообщение «Приглашения недоступны» связано с возможностями текущего рабочего пространства и требует проверки его условий, а не изменения написания адреса.

У одной отправки общая роль. Если часть коллег должна стать администраторами, а остальные — участниками, разделите их на отдельные списки. Не выбирайте административную роль для всего отдела ради нескольких руководителей: должность и право управлять командой — разные признаки.

Перед отправкой уберите лишние пробелы и проверьте корпоративные адреса. Форма распознаёт повторяющиеся одинаковые адреса и показывает некорректные; до исправления ошибочных значений отправка недоступна. Это проверка записи адреса, а не существования почтового ящика или трудоустройства получателя.

Для массового приглашения заранее подготовьте список ожидаемых получателей и после завершения сопоставьте его с результатом. Если часть приглашений уже создана, сначала посмотрите «Приглашенные» и «Активные», затем повторяйте только оставшиеся. Отправленное приглашение, доставленное письмо и принятый доступ — разные этапы. Человек появляется среди активных участников после принятия приглашения; его принадлежность к дирекции и доступ к материалам проверяются отдельно.

Назначьте роли

На уровне команды предусмотрены владелец, администратор и участник. Владелец управляет командой и платёжными действиями. Администратору доступны общие настройки, участники, приглашения и группы в пределах его разрешений. Обычный участник работает с продуктом без этих административных полномочий.

В таблице активных участников нажмите текущую роль и выберите доступный вариант. Роль владельца не меняется обычным переключателем администратора и участника. Перед назначением администратора определите, нужен ли человеку доступ к управлению всей командой или достаточно участия в конкретном внутреннем пространстве.

Роли команды и роли внутреннего пространства — отдельные настройки. Приглашение в команду не должно служить проверкой того, что сотрудник видит каждое пространство. Добавьте его в нужный контекст через вкладку «Доступ» этого пространства, затем проверьте один рабочий сценарий под его учётной записью.

Организуйте группы

Если функция доступна, откройте **«Группы»** и нажмите **«Создать группу»**. Введите название, например «Закупки», и создайте запись. Затем вернитесь к участникам и выберите группу в соответствующей колонке. Для снятия назначения используйте **«Убрать из группы»**.

В колонке участника выбирается одна группа. Её удобно использовать как основную дирекцию или подразделение. Сотрудник при этом может участвовать в нескольких проектных пространствах с разными правами. Межфункциональный проект не требует переименовывать основную дирекцию человека или выдавать ему роль администратора всей команды.

Группы связывают состав подразделений с административной матрицей и сводками активности. Название группы само по себе не открывает документы, ассистентов или инструменты: для этих объектов нужны соответствующие назначения. Переименование «Закупки» в «Снабжение» также не следует использовать как способ скрытого изменения полномочий.

Группу можно переименовать. Удаление группы с назначенными участниками блокируется сообщением **«Невозможно удалить группу с назначенными участниками»**. Сначала перенесите людей в другую группу или снимите назначения, затем удалите пустую запись.

Настройте доступ по дирекциям

Дирекция описывается тремя связанными наборами: **кто входит**, **с какими материалами работает**, **какие действия выполняет**. Это позволяет дать отделу нужных ассистентов и источники, не делая каждого сотрудника администратором. Настройку выполняет владелец или администратор с правом управлять участниками; правила обработки данных определяются действующей политикой организации.

Сначала составьте небольшую матрицу. Ниже — условный пример настройки, а не предустановленные роли Nevel.

Дирекция	Рабочее пространство	Ассистент	Источники
Закупки	«Работа с поставщиками»	«Помощник закупок»	Регламенты закупки, справочник поставщиков
Финансы	«Финансовое планирование»	«Помощник финансов»	Утверждённые формы и бюджетные материалы
Общий проект	«Подготовка плана»	«Редактор проекта»	Только материалы общего проекта

Отдельно укажите разрешённые действия: доступ к инструменту не открывает его источники автоматически.

Назначение	Разрешённые действия в примере	Граница доступа
Закупки	Читать регламенты, готовить документы, работать с таблицами	Не открывает бюджетные материалы финансов
Финансы	Читать свои источники, создавать документы и таблицы	Не даёт управления участниками команды
Участник общего проекта	Работать с опубликованными материалами проекта	Не открывает личные диалоги авторов

Для каждой строки выполните следующие действия:

1. **Создайте группу и включите сотрудников.** В «Участники» проверьте активный аккаунт каждого человека, основную группу и роль. Руководителю направления может хватать прав на его проект; роль администратора команды назначается для административных задач.
2. **Подготовьте рабочее пространство.** Создайте его или откройте «Управление», затем настройте вкладку **«Доступ»** для нужных участников. Порядок работы описан в пространствах.
3. **Подключите только нужные источники.** Проверьте содержимое каждого файла или набора знаний до предоставления доступа. Общая папка с материалами нескольких дирекций не становится подходящим источником только из-за удобного названия.
4. **Предоставьте ассистента.** Определите, кто может пользоваться им и кто поддерживает инструкции. Доступ к ассистенту, документам и подключённым сервисам проверяется отдельно; см. настройку ассистентов.
5. **Примените разрешения на функции и модели.** В административной политике установки задайте доступные дирекции инструменты, подключения и маршруты обработки. Если разрешена работа с таблицами, это не разрешает извлекать любой финансовый файл. Если разрешён веб-поиск, отдельно определите, какие данные могут попасть в поисковый запрос.
6. **Проверьте итог от лица сотрудника.** Откройте нужного ассистента, выполните запрос по разрешённому источнику и создайте требуемый результат. Затем проверьте, что источник другой дирекции и недоступное административное действие остаются

закрыты.

Одна строка матрицы может требовать нескольких назначений: группы участника, доступа пространства, публикации ассистента и политики функций. Не считайте работу законченной только после выбора группы в таблице участников. Критерий завершения — рабочий сценарий выполняется с нужными источниками и без лишних полномочий.

Импортируйте состав и назначения дирекций

Импорт оргструктуры применяется, когда нужно массово создать или обновить состав подразделений и связанные назначения. Он отличается от приглашения списка адресов: приглашение открывает путь в команду, а импорт сопоставляет сотрудников, дирекции и доступ. Процесс состоит из подготовки, сопоставления, проверки, предпросмотра, применения и сверки результата.

Администратор передаёт подготовленную таблицу в настроенный для организации процесс импорта. Используйте формат и способ передачи вашей поставки. Названия столбцов кадровой системы сначала сопоставляются со сведениями Nevel; перечисленные ниже данные описывают смысл, а не обязательные имена полей API или готовый формат загрузки.

Подготовьте исходные данные

Для каждого сотрудника нужны:

- устойчивый идентификатор из принятого справочника, позволяющий отличить его от однофамильца;
- корпоративный адрес для приглашения и сопоставления с аккаунтом;
- основная дирекция и, при необходимости, отдельные проектные назначения;
- требуемая роль команды и права в указанных пространствах;
- действие над существующим назначением: добавить, сохранить, изменить или явно отозвать.

Не используйте отображаемое имя как единственный ключ. Два одинаковых имени не означают одного человека, а смена фамилии не должна создавать второй аккаунт. Если сотрудник меняет адрес, свяжите изменение с его устойчивым идентификатором и проверьте существующий аккаунт до отправки нового приглашения.

Матрицу доступа подготовьте отдельно от кадрового списка. Например, строка «Закупки» в справочнике людей ссылается на уже определённые пространства, ассистентов и функции закупок. Должность «директор» сама по себе не преобразуется в административную роль Nevel.

Сопоставьте существующих и новых сотрудников

В ходе сопоставления отделите уже активных участников, людей с отправленным приглашением и новых получателей. Для существующего сотрудника меняются нужные назначения, без создания дубликата. Для ожидающего приглашения сохраняется связь с текущим приглашением, если получатель и команда верны. Новому сотруднику требуется приглашение, а окончательный доступ проверяется после его принятия.

Если один идентификатор соответствует разным адресам или один адрес связан с несколькими строками с разными ролями, сначала разрешите конфликт в исходных данных. Совпадение имени или похожий адрес не подходят для автоматического объединения. Неясную строку оставьте неприменённой; продолжить её обработку можно после установления правильного аккаунта.

Выберите область обновления

Добавление и обновление, merge: импорт обрабатывает указанных сотрудников и назначения, сохраняя остальные. Отсутствие человека в файле не означает увольнение или исключение из команды. Такой режим подходит для новых сотрудников, смены адресов и уточнения отдельных подразделений.

Замена состава выбранной дирекции: сравнивается полный состав только указанной области. До применения отдельно просмотрите, какие прежние назначения будут сняты. Отзыв права на проект и удаление из всей команды — разные действия; замена состава подразделения не должна смешивать их. Несвязанные назначения общего проекта остаются предметом отдельного решения.

Для обычного обновления начинайте с merge. Полную замену используйте, когда источник действительно содержит полный и актуальный состав выбранной дирекции. Укажите дату выгрузки и область, чтобы неполный кадровый файл не превратился в массовый отзыв.

Просмотрите изменения до применения

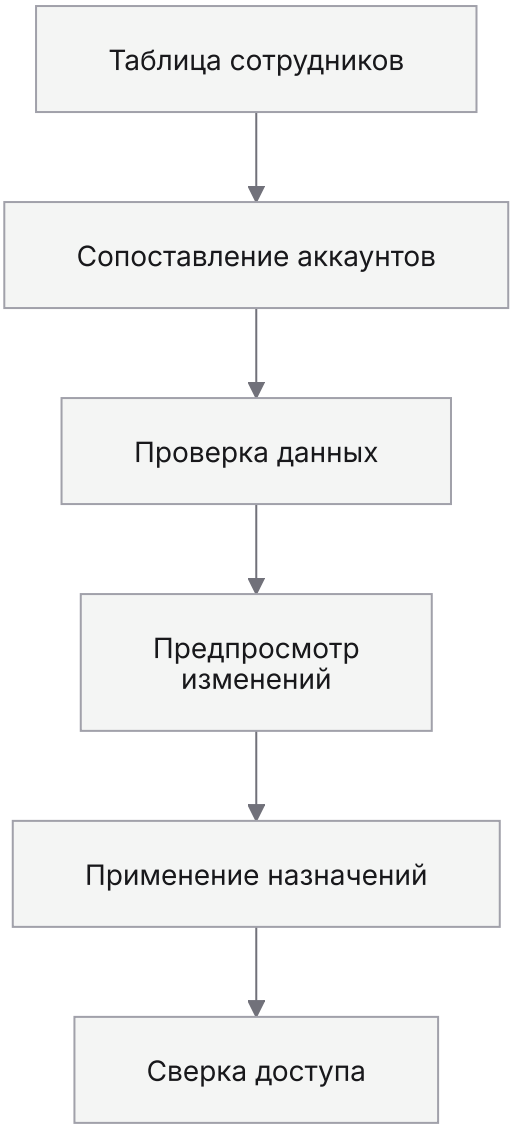
Предпросмотр позволяет сравнить текущее и требуемое состояние. Проверяйте не только количество строк, но и смысл изменений:

Что просмотреть	Что проверить
Новые сотрудники и приглашения	Адрес, команда, роль, отсутствие прежнего аккаунта
Изменение дирекции	Прежняя и новая группа, связанные пространства
Новые полномочия	Источники, ассистенты, инструменты, административные права
Отзыв	Какие назначения снимаются и у кого
Строки без изменений	Повторная загрузка не создаёт новые приглашения и дубликаты
Необработанные строки	Причина конфликта и требуемое исправление

Отдельно просмотрите повышения до администратора, изменение владельцев и массовое снятие доступа. Эти действия не должны быть случайным следствием пустой ячейки или нового названия отдела. Сохраните исходную выгрузку и просмотренный перечень изменений для последующей сверки.

СХЕМА 1

Импорт состава дирекции



Примените и проверьте результат

После применения сопоставьте отчёт с предпросмотром: какие приглашения созданы, какие назначения изменены, какие строки остались без изменений и какие не выполнены. Успешный приём таблицы ещё не означает успешную обработку всех сотрудников. Если операция завершилась частично, сохраняйте результаты успешных строк и повторяйте исправленные строки после сверки текущего состояния.

Повтор той же таблицы должен приводить к тому же составу и полномочиям, без накопления приглашений или повторных назначений. Перед повтором проверьте, не изменилась ли сама исходная таблица и не был ли выбран режим замены вместо merge. Если отчёт отсутствует из-за обрыва связи, сначала посмотрите фактический состав участников; не исходите из предположения, что ничего не применилось.

Проверку завершите несколькими представителями разных изменений: новым сотрудником, переведённым в другую дирекцию и человеком, у которого снят доступ. Проверяется вход, нужный ассистент, разрешённый источник, требуемый инструмент и недоступность прежнего закрытого материала. Правила контроля прав подробно описаны в идентификации и доступе.

Исправьте ошибки импорта

Ситуация	Действие администратора
Повтор сотрудника с разными ролями	Установите правильную запись и устраните противоречие до применения
Неизвестная дирекция	Сопоставьте с существующей группой или создайте нужную группу осознанно
Адрес уже связан с другим аккаунтом	Проверьте идентичность; не объединяйте по имени автоматически
Не найдено пространство или ассистент	Исправьте назначение и убедитесь, что объект существует в нужной команде
Недостаточно полномочий у выполняющего импорт	Передайте конкретные изменения уполномоченному администратору
Ошибка отправки приглашения	Проверьте вкладку приглашений и повторите неуспешную отправку
Обнаружено ошибочное назначение после применения	Снимите именно ошибочные права и проверьте новый запрос под затронутым аккаунтом

Для восстановления используйте сохранённое состояние и обратные изменения назначений. Не удаляйте аккаунты, материалы или всю дирекцию ради исправления одной строки. Исправление состава должно сохранять рабочие объекты команды и происхождение изменения.

Переведите сотрудника в другую дирекцию

Начните с перечня текущих назначений: основная группа, проектные пространства, ассистенты, источники и подключения. Отдельно отметьте временные межфункциональные проекты, которые должны сохраниться после перевода. Затем измените группу, снимите прежние полномочия, которые больше не нужны, и предоставьте новые по матрице принимающей дирекции.

После изменения откройте новый рабочий сценарий под аккаунтом сотрудника. Проверьте и разрешённое действие новой дирекции, и отсутствие доступа по старой ссылке. Изменение названия группы в списке не является достаточной проверкой отзыва. Уже открытая страница может сохранять ранее загруженное отображение; проверяйте новый запрос к источнику, скачивание или запуск инструмента.

Если ошиблись при переводе, восстановите конкретные прежние назначения по сохранённому перечню и снимите лишние новые. Роль администратора команды не используется как временная замена правильно настроенному доступу.

Удалите участника корректно

В меню роли участника выберите **«Удалить»**. Диалог предупреждает об удалении из рабочего пространства и предлагает заранее дать человеку возможность экспортировать нужный контент. Сначала определите, какие материалы нужны команде, и сохраните разрешённые копии.

Удаление из команды не равно удалению всего аккаунта Nevel и отличается от исключения из одного внутреннего пространства. После завершения проверьте состав команды и нужных пространств. Для связанных приложений отдельно пересмотрите API-ключи и интеграции: уход человека не заменяет разбор того, от чьего имени работает автоматизация.

Настройте команду и подписку

В **«Пространство»** владелец или администратор может изменить название команды и аватар. Для пользователя с платёжными правами отображаются сведения о подписке и доступные действия управления. Часть элементов зависит от платёжного способа и состояния подписки.

Тариф, управление подпиской и отмена — отдельные действия. Смена названия команды не меняет её условия. Актуальное описание вариантов приведено в тарифах; для корпоративного договора используйте условия вашей организации. Для программного доступа перейдите к управлению ключами, для контроля использования — к аналитике.

Тема, обновления и поддержка

В меню аватара выберите **«Тема»**: **«Светлая»**, **«Тёмная»** или **«Системная»**. Системная тема следует настройке устройства. Изменение темы касается отображения интерфейса и не меняет созданные документы.

Пункт **«Что нового»** открывает список обновлений; индикатор рядом с аватаром и пунктом меню показывает непрочитанные записи. Если привычная функция переехала, посмотрите обновления и текущий раздел документации. **«Скачать приложение»** открывает страницу загрузок; браузерная работа не требует установки приложения.

Рядом с **«Поддержка»** указан адрес обращения. Для ошибки передайте название раздела, время, последовательность действий и видимый текст ошибки; для чата добавьте ID запроса, если он показан. **«Выйти»** завершает текущую сессию. Это не удаление аккаунта и не отмена подписки.

Удаление аккаунта и варианты размещения

«Удалить аккаунт» находится в профиле и требует ввода **УДАЛИТЬ**. Операция необратима. Если вы владеете рабочим пространством, интерфейс может потребовать сначала удалить или передать его; обычный выход из браузера эту задачу не решает.

В облаке используются включённые организацией способы входа и административные возможности. **При установке Platform в контуре** доступ и администрирование выполняются через инфраструктуру заказчика; модели и внешние подключения разрешаются отдельной политикой. В обоих вариантах размещения можно выбрать допустимые для организации маршруты выполнения моделей.

Наличие Email или Telegram в облачной форме не означает обязательную доступность внешнего провайдера в **изолированном профиле**. Используйте выданный организации адрес и настроенный способ входа. В облаке и on-prem одинаково проверяйте идентичность, группу и фактические назначения; место размещения не заменяет матрицу доступа.

Популярные модели

Десять моделей из недельного рейтинга: возможности и базовые цены для сравнения.

Облако / В контуре

https://docs.nevel.sh/platform/model-catalog/

Сравните популярные модели по поддерживаемым входным данным, работе с инструментами и базовой цене. Место в рейтинге отражает использование в источнике; качество для вашей задачи проверяется на собственных примерах.

Топ-10 за неделю

Рейтинг на 6 ноября 2026. Цены на 7 ноября 2026, 12:25 МСК.

Цена указана в долларах США за 1 млн токенов: **вход / выход**. Это базовые цены моделей в источнике, без наценки Nevel; условия использования через Nevel API описаны отдельно.

№	МОДЕЛЬ	ВОЗМОЖНОСТИ	ВХОД / ВЫХОД, $
1	**DeepSeek V4.1 Flash 1017**	текст, изображения вызов инструментов, рассуждение, структурированный ответ	от $0,15 / от $0,6
2	**DeepSeek V4.1 Flash**	текст, изображения вызов инструментов, рассуждение, структурированный ответ	$0,051 / $0,567
3	**GLM 5.3 Flash**	текст, изображения, видео вызов инструментов, рассуждение, структурированный ответ	$0,15 / $0,5
4	**GPT-6 Luna**	файлы, изображения, текст вызов инструментов, рассуждение, структурированный ответ	от $0,1 / от $0,5
5	**Hy4**	текст, изображения вызов инструментов, рассуждение, структурированный ответ	$0,417 / $1,2505
6	**Step 5 Preview**	текст, изображения, видео вызов инструментов, рассуждение, структурированный ответ	$1 / $2,7
7	**Claude Opus 5.5**	текст, изображения, файлы вызов инструментов, рассуждение, структурированный ответ	$4 / $20
8	**MiniMax M3.1 Flash**	текст, изображения, видео вызов инструментов, рассуждение, структурированный ответ	от $0,075 / от $0,3
9	**Hy4 preview**	текст вызов инструментов, рассуждение, структурированный ответ	от $0,7506 / от $2,2509
10	**GPT-5.6 Luna**	файлы, изображения, текст вызов инструментов, рассуждение, структурированный ответ	от $0,2 / от $1,2

Источник рейтинга: OpenRouter, 2026-11-06 · CC BY 4.0.

Как читать сравнение

- **Вход** — данные запроса, **выход** — сгенерированный ответ. Кэш, дополнительные инструменты, отдельные операции и условия выбранного провайдера могут менять итоговую стоимость. «От» отмечает модель с дополнительными ценовыми условиями, например для длинного запроса или времени обработки.
- **Вызов инструментов** означает поддержку соответствующего интерфейса моделью. Доступ к файлам, источникам и действиям в Nevel определяется правами пользователя и настройками организации.
- **Структурированный ответ** означает заявленную поддержку параметра structured outputs. Это характеристика API модели, а не гарантия правильности данных в ответе.
- **Бесплатный вариант** — конкретный endpoint в источнике. Нулевая базовая цена не обещает бесплатное использование Platform или доступность этого варианта в вашей установке.

Выбор модели в организации

Этот рейтинг помогает сравнить модели и не является списком разрешённых обработчиков вашей организации. В меню Nevel показывается каталог установки; внешний API может выдавать иной набор доступных идентификаторов. Для рабочего запроса выберите разрешённую модель или режим Auto.

Размещение Platform и место выполнения модели выбираются отдельно: доступны модели заказчика, модели Nevel на серверах в РФ, российские и глобальные провайдеры в пределах политики организации. Популярность модели не меняет правила маршрутизации и защиты данных.

Источники характеристик

Рейтинг охватывает семь завершённых суток публичного трафика источника. Цены и возможности: каталог моделей и карточки отдельных вариантов.

- DeepSeek V4.1 Flash 1017
- DeepSeek V4.1 Flash
- GLM 5.3 Flash
- GPT-6 Luna
- Hy4
- Step 5 Preview
- Claude Opus 5.5
- MiniMax M3.1 Flash
- Hy4 preview
- GPT-5.6 Luna

БЕЗОПАСНОСТЬ

Безопасность Nevel

Модель защиты Platform и Guardrail/API, состав контролей, ответственность и программа проверки внедрения.

Облако / В контуре

https://docs.nevel.sh/security/overview/

Nevel защищает работу сотрудников и приложений с моделями на нескольких уровнях: проверяет права пользователя, ограничивает доступ к контексту, обнаруживает чувствительные сведения, определяет допустимый маршрут обработки и регистрирует события. Guardrail — часть этой системы. Он не заменяет контроль доступа к документам, защиту серверов, правовое основание обработки или проверку результата человеком.

Раздел описывает облачное размещение Platform и on-prem — размещение Platform в инфраструктуре заказчика. Модельный маршрут выбирается отдельно: локальная модель заказчика, российский провайдер, self-hosted модель Nevel на серверах в РФ или глобальная модель. On-prem может использовать разрешённые внешние маршруты. Полностью внутренняя обработка задаётся отдельным изолированным профилем.

Как устроена защита

Защита начинается до обращения к модели. Запрос получает идентификатор, привязывается к пользователю или сервисному ключу и рабочему пространству. Сервер проверяет полномочия на диалог, проект, файл и подключение. Только разрешённые данные попадают в контекст. Затем Guardrail анализирует весь подготовленный контекст, применяет маскирование и классификацию и выбирает разрешённый маршрут. Ответ проходит обратную обработку; события решения остаются в журнале.

СХЕМА 1

Уровни защиты Nevel

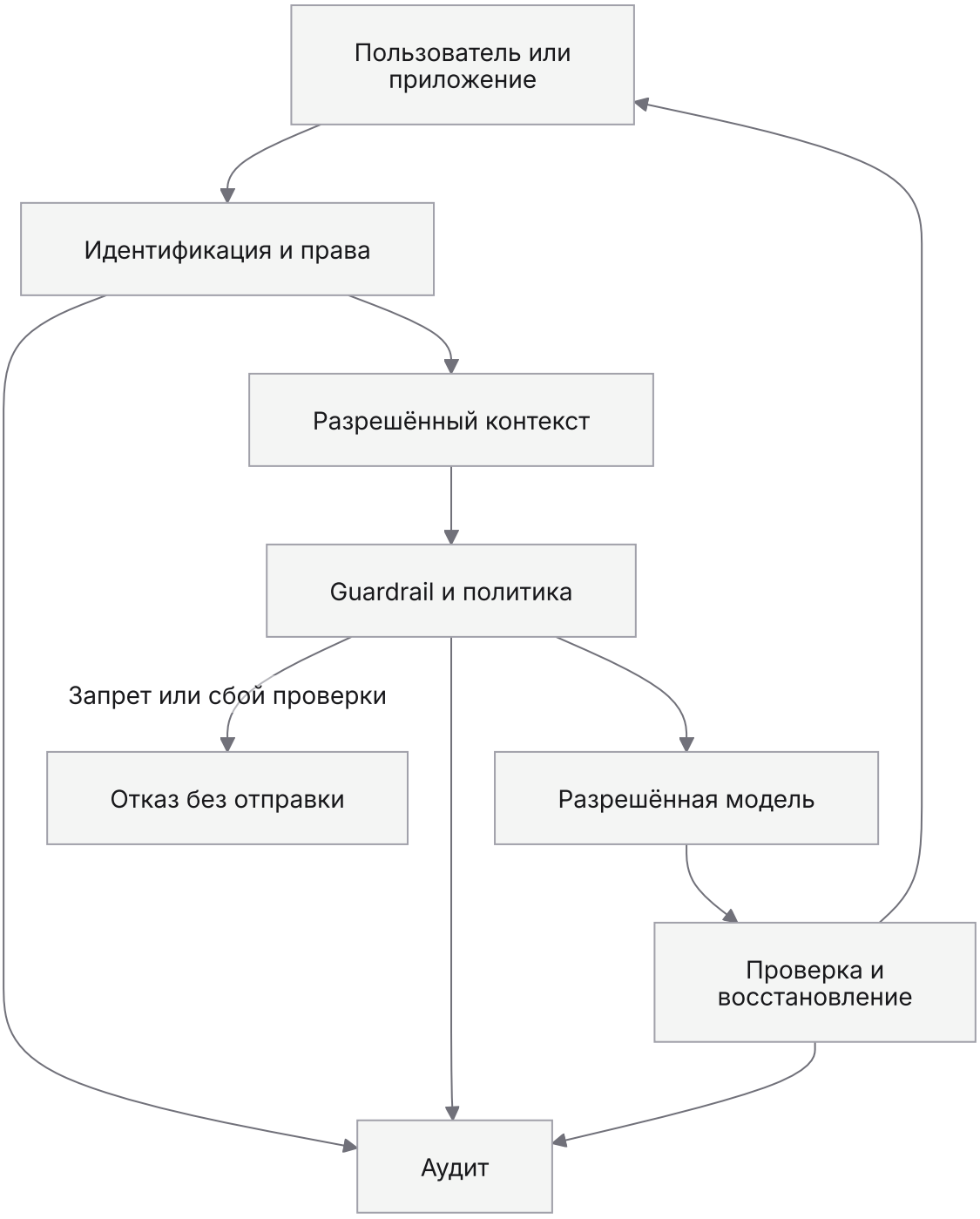


Схема показывает логические проверки, а не обещание одинаковой структуры серверов во всех поставках. Для архитектурной экспертизы существенны точки доступа к открытому тексту, получатели данных и возможность обхода обязательного контроля. Эти границы раскрыты в архитектуре и потоках данных.

Что именно защищается

Объект	Примеры чувствительного содержания	Основной контроль
Сообщение и история	ФИО, условия сделки, реквизиты, внутренние решения	Права на диалог, анализ полного контекста, политика маршрута
Файл и извлечённый текст	Договор, таблица выплат, комментарии, скрытые листы	Изолированная обработка, классификация, ограничение извлечения
База знаний	Фрагменты документов, векторы, заголовки, ссылки на источник	Фильтрация по правам до выдачи контекста, удаление производных данных
Результат инструмента	Ответ CRM, SQL-выборка, письмо, запись календаря	Полномочия подключения, минимизация результата, повторный контроль
Таблица соответствий масок	Исходное значение и его заменитель	Изоляция, ограниченный срок жизни, доступ только к своему запросу
Секреты	Ключ модели, токен подключения, сертификат	Шифрованное хранение, ограничение использования, отзыв
Журнал и резервная копия	Идентификаторы, метаданные, иногда содержание	Отдельные права, сроки хранения, защита экспорта и восстановления

Даже после удаления фамилии содержание может оставаться коммерческой тайной или позволять определить человека по контексту. Документ о единственном сотруднике с редкой должностью не становится безопасным только из-за замены его имени. Для таких случаев используется ограничение маршрута целого документа, вплоть до обработки локальной моделью или блокировки. Подробнее — Guardrail и 152-ФЗ.

Место платформы и место модели

Вопрос	Облако	On-prem
Где работают компоненты Platform	В инфраструктуре Nevel	В инфраструктуре заказчика
Какие модельные маршруты используются	Подключённые и разрешённые политикой	Подключённые и разрешённые политикой, включая внешние
Кто контролирует серверы платформы	Эксплуатация облачной поставки	Заказчик или его назначенный исполнитель
Где находятся рабочие данные и соответствия масок	В защищённых хранилищах облачной поставки в РФ	В хранилищах установленной платформы у заказчика
Кто разрешает подключения	Администратор в пределах политики организации	Администратор заказчика в пределах политики организации
Как проверяется граница	Фактические получатели и сетевой след	Внутренние и разрешённые исходящие обращения, фактические получатели

Маршрут может вести к модели внутри контура заказчика, к Яндексу или Сберу, к self-hosted экземпляру Nevel в РФ, например DeepSeek, либо к глобальному провайдеру. Их различия и типовые сочетания с местом установки показаны в матрице вариантов. Доступность конкретного маршрута зависит от подключений и политики.

В изолированном профиле все рабочие этапы остаются внутри заданного контура: модели, OCR, embeddings, инструменты, данные, аудит и копии. Отсутствие внешних обращений проверяется именно для этого профиля. Для обычного on-prem проверяется, что наружу идут только разрешённые запросы, а не полное отсутствие внешней связи.

Размещение сервера в России не определяет местоположение выбранной модели. Передача российскому провайдеру может пересекать границу инфраструктуры заказчика, оставаясь обработкой в РФ. Юридическая оценка трансграничной передачи зависит от фактической цепочки, а не от слова on-prem. Маскирование само по себе не определяет её допустимость.

Карта технических мер

Задача корпоративной ИБ	Где получить самостоятельный ответ
Понять, где появляется открытый текст и кто его видит	Архитектура, потоки
Разобраться в TLS, шифровании хранилищ и ключах	Шифрование и ключи
Подключить корпоративный вход и разделить полномочия	Идентификация и SSO
Проверить маскирование, классификацию и ошибки	Guardrail и маршрутизация
Проверить файлы, базы знаний и защиту от отравления контекста	Файлы и RAG
Встроить собственный AI firewall и закрыть обходные пути	Сетевая защита
Включить российские SIEM, DLP, сетевой шлюз и IdP	Корпоративная среда
Ограничить prompt injection в файлах, веб-страницах и инструментах	Проверка инструкций и действий
Получить события для расследования	Доступ и аудит
Настроить доставку, нормализацию и проверку событий SIEM	Мониторинг и SIEM
Установить сроки и проверить уничтожение	Хранение и удаление
Проверить сценарии атак и остаточный риск	Модель угроз
Соотнести технические меры с российскими требованиями	Меры защиты в РФ, 152-ФЗ

Распределение ответственности

Владелец бизнес-процесса определяет цель использования, допустимые сведения и операции, которые требуют проверки человеком. Юридическая служба определяет основания обработки, роли сторон и допустимость передачи. Служба ИБ задаёт классификацию, обязательные маршруты, требования к контролю и критерии приёмки. ИТ отвечает за учётные записи, рабочие устройства и интеграции; при установке в контуре — также за серверы, сеть, обновления и восстановление.

Команда Nevel обеспечивает описание и настройку компонентов поставки, исправление продукта и сопровождение в пределах договора. Ни администратор платформы, ни инженер поддержки не должны автоматически получать право читать все корпоративные документы только в силу технической роли. Административное управление, доступ к содержимому и расследование разделяются.

Правовые роли зависят от фактических целей и действий сторон. Заказчик обычно определяет цели использования корпоративных данных, а исполнитель действует в пределах поручения. Для конкретного внедрения это закрепляется в пакете документов; поручение должно описывать фактические операции и участников.

Программа приёмки

Испытание строится на синтетических данных. Один положительный запрос подтверждает доступность сервиса, но не подтверждает изоляцию, отсутствие обхода или качество распознавания. В минимальный набор входят:

1. Обычный запрос с разрешённой моделью и воспроизводимым результатом аудита.
2. Запрос с несколькими типами чувствительных сведений в сообщении, истории и файле.
3. Попытка получить чужой документ, указав его известный идентификатор.

4. Запрет внешней модели для локального класса данных, включая автоматический выбор и повторную попытку.
5. Отключение DLP, классификатора, обязательного средства защиты и хранилища соответствий.
6. Отзыв пользователя и ключа во время уже начатой рабочей сессии.
7. Удаление документа с последующей проверкой поиска, контекста, ссылки на файл и резервного восстановления.
8. Проверка журналов на отсутствие секретов и исходных значений масок.

Протокол содержит версию продукта, конфигурацию политики, набор входов, ожидаемый и фактический результат, время и доказательство. После изменения модели, извлечения файлов, прав или сетевого маршрута повторяются затронутые проверки. Согласование архитектуры не заменяет испытание конкретной установленной конфигурации.

Показатели поставки и подтверждающие документы

Числовые параметры базового профиля применяются только при указанных для него условиях и не считаются автоматически показателями любого окружения заказчика. Результаты конкретной поставки подтверждаются её испытаниями. Сертификаты, аттестаты, уровень защищённости и соответствие требованиям подтверждаются отдельными документами с областью действия; наличие Guardrail само по себе их не доказывает.

Применимые нормативные основания и версии актов приведены в мерах РФ и юридическом разделе.

БЕЗОПАСНОСТЬ

Архитектура и границы доверия

Компоненты Platform и Guardrail, разделение контуров управления и данных, изоляция организаций и два варианта размещения.

Облако / В контуре

https://docs.nevel.sh/security/architecture/

Архитектура безопасности Nevel описывает не только соединение интерфейса с моделью. В обработке участвуют авторизация, история диалога, файлы, поиск по знаниям, подключения, Guardrail, модель, обратная обработка и аудит. Каждый компонент имеет собственный доступ к данным и может стать точкой обхода, если рассматривать только основной чат.

Для архитектурного обследования используются три понятия. **Контур данных** обрабатывает сообщения, файлы и ответы. **Контур управления** изменяет пользователей, роли, подключения, модели и политики. **Контур эксплуатации** обслуживает серверы, секреты, журналы и резервные копии. Их логическое разделение позволяет не выдавать человеку право читать документы только потому, что он управляет настройками.

Компоненты и полномочия

Компонент	Что получает	Что выполняет	Какие полномочия ему не нужны
Веб-приложение	Ввод пользователя и доступные ему данные	Отображает диалоги, проекты, файлы и результаты	Секреты провайдеров и доступ к серверным хранилищам
Сервер Platform	Идентичность, контекст пространства, выбранные объекты	Проверяет права, собирает контекст, запускает инструменты	Неограниченный доступ к чужим организациям в пользовательской операции
API Gateway	Запрос, ключ, метаданные, параметры модели	Аутентификация, лимиты, идентификатор запроса	Разрешение менять политики на основании текста запроса
DLP/NER	Текст, подлежащий отправке	Находит сущности, заменяет значения и создаёт соответствия	Управление пользователями или маршрутом вне политики
Классификатор	Содержание и обязательные метки	Определяет категорию документа и допустимость маршрута	Снятие административного запрета по просьбе модели
Model Router	Решение политики, подготовленный запрос	Выбирает разрешённый обработчик и резервный маршрут	Выбор запрещённого получателя ради доступности
Хранилище соответствий	Маски и исходные значения в границе запроса	Даёт восстановить разрешённый ответ	Публичный поиск по любому идентификатору маски
Audit	Метаданные решения и действий	Обеспечивает расследование и экспорт событий	Неограниченное накопление содержимого по умолчанию
Обработчики файлов и инструменты	Только необходимые входные данные	Извлечение, поиск, вычисления, подготовка артефактов	Неограниченная сеть, секреты других подключений

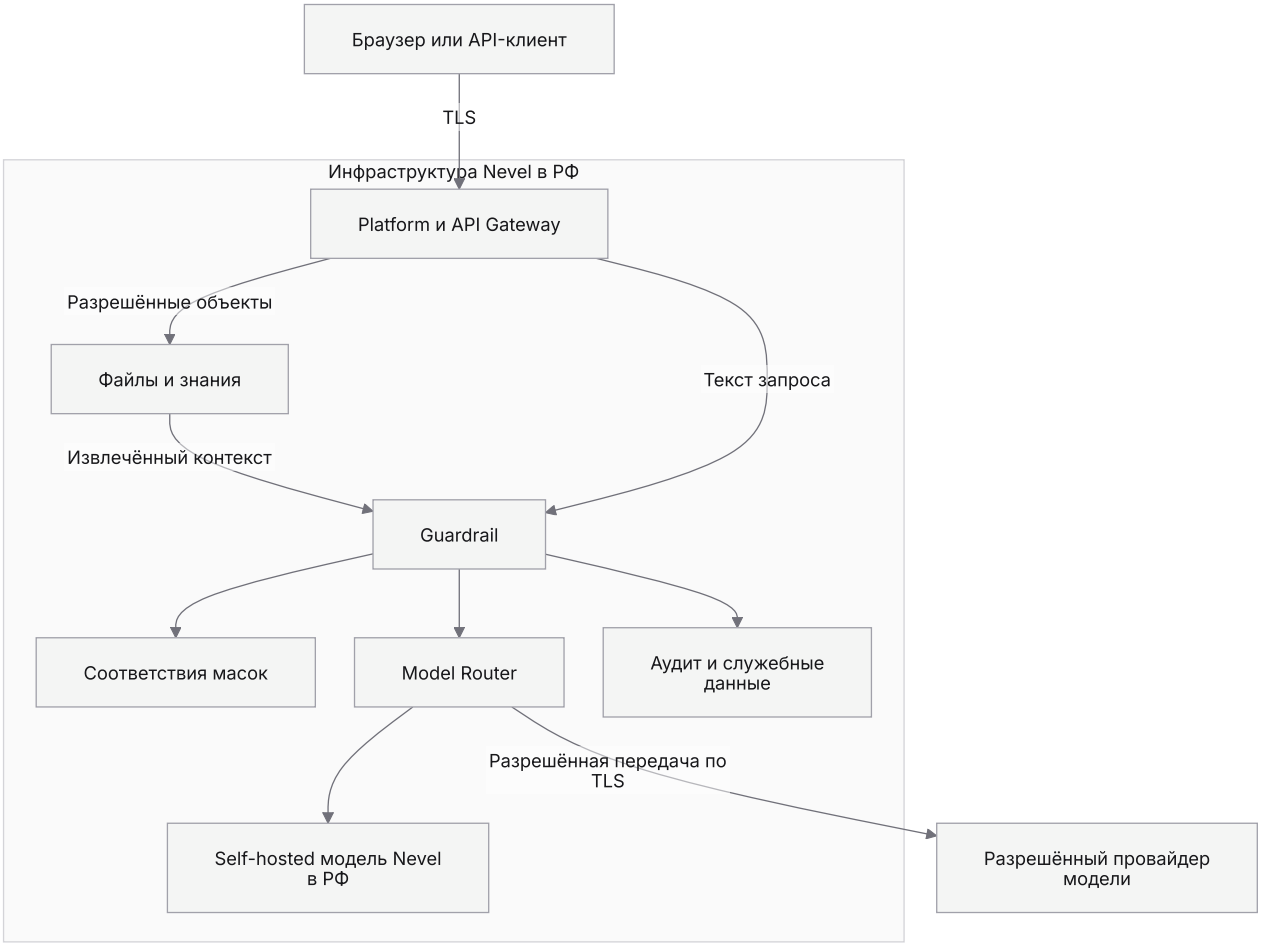
Конкретные процессы могут размещаться совместно или раздельно. Совместное размещение не отменяет разграничения сервисных полномочий и не делает внутренний сетевой вызов доверенным автоматически.

Облачная архитектура

В облачной поставке основные компоненты Platform и Guardrail, хранилища соответствий и служебные данные размещаются на российской инфраструктуре. Модельный маршрут может использовать self-hosted экземпляр Nevel на российских серверах, например DeepSeek, российского провайдера, например Яндекс или Сбер, либо разрешённую глобальную модель. Получатель и место обработки фиксируются отдельно от места Platform. Обращение к отдельному провайдеру учитывается как выход за границу инфраструктуры платформы; его география проверяется отдельно.

СХЕМА 1

Облачная архитектура



Стрелка к внешнему получателю обозначает реальную передачу подготовленного содержания. TLS защищает канал, но получатель расшифровывает запрос для обработки. Маски уменьшают объём раскрываемых идентификаторов, однако сам текст, структура таблиц, смысл документов и ответы инструментов тоже могут быть чувствительными.

Если между маршрутизатором и моделью присутствует дополнительный сервис, он включается в ту же цепочку получателей. Для каждого звена нужны назначение, география, состав получаемых данных и правила хранения. Утверждение о режиме конечной модели нельзя автоматически распространить на промежуточное звено. Результат обработки проверяется по фактическому маршруту запроса, а не по названию модели в интерфейсе.

Архитектура on-prem

При on-prem Platform, её установленные защитные компоненты и рабочие хранилища находятся в инфраструктуре заказчика. Модели размещаются независимо: внутри контура, у российского провайдера, в self-hosted инфраструктуре Nevel в РФ или у глобального провайдера. Исходящие обращения разрешаются по категории данных и конкретной цепочке получателей. Такая конфигурация остаётся on-prem, поскольку термин определяет место платформы.

Схема включает внутренний адрес модели и каждый разрешённый внешний endpoint, точки TLS, секреты, Guardrail и аудит. Если организация требует собственный AI firewall, он включается в обязательный маршрут. Российский провайдер за пределами инфраструктуры заказчика является внешним получателем по этой границе, хотя обработка может оставаться в России. Схема on-prem с несколькими маршрутами приведена в инструкции установки.

Изолированный профиль

Изолированный профиль является частным случаем on-prem. В контуре находятся приложение, шлюз, Guardrail, модели, извлечение текста, OCR, embeddings, поиск, инструменты, хранилища, секреты, аудит и резервирование. Все нужные рабочие сценарии завершаются внутри этой границы.

Внешние модели, облачный OCR, внешний поиск и телеметрия с содержимым не входят в изолированный рабочий маршрут. Если внутреннего средства недостаточно, операция завершается отказом либо выполняется другой разрешённой внутренней моделью. Недоступность компонента не меняет этот профиль на внешнюю обработку.

Обновления и поддержка являются отдельными административными операциями. Пакеты переносятся по процедуре заказчика, а доступ выдаётся конкретному исполнителю на ограниченную задачу и протоколируется. В обычном on-prem разрешённые каналы обновления также фиксируются, но постоянное отсутствие внешних соединений не является свойством всех установок.

Где существует открытый текст

Открытый текст доступен пользователю, серверу при подготовке запроса, DLP и классификатору, разрешённому обработчику файлов и модели, которая должна его обработать. В защищённом внешнем маршруте модель получает результат маскирования; в хранилище соответствий остаётся связь с исходными значениями. При восстановлении результата исходные значения снова появляются на стороне обратной обработки и в ответе уполномоченному пользователю.

Шифрование диска не скрывает данные от работающего приложения с разрешённым ключом. TLS не скрывает содержание от точки завершения TLS. Поэтому архитектурная защита включает права на сервисы, ограничение административного доступа, контроль памяти и диагностических выгрузок, минимизацию логов и разделение секретов. Подробности — шифрование и ключи.

Изоляция организаций и пространств

Граница доступа устанавливается сервером по проверенной учётной записи и членству. Переданный клиентом идентификатор организации, проекта или файла служит ссылкой на объект, но не доказательством права. Проверка выполняется при чтении, поиске, скачивании, изменении, экспорте и выполнении инструмента.

Изоляция распространяется на менее заметные объекты: кэш результатов, векторные фрагменты, соответствия масок, очередь фоновых задач и временные ссылки на файлы. Кэш без контекста полномочий может раскрыть результат, полученный ранее другим пользователем. Фоновая задача не получает бессрочное право только потому, что её когда-то создал администратор: перед выполнением учитываются актуальные права и состояние подключения.

Организационная группировка не подменяет права. Название отдела или тег пользователя помогает администрированию и аналитике, но доступ к данным определяется явным правилом. Публичность ассистента внутри пространства также не делает автоматически публичными его исходные документы или персональные подключения.

Ошибки и резервные маршруты

Недоступность основной модели допускает только резервный вариант из того же разрешённого класса обработки. Нельзя переключать документ с обязательной локальной обработкой на внешнюю модель. Тайм-аут DLP, неизвестная категория или отказ обязательного AI firewall не превращаются в разрешение отправить исходный текст.

Ошибка должна сохранять идентификатор запроса и стадию отказа, не раскрывая секреты и исходные значения в сообщении клиенту. Повторная попытка проходит те же проверки; изменение модели, тела или подключённого контекста является новым решением политики. Отмена пользователем прекращает дальнейшую работу насколько это возможно, но не означает отзыв данных, уже полученных внешним обработчиком.

Что проверяется в архитектуре

Контрольная схема содержит список компонентов, адресуемые границы, хранилища, точки TLS, сервисные идентичности, источники секретов и разрешённые исходящие соединения. Для каждой стрелки фиксируются категория данных, инициатор, получатель, протокол, возможность хранения и поведение при отказе.

Практическая проверка включает обращение к чужому объекту, работу после отзыва роли, запрос через альтернативный API-маршрут, принудительный отказ DLP и модели, а также восстановление из резервной копии. Для закрытого контура дополнительно просматриваются сетевые события при обработке файла, создании векторов, использовании инструмента и генерации артефакта. Проверка только обычного текстового ответа оставляет эти маршруты непроверенными.

Смежные статьи: движение данных, сетевая защита, файлы и RAG, модель угроз. Архитектурные механизмы не являются самостоятельным заключением о соответствии законодательству; применимость требований рассмотрена в мерах РФ.

БЕЗОПАСНОСТЬ

Потоки данных и маршруты обработки

Путь запроса в облаке и on-prem, разрешённые модельные маршруты, восстановление масок и отдельный изолированный профиль.

Облако / В контуре

https://docs.nevel.sh/security/data-flow/

Единицей проверки в Nevel является полный запрос к обработчику. Он может содержать сообщение пользователя, историю диалога, инструкции проекта, найденные фрагменты базы знаний, извлечённый текст файлов и результаты инструментов. Защита только последней фразы пользователя оставила бы остальной контекст за пределами контроля.

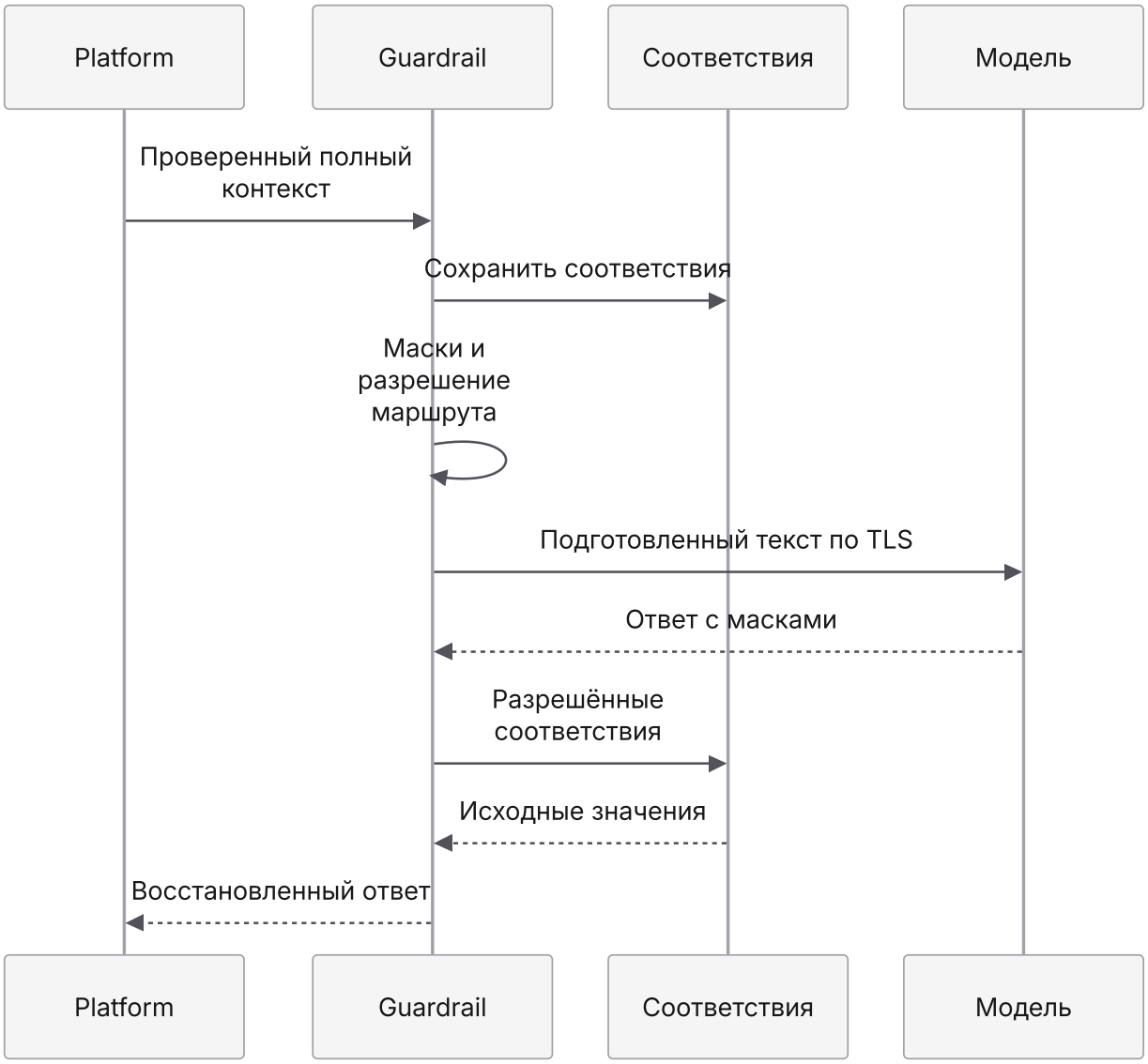
Поток описывается в двух направлениях: подготовка данных перед моделью и обработка результата после модели. Отдельно учитываются служебные ответвления — журналы, временные файлы, индексы, кэши и резервные копии. Они тоже могут содержать корпоративные сведения, хотя пользователь не видит их в чате.

Путь защищённого запроса

1. **Приём.** Gateway проверяет ключ или пользовательскую сессию, состояние доступа, рабочее пространство и ограничения запроса. Создаётся идентификатор для связывания дальнейших событий.
2. **Проверка объектов.** Platform проверяет права на историю, проект, файлы и подключения. Данные недоступного объекта не извлекаются для модели.
3. **Сборка контекста.** Из разрешённых источников формируется фактическое содержание запроса. Результат инструмента считается новым входом, а не доверенной инструкцией.
4. **Обнаружение и маскирование.** DLP/NER выделяет поддерживаемые чувствительные сущности; исходные значения заменяются типизированными идентификаторами. Соответствия сохраняются в защищённой области текущего запроса.
5. **Классификация и маршрут.** Содержание, метки, политика организации и ограничения модели определяют допустимый периметр. Запрет имеет приоритет перед предпочтением пользователя и автоматическим выбором модели.
6. **Обработка моделью.** Разрешённый получатель получает подготовленный контекст по защищённому каналу. При ошибке резервный маршрут обязан удовлетворять тем же ограничениям.
7. **Обратная обработка.** В ответе восстанавливаются только действительные соответствия разрешённого запроса. Незнакомая или просроченная маска не вызывает поиск по чужим данным.
8. **Возврат и аудит.** Результат получает уполномоченный клиент. Журнал связывает идентичность, политику, маршрут, маскирование, результат и ошибку без обязательной записи полного содержания.

СХЕМА 1

Маскирование и восстановление



Пример преобразования

Ниже используются вымышленные данные. Формат масок иллюстрирует принцип и не является обязательной схемой ответа API.

```
Исходный запрос:
Подготовь письмо Анне Примеровой по адресу anna@example.invalid
о переносе встречи. Телефон для связи: +7 000 000-00-00.

Содержание после обнаружения сущностей:
Подготовь письмо [PERSON_1] по адресу [EMAIL_ADDRESS_1]
о переносе встречи. Телефон для связи: [PHONE_NUMBER_1].

Ответ модели:
[PERSON_1], сообщаем о переносе встречи ...

После разрешённого восстановления:
Анна Примерова, сообщаем о переносе встречи ...
```

Модель получает смысл задачи и заменители, а таблица исходных значений остаётся внутри защищённой области. Но маскирование не скрывает факт переноса встречи, коммерческие условия или другие немаскированные детали. Для документа, содержание которого целиком запрещено передавать наружу, правильное решение — локальный маршрут либо отказ, даже если все известные идентификаторы обнаружены.

Что меняется в облаке

В облаке Platform и Guardrail размещены в российской инфраструктуре Nevel. Модель может быть self-hosted у Nevel в РФ, например DeepSeek, либо находиться у российского или глобального провайдера. При обращении к отдельному провайдеру подготовленный запрос пересекает границу инфраструктуры Nevel. Это описание сетевой границы; фактическая страна обработки определяется отдельно для всей цепочки, включая доставку запросов, поиск, преобразование файлов и инструменты.

Для запроса составляется цепочка: инициатор → Platform → Guardrail → каждый промежуточный обработчик → модель → обратная обработка. Применение одним участником режима без обучения не доказывает такой же режим у остальных. Отсутствие сохранения истории в интерфейсе также не подтверждает отсутствие технического хранения у получателя.

Данные соответствий не включаются в запрос к модели. Их содержимое не должно попадать в заголовки, диагностический ответ, трассировку, события SIEM или текст сообщения об ошибке. Результат восстановления возвращается только по исходному авторизованному маршруту.

Что меняется в on-prem

При on-prem Platform и её рабочие хранилища находятся у заказчика. После сборки контекста и проверки Guardrail запрос направляется внутренней модели или разрешённому получателю за пределами инфраструктуры заказчика: российскому провайдеру, self-hosted модели Nevel в РФ либо глобальному провайдеру. Место Platform не снимает требований к данным и получателям.

В реестре отдельно фиксируются место выполнения модели, владелец endpoint и география. Например, DeepSeek, размещённый Nevel в РФ, и внешний API с той же моделью имеют разные цепочки обработки. Ответ из внешнего маршрута возвращается в установленную у заказчика платформу для обратной обработки и выдачи уполномоченному клиенту.

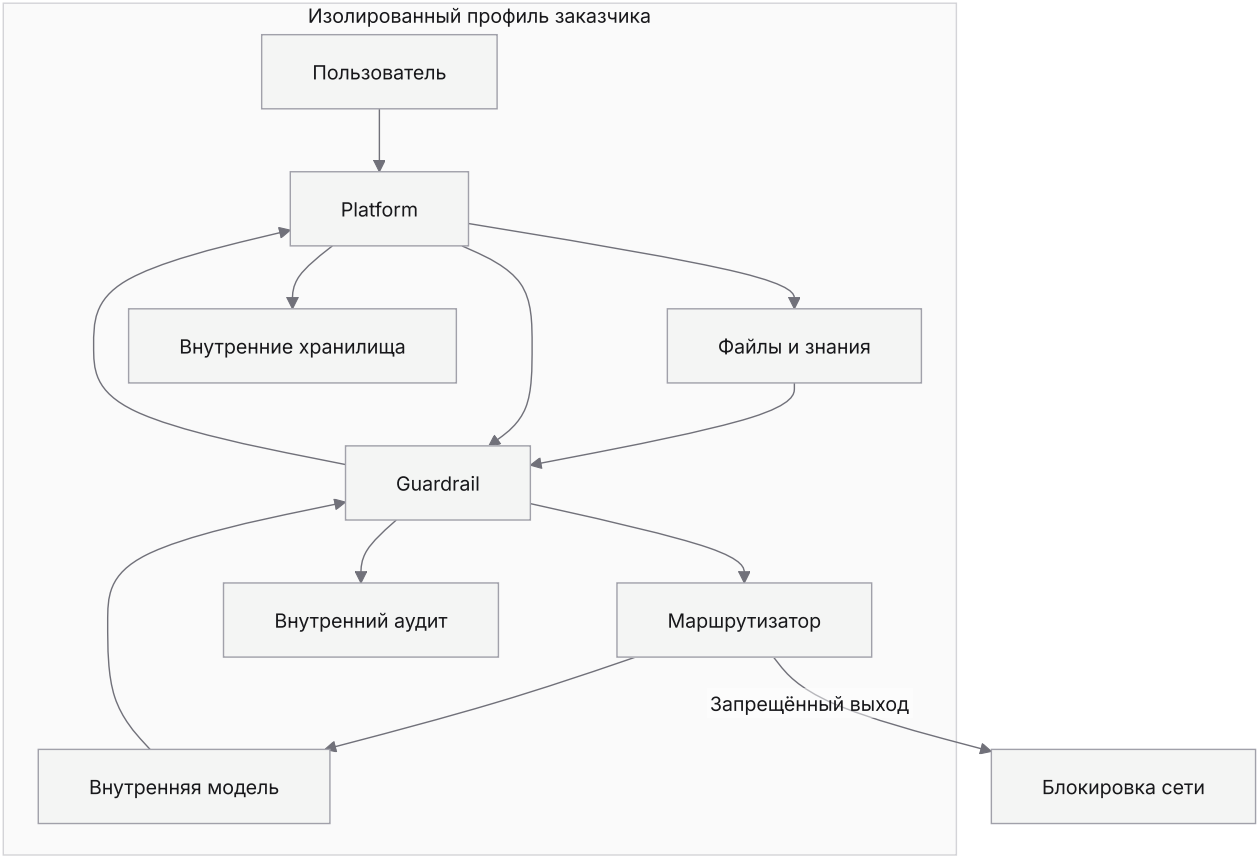
Резервный маршрут выбирается только из вариантов, заранее разрешённых для категории данных. Его использование отражается в аудите. При отсутствии допустимой альтернативы запрос завершается отказом; SDK и фоновые задачи не должны самостоятельно включать скрытый внешний fallback. Матрица вариантов связывает место платформы и модельные маршруты.

Изолированный профиль

Внутри контура находится вся рабочая цепочка. Используются внутренние модели, внутренний обработчик файлов, внутреннее векторное хранилище и внутренние инструменты. Даже если запрос выглядит безобидным, политика закрытого контура не разрешает автоматический внешний fallback.

СХЕМА 2

Изолированный рабочий маршрут



Проверка отсутствия внешней обработки проводится по всем сценариям: OCR изображения, разбор сложного PDF, построение embeddings, вызов инструмента, создание файла, повторная попытка и диагностика сбоя. On-prem с разрешённым облачным векторизатором остаётся локальной установкой платформы, но не соответствует изолированному профилю.

Файлы, история и повторные запросы

Файл может породить несколько объектов: оригинал, извлечённый текст, изображения страниц, фрагменты, векторы, предварительный просмотр и итоговый артефакт. Каждый наследует контекст организации и исходные ограничения доступа. При изменении метки исходного документа производные представления учитывают новое ограничение до дальнейшей выдачи.

История диалога повторно передаёт модели ранее написанные сообщения. Поэтому правило защиты применяется к собранной истории, а не только к новым данным. Старое решение «разрешено» не должно бессрочно действовать после смены политики или модели. Если срок хранения соответствия маски истёк, это не повод возвращаться к исходным значениям из чужого запроса или отправлять их во внешнюю модель без повторного контроля.

Вызов инструмента продолжает тот же процесс. Модель предлагает параметры, сервер проверяет разрешённость действия, инструмент возвращает результат, после чего этот результат проходит контроль перед следующим модельным вызовом. Полномочия инструмента не определяются текстом модели. Для фоновых и запланированных запусков учитываются текущий доступ владельца и актуальность секрета подключения.

Потоковый ответ и частичный результат

Потоковая передача требует аккуратной обработки границ фрагментов: маска может разделиться между несколькими порциями текста. Восстановление выполняется по целой распознанной маске в пределах запроса. Неполный фрагмент не используется для приблизительного подбора значения. Если соединение прерывается, журнал отличает частичный ответ от успешного завершения.

Повторный запрос может привести к новому вызову модели и новой стоимости. Поэтому клиент связывает повторы с исходной операцией, а расследование учитывает количество фактических отправок. Отмена на стороне браузера не означает, что уже отправленный внешний запрос был отменён у получателя или что все временные объекты уничтожены мгновенно.

Контрольная матрица

Проверка	Ожидаемый результат	Доказательство
Чувствительное значение только в старом сообщении	Оно также проходит контроль	Синтетический захват перед моделью и событие маскирования
Запрещённый документ в результате инструмента	Следующий вызов не уходит по запрещённому маршруту	Решение политики и отсутствие исходящего вызова
Поддельная маска из другого запроса	Чужое значение не восстанавливается	Ответ, событие отказа, изоляция соответствий
Локальная категория и отказ внутренней модели	Внешний fallback не происходит	Сетевой журнал и событие маршрутизатора
Отказ DLP или хранилища соответствий	Защищённый маршрут не отправляет исходный текст	Ошибка, идентификатор, отсутствие вызова модели
Запрос с удалённым файлом	Файл и его фрагменты больше не попадают в контекст	Проверка доступа и поиск по индексу

Подробности: Guardrail, хранение, файлы и RAG. Юридическая оценка внешней передачи изложена в 152-ФЗ; техническая замена идентификаторов не подменяет её.

БЕЗОПАСНОСТЬ

Guardrail: маскирование и маршрутизация

15 типов сущностей, режимы распознавания, классификация документов, восстановление масок и поведение при ошибках.

Облако / В контуре

https://docs.nevel.sh/security/guardrail/

Guardrail контролирует подготовленный запрос до передачи модели. Он объединяет обнаружение чувствительных сущностей, замену значений, классификацию документа и выбор допустимого маршрута. Эти операции решают разные задачи: замена имени сокращает раскрытие идентификатора, а классификация финансового отчёта может запретить передачу всего отчёта независимо от найденных имён.

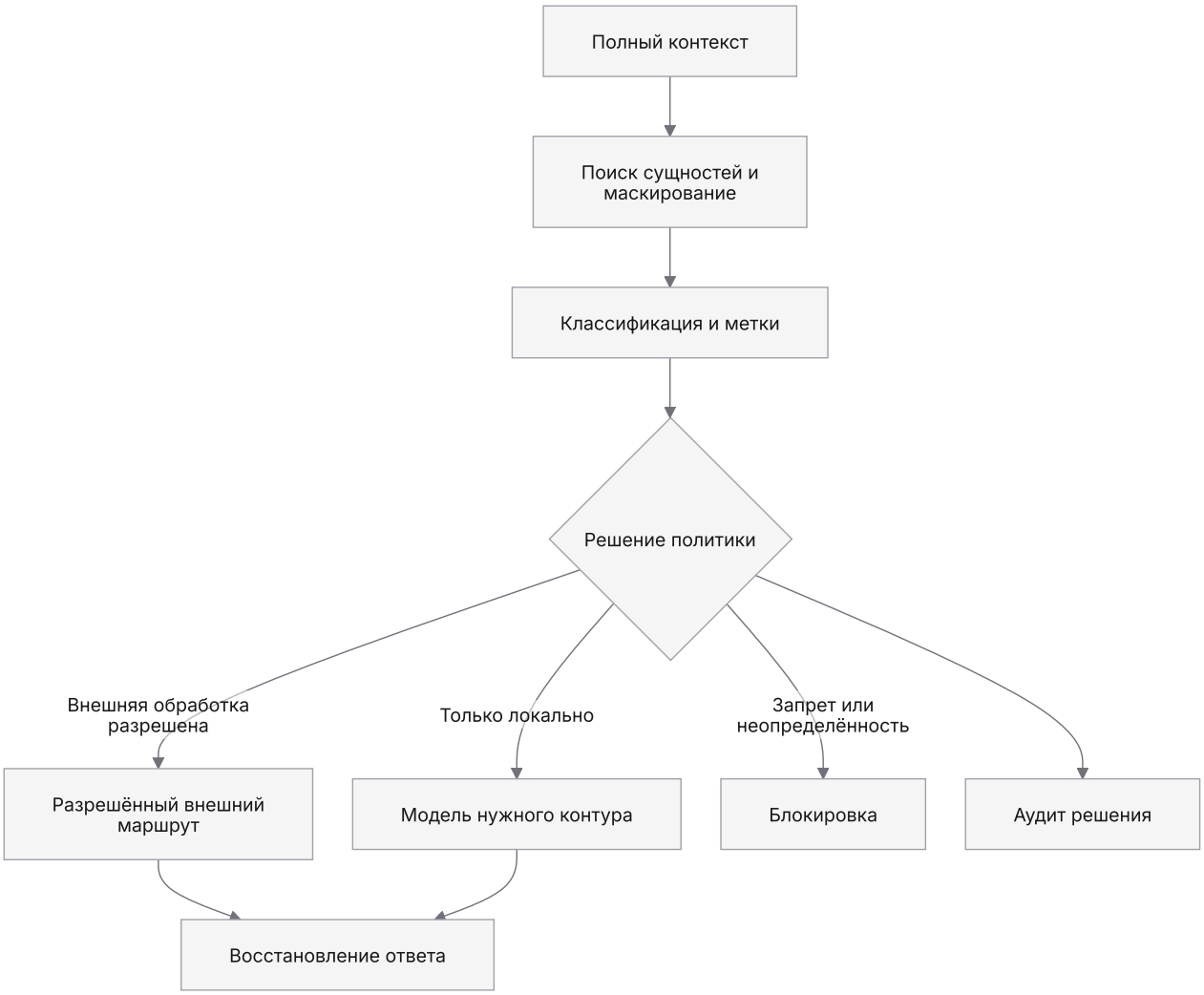
Защищённый API-маршрут использует префикс `/secure/v1`. Наличие обычного совместимого API-маршрута не означает, что к нему автоматически применяется тот же набор DLP-проверок. Для обязательной защиты администратор закрепляет защищённый маршрут в клиентах и закрывает обходные пути сетевыми и серверными правилами. Доступность конкретных методов, параметров и потокового режима проверяется по документации API, а не по одному сходству URL.

Порядок обработки

Gateway проверяет доступ и ограничения запроса. DLP анализирует текст, заменяет обнаруженные сущности и сохраняет соответствия. Классификатор оценивает содержание целиком и обязательные метки, после чего маршрутизатор выбирает разрешённый периметр и модель. После ответа выполняется восстановление допустимых масок и запись события обработки. История, извлечённый текст и результаты инструментов включаются в эту последовательность перед соответствующим модельным вызовом.

СХЕМА 1

Решение Guardrail



Как обнаруживаются сведения

Обнаружение сочетает правила для структурированных последовательностей и модели распознавания именованных сущностей, адаптированные к русскоязычным данным. Правила полезны для форматов документов, счетов и контактов; контекстное распознавание — для имён, адресов и неоднозначных фрагментов. Более строгая обработка использует дополнительный анализ сложного контекста.

В базовом текстовом обработчике используются Microsoft Presidio, NLP-обработка spaCy и собственные распознаватели русскоязычных форматов, включая телефоны, электронную почту и имена с инициалами. Распознаватель сочетает шаблон, контекстные слова и оценку уверенности. Этот состав объясняет принцип обработки; полный профиль Guardrail определяется также расширенным набором сущностей, классификацией и маршрутизацией. Наличие одной библиотеки само по себе не подтверждает весь профиль защиты.

Ни одна из этих стадий не является доказательством отсутствия всех персональных данных. Текст с ошибками OCR, транслитерацией, вставленными пробелами, нетипичным сокращением или косвенным описанием человека может потребовать другого правила или локального маршрута. Название конкретной библиотеки не определяет качество защиты: проверяется настроенный набор распознавателей на реальных форматах организации и синтетических примерах.

Тип сущности	Что распознаётся	Существенное ограничение проверки
PERSON	Имена, фамилии, инициалы	Совпадение с названием организации или термином может давать ложное срабатывание
PHONE_NUMBER	Телефонные номера, в том числе российские формы записи	Добавочные номера, разделители и ошибки распознавания требуют примеров
EMAIL_ADDRESS	Адреса электронной почты, включая кириллические варианты	Косвенная запись словами и разорванный адрес сложнее обычного формата
CREDIT_CARD	Номера платёжных карт	Номер в изображении становится текстом только после распознавания
IP_ADDRESS	IP-адреса	Адрес может быть техническими данными, а его чувствительность зависит от контекста
INN	ИНН физических и юридических лиц	Сам по себе ИНН организации не равен персональным данным физического лица
SNILS	СНИЛС	Пробелы, дефисы и ошибки OCR проверяются отдельно
PASSPORT	Данные внутреннего паспорта	Контекст серии и номера важен для отличия от других последовательностей
ADDRESS	Почтовые и фактические адреса	Неполные адреса и свободная форма повышают неопределённость
DATE_OF_BIRTH	Даты рождения	Обычная дата события не должна автоматически трактоваться как дата рождения
BANK_ACCOUNT	Банковские счета	Счёт, БИК и другие реквизиты различаются по назначению
DRIVER_LICENSE	Водительские удостоверения	Проверяются используемые варианты записи
FOREIGN_PASSPORT	Заграничные паспорта	Форматы документа и контекст отличаются от внутреннего паспорта
VEHICLE_PLATE	Регистрационные номера автомобилей	Нужны проверки регионов, похожих символов и OCR
MEDICAL_POLICY	Номера полисов ОМС и ДМС	Вид номера не раскрывает автоматически всю медицинскую информацию в документе

Перечень показывает покрываемые классы, а не исчерпывающее юридическое определение персональных данных. Сведения о здоровье, зарплате, поведении или трудовой оценке могут оставаться чувствительными без любого из этих идентификаторов. Специальные категории оцениваются по содержанию и цели обработки, см. 152-ФЗ.

Режимы распознавания

Режим	Поведение	Подходящий контрольный сценарий
Balanced	Баланс полноты обнаружения и числа лишних замен; базовый режим	Обычные деловые тексты с контактами, договорами и перепиской
Strict	Более агрессивная обработка неоднозначных фрагментов	Повышенная чувствительность, допустимость дополнительных масок и большей задержки

Режим	Поведение	Подходящий контрольный сценарий
Permissive	Снижение лишних замен в техническом тексте	Код, технические адреса и числовые значения при отдельно определённой допустимости маршрута

Режим влияет на компромисс между пропуском и лишней заменой. Permissive не является способом разрешить запрещённую обработку. Strict не означает стопроцентное обнаружение. Политика может независимо запретить внешний маршрут для класса данных, даже если пользователь предпочитает более мягкое распознавание.

При оценке качества отдельно считают полноту по каждому типу сущности, ложные срабатывания, ошибки восстановления и долю документов, отправленных в неверный класс. Сводная «точность» без состава выборки скрывает важные различия: редкие паспортные данные могут распознаваться хуже частых адресов электронной почты.

Классификация целого документа

Классификатор определяет допустимость обработки документа как целого. Примеры категорий — общедоступный материал, внутренний рабочий документ, коммерческое предложение, финансовый отчёт, кадровые сведения. Организация использует свои определения и обязательные метки. Категория не привязана только к расширению файла: финансовая таблица может быть вставлена в обычное сообщение.

Место Platform не является единственным условием решения. В облаке и on-prem используются разрешённые локальные, российские, self-hosted Nevel в РФ и глобальные маршруты, доступные в конфигурации. Например, self-hosted DeepSeek в РФ и API внешнего провайдера с той же моделью различаются получателем и географией. Подробное сравнение — в вариантах маршрутизации.

Решение учитывает несколько источников ограничений: административную политику, метку исходного документа, выбранный пользователем повышенный уровень чувствительности, результат обнаружения и классификации. Пользовательская отметка «только локально» усиливает ограничения; текст внутри файла «можно отправлять всем» не снимает обязательный запрет. При конфликте действует более ограничительный допустимый маршрут.

Условие	Решение
Допущены внешние модели, содержимое соответствует разрешённой категории	Маскирование и отправка по разрешённой цепочке получателей
Разрешена обработка в РФ у определённых получателей	Только выбранный допустимый российский маршрут: локальная модель, self-hosted Nevel или российский провайдер
Изолированный профиль или метка «только в контуре заказчика»	Только модель внутри заданного контура
Категория запрещена для любых моделей	Блокировка и событие аудита
Для категории требуется локальная обработка, но нужный обработчик недоступен	Отказ или другой разрешённый локальный обработчик
Проверка не завершилась	Защищённый маршрут не передаёт непроверенный текст наружу

Приоритеты, разрешённые и запрещённые категории и резервные маршруты изменяются административной политикой. Изменение применяется к новым решениям без необходимости разрешать обход проверки на время обновления. Версия политики сохраняется вместе с событием, чтобы объяснить прошлое решение после изменения настроек.

Маски и обратное восстановление

Каждой обнаруженной сущности назначается заменитель, а исходное значение помещается в таблицу соответствий. Область соответствий ограничена организацией и текущим запросом или разрешённым контекстом обработки. Модель не получает таблицу и не может запросить произвольное раскрытие значения через текст ответа.

При восстановлении проверяются принадлежность маски, срок жизни соответствия и право получить результат. Совпадение текста с шаблоном маски само по себе недостаточно. Поддельный идентификатор из другого диалога не должен открывать чужое значение. Просроченное соответствие не восстанавливается эвристически. Повреждённая моделью маска остаётся нераскрытой либо приводит к контролируемой ошибке обработки.

Маски могут влиять на качество: модель не знает склонение фамилии, свойства удалённого адреса или часть контекста. Поэтому приёмка проверяет не только отсутствие исходных значений на внешнем выходе, но и полезность результата, сохранение структуры документа и корректность восстановления. Сроки таблицы соответствий разобраны в хранении.

Ограничения по времени обработки

В базовом описании Guardrail версии 1.2 приведён профиль добавочной задержки шлюза: p50 до 150 мс и p99 до 500 мс без времени ответа LLM. В обсуждении практического сценария Balanced также использовался ориентир около 500 мс дополнительной обработки. Это разные контексты оценки, которые нельзя превратить в одно безусловное обещание для любого текста, файла и нагрузки.

Для испытания отдельно измеряются приём, DLP, классификация, ожидание модели и восстановление. Фиксируются размер входа, наличие OCR, режим, параллелизм и оборудование. Процентили рассчитываются по достаточной серии запросов, а не по одному успешному ответу. Обязательные показатели конкретной поставки определяются её SLA и программой испытаний.

Негативные проверки

Проверочный корпус содержит все 15 классов, несколько сущностей одного типа, кириллицу, транслитерацию, пробелы внутри номера, ошибку OCR, ложные совпадения в коде и чувствительные сведения только в истории. Отдельный набор содержит документ без явных идентификаторов, который всё равно запрещён к внешней передаче.

Далее проверяются: поддельная и разделённая потоком маска; повтор чужого идентификатора; истёкшее соответствие; отказ DLP; отказ локальной модели; попытка вызвать обычный API вместо защищённого; смена модели автоматическим режимом; запрещённый результат инструмента. Успех означает отсутствие обхода и понятный след решения, а не только красивый ответ в интерфейсе.

Смежные меры: потоки, AI firewall и сеть, аудит. Обезличивание и правовой статус передачи рассматриваются отдельно в 152-ФЗ.

БЕЗОПАСНОСТЬ

Шифрование и управление ключами

TLS, защита хранилищ и резервных копий, секреты подключений, ротация и границы доступа к открытому тексту.

Облако / В контуре

https://docs.nevel.sh/security/encryption/

Шифрование в Nevel защищает данные при передаче и хранении. Оно применяется вместе с разграничением доступа, контролем маршрута и ограничением сроков. Для оценки защиты нужно различать три объекта: ключи шифрования данных, ключи доступа к API и соответствия масок. Они решают разные задачи и не заменяют друг друга.

API-ключ даёт право выполнить запрос. Ключ шифрования позволяет расшифровать защищённое представление. Таблица соответствий связывает заменитель с исходным значением; сама по себе такая замена не является криптографическим шифрованием. Утечка любого из этих объектов имеет собственные последствия и требует отдельного отзыва или удаления.

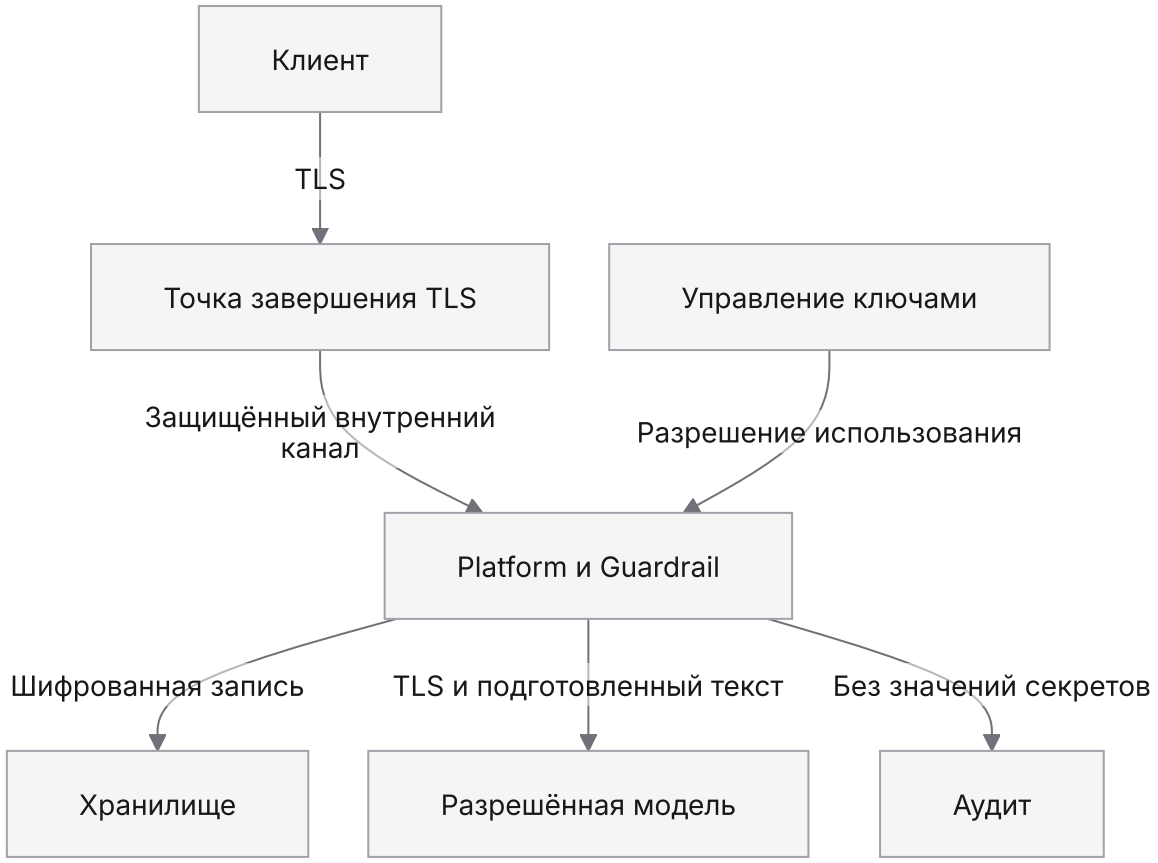
Защита каналов

Базовое требование транспортной защиты — TLS 1.2 или выше. Оно относится к пользовательскому входу, API, соединениям между компонентами, хранилищам и разрешённым внешним получателям. При проверке оцениваются не только версия протокола, но и доверие к сертификату, соответствие имени узла, срок действия и отсутствие автоматического понижения до незащищённого канала.

TLS завершается на конкретном компоненте: балансировщике, reverse proxy, Gateway или средстве защиты. В этой точке запрос становится доступен в открытом виде для разрешённой обработки. Если после завершения TLS есть следующий сетевой участок, он проверяется отдельно. Зелёный значок в браузере подтверждает только участок до выбранной точки завершения, а не всю цепочку до модели.

СХЕМА 1

Где данные расшифровываются



В on-prem внутренние сервисы используют доверенные сертификаты инфраструктуры заказчика, а соединения с разрешёнными модельными провайдерами отдельно проверяют сертификат и имя внешнего endpoint. Подмена доверия ради удобства — например, отключение проверки сертификата у модельного клиента — создаёт обход транспортной защиты. При подключении собственного AI firewall его точка расшифрования включается в схему доступа к открытому тексту; см. сетевую архитектуру.

Защита хранилищ

Шифрованное хранение охватывает служебные данные и секреты в пределах выбранной поставки. Для резервных копий в базовом профиле Guardrail версии 1.2 указан AES-256. Это точное указание относится к профилю резервирования; его нельзя автоматически переносить на алгоритм каждой базы, файла или поля приложения без паспорта установленной конфигурации.

У шифрования несколько уровней. Шифрование диска защищает носитель вне работающей системы. Шифрование объекта защищает сохранённое представление в файловом хранилище. Шифрование отдельных секретов отделяет их от обычных записей приложения. Эти уровни могут сочетаться, но доступ приложения к рабочему ключу позволяет ему обрабатывать данные в открытом виде.

Данные	Что проверяется	Какой риск остаётся
История и файлы	Шифрование тома или хранилища, сервисные права, защита экспорта	Авторизованный процесс приложения может читать разрешённое содержимое
Соответствия масок	Отдельная область, шифрованное хранение, ограниченный доступ	Компрометация процесса восстановления может раскрыть доступные ему значения
Токены интеграций	Шифрованное хранение и выдача только нужному исполнителю	Избыточные права самого токена позволяют действие во внешней системе

Данные	Что проверяется	Какой риск остаётся
Резервные копии	Защита копии, отдельное хранение ключа, проверка восстановления	Доступ к копии вместе с рабочим ключом позволяет восстановить сведения
Журналы	Минимизация содержания, доступ и защита экспорта	Даже без тел запросов метаданные могут раскрывать активность людей
Временные файлы	Изоляция рабочего каталога и очистка	Аварийный процесс или дамп памяти может сохранить фрагмент данных

Шифрование не исправляет открытый бакет, чрезмерные полномочия или публикацию расшифрованного экспорта. Поэтому проверка зашифрованного носителя дополняется попыткой неавторизованного чтения через прикладной API.

Реестр ключей и секретов

У каждого секрета фиксируются назначение, владелец, среда, область действия, потребители и порядок отзыва. Разделяются производственная среда, тестирование и разработка. Ключ модели не передаётся в браузер; пользователь обращается к Platform или Gateway под своей идентичностью. Секрет подключения доступен исполнителю только для разрешённой операции.

Объект	Назначение	Действие при компрометации
Сертификат и закрытый ключ TLS	Подлинность узла и защищённый канал	Замена ключевой пары и сертификата, проверка доверия клиентов
Ключ шифрования хранилища	Защита сохранённых данных	Ограничение доступа, ротация или повторное шифрование по механизму хранилища
Секрет подписи сессий	Проверка подлинности пользовательской сессии	Замена и прекращение доверия к затронутым сессиям
Ключ доступа к Nevel API	Доступ приложения к разрешённым моделям и операциям	Немедленный отзыв ключа, новый ключ с минимальными правами
Ключ провайдера модели	Оплата и доступ к модельному обработчику	Отзыв у провайдера, замена в защищённой конфигурации, проверка расходов
OAuth-токен подключения	Действия в подключённой системе	Отзыв подключения и токена у источника, повторная авторизация
Доступ к хранилищу соответствий	Восстановление масок	Отзыв сервисного доступа и оценка раскрытых областей запросов

Отзыв ключа Nevel не отзывает автоматически уже скомпрометированный ключ внешнего провайдера. Удаление подключения из интерфейса также проверяется на стороне источника: отдельный внешний токен может иметь свой срок действия и механизм отзыва.

Ротация без потери управляемости

Ротация начинается с выпуска нового секрета в той же или более узкой области. Разрешённые потребители переводятся на новую версию, выполняется контрольный запрос, затем старый секрет отзывается. Переходный период ограничен и отражён в журнале. Если источник не позволяет безопасное одновременное действие двух ключей, переключение выполняется как управляемая операция с проверкой доступности.

Ключи шифрования требуют отдельного внимания к старым данным и резервным копиям. Простая замена значения переменной окружения не гарантирует доступность ранее зашифрованного содержимого. Механизм хранения должен связывать данные с версией ключа или выполнять предусмотренное повторное шифрование. Проверяется чтение текущих данных и восстановление контрольной копии после ротации.

Удаление последнего ключа может сделать данные невосстановимыми, но такой результат не объявляется универсальным способом исполнения любого запроса на уничтожение. Нужно знать, какие копии использовали этот ключ, существовали ли расшифрованные экспорты и сохранились ли другие версии. Порядок удаления описан в жизненном цикле данных.

Облако и ключи заказчика

В облаке эксплуатационная команда управляет инфраструктурными ключами в пределах поставки. Корпоративный API-ключ, персональный токен подключения и ключ провайдера с оплатой заказчика сохраняют отдельного владельца и область ответственности. Использование собственного ключа модели меняет учёт и полномочия обращения к провайдеру, но не переносит автоматически обработку внутрь компании.

В on-prem источники секретов установленной платформы находятся под управлением заказчика в рамках выбранной конфигурации. Это включает ключи разрешённых внешних моделей: российский провайдер, self-hosted Nevel в РФ или глобальная цепочка API не становятся внутренними сервисами из-за локального хранения их ключа. В изолированном профиле секреты и все рабочие потребители остаются внутри контура. Заказчик контролирует выдачу, ротацию, резервирование и аварийное восстановление своей части системы. Необходимость HSM, сертифицированного СКЗИ или конкретного корпоративного KMS определяется требованиями к установленной системе и интеграцией; совместимость проверяется для выбранных средства, версии и конфигурации.

Диагностика и защита от случайного раскрытия

В журнал попадает идентификатор ключа или его безопасный отпечаток, но не значение. Заголовки авторизации, cookie, строки подключения и исходные соответствия масок исключаются из сообщений об ошибках, трассировок и экспортов поддержки. Скриншоты настроек и диагностические архивы рассматриваются как отдельные передаваемые объекты.

Воспроизводимая проверка использует специальный тестовый секрет и поиск его точного значения в логах, ответах, сохранённых артефактах и диагностических выгрузках. Настоящие производственные ключи для такого теста не нужны. Дамп памяти работающего процесса способен содержать расшифрованные данные, поэтому доступ к дампам и их срок хранения ограничиваются так же, как доступ к самим данным.

Приёмка

Проверьте отказ при недоверенном сертификате, невозможность незашифрованного соединения и отсутствие понижения протокола. Затем отзовите тестовый API-ключ и токен интеграции, убедитесь в прекращении новых операций и сохранении события. После ротации ключа хранения проверьте чтение данных и восстановление копии; после отказа источника секретов защищённый маршрут не должен продолжать работу с неразрешённым запасным секретом.

Для регуляторной оценки важно различать применение современных алгоритмов и применение конкретных средств, прошедших требуемую оценку соответствия. TLS и AES сами по себе не являются сертификатом СКЗИ или заключением о выполнении требований ФСТЭК. Применимость таких требований рассмотрена в мерах защиты РФ.

БЕЗОПАСНОСТЬ

Идентификация, SSO и права

Корпоративный вход, права дирекций на источники и функции, массовые назначения, проверка полномочий и отзыв доступа.

Облако / В контуре

https://docs.nevel.sh/security/identity/

Nevel различает идентификацию пользователя, аутентификацию и разрешение конкретного действия. Успешный вход подтверждает, кто обращается к системе. Доступ к проекту, файлу, ассистенту, модели или подключению определяется отдельно. Корпоративный SSO не делает сотрудника администратором и не открывает ему все документы компании.

Внешнее приложение работает под сервисным API-ключом, а сотрудник — под личной учётной записью. Общая учётная запись отдела лишает журнал возможности установить инициатора и усложняет отзыв доступа при увольнении. Поэтому персональные пользователи, сервисные интеграции и административные операции разделяются.

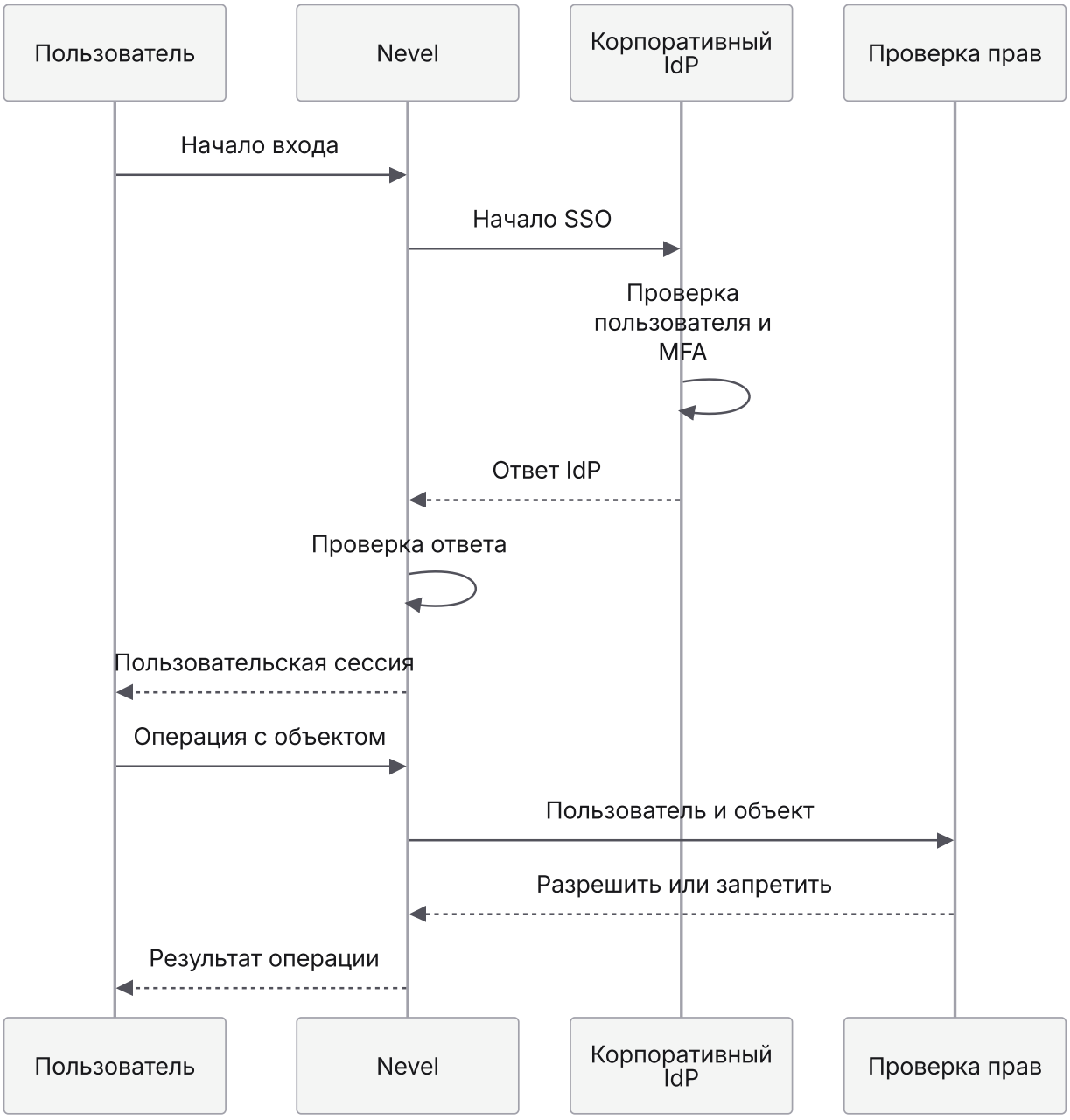
Корпоративная аутентификация

Для корпоративной интеграции используются предусмотренные поставкой механизмы OAuth 2.0 и SAML. OAuth 2.0 описывает делегирование доступа и сам по себе не является универсальным доказательством пользовательской идентичности. Конкретный сценарий входа должен определять доверенный источник идентичности, атрибут стабильного пользователя и проверку полученного ответа. SAML-сценарий использует доверие к корпоративному поставщику идентичности и проверку утверждения о пользователе.

При SSO учитываются издатель, адрес назначения, получатель, срок действия, связь с начатым входом и защита от повторного использования ответа. Для SAML проверяются подпись и доверенный сертификат поставщика идентичности. Недействительный, просроченный или предназначенный другой системе ответ не создаёт сессию. Совпадение адреса электронной почты само по себе не позволяет связать неподтверждённую внешнюю учётную запись с привилегированным пользователем.

СХЕМА 1

Вход через корпоративный источник



Требование MFA может исполняться корпоративным IdP в процессе входа. Приёмка проверяет, что приложение использует именно этот обязательный маршрут и что альтернативная учётная запись не даёт обойти корпоративное требование. Применение конкретного продукта IdP, автоматического создания пользователей или отдельного протокола синхронизации определяется конфигурацией интеграции; наличие SSO не означает автоматически поддержку любого каталога и метода provisioning.

Роли и административные полномочия

В модели рабочего пространства различаются OWNER, ADMIN и MEMBER: владелец, администратор и участник. Владение определяет максимальный уровень управления пространством; администратор выполняет разрешённые операции настройки; участник использует доступные ему функции. Практическое право определяется матрицей разрешений, состоянием пользователя и областью операции.

Область	Обычный участник	Административная операция
Собственные диалоги	Создание и использование в пределах политики	Не требует права читать чужие диалоги
Общие проекты и файлы	Только предоставленный уровень доступа	Выдача и отзыв совместного доступа
Пользователи	Работа под своей идентичностью	Приглашение, изменение роли, прекращение доступа
Модели и инструменты	Использование разрешённых вариантов	Настройка доступных функций и маршрутов обработки
Политики Guardrail	Исполнение установленных правил	Изменение категорий, маршрутов и режима защиты
Подключения	Личное подключение или явно разрешённое общее	Разрешение интеграции и управление областью использования
Аудит	Не равен общему доступу к журналу	Просмотр и экспорт в отдельной области полномочий

Управление пользователями не должно автоматически означать право на содержимое всех файлов. Разграничение проверяется по конкретной операции и объекту, а не только по видимости кнопки в интерфейсе. Сервер повторяет проверку даже при прямом API-запросе с известным идентификатором.

Пространства, проекты, ассистенты и группы

Рабочее пространство задаёт организационный контекст. Внутри него проекты, файлы и ассистенты могут иметь собственные правила совместного использования. Личный ассистент доступен владельцу; общий предоставляется определённым участникам; доступность внутри организации ограничена членством и политикой. Ссылка является способом найти ресурс, а право определяется правилами публикации и авторизацией.

При передаче ассистента проверяются отдельно его инструкции, исходные документы и подключения. Нельзя считать, что публикация конфигурации ассистента разрешает использование личного токена автора всеми сотрудниками. Аналогично совместный проект не обязан открывать любой диалог автора, если общий доступ предоставлен только файлам или инструкциям.

Группы и организационные признаки связывают сотрудников с дирекциями и административной матрицей. Они не образуют автоматически строгую иерархию подразделений. В таблице участников основная группа выбирается отдельно от роли команды; проектные назначения позволяют одному человеку работать с несколькими направлениями. Название должности или группы само по себе не выдаёт права владельца, доступ к источникам или разрешение инструмента.

Для совместной работы нужны одновременно подходящий состав участников, права на объекты и разрешённые действия. Например, сотрудник финансов может участвовать в общем проекте закупок, не получая доступа ко всем материалам закупочной дирекции. Общий ассистент не должен извлекать для него документы, недоступные в его текущем контексте.

Как применяется матрица дирекций

Матрица фиксирует связь **сотрудник → дирекция и проекты → ассистенты и источники → разрешённые действия**. Для каждого набора назначений определяются ответственный за состав людей, владелец материалов и администратор политики. Эти роли могут принадлежать разным сотрудникам: руководитель дирекции подтверждает рабочую потребность, а право изменить настройки принадлежит уполномоченному администратору.

При проверке действия учитываются несколько независимых условий:

1. **Идентичность и членство.** Пользователь действует под правильным аккаунтом в нужной организации; приглашение ещё не подменяет принятого членства.
2. **Право на объект.** Доступны конкретный проект, диалог, ассистент, файл или подключение, участвующие в операции.
3. **Разрешённое действие.** Пользователю разрешено читать, изменять, публиковать, запускать инструмент либо выполнять административную операцию — в зависимости от запроса.
4. **Политика обработки.** Для этих данных разрешены модель, получатель и вспомогательные обработчики. Доступ к документу не означает разрешения передать его любому провайдеру.

Разрешение одной части не компенсирует отсутствие другой. Например, сотрудник может читать финансовый документ, но не публиковать его как общий источник. Или ему доступен веб-поиск, но конкретные сведения нельзя включать во внешний поисковый запрос. Доступность кнопки подтверждает лишь видимость элемента; сервер проверяет фактическую операцию.

Когда у сотрудника несколько проектных назначений, каждое основание предоставления права должно оставаться понятным. Снятие доступа одного проекта не отменяет другое действующее назначение автоматически. Поэтому при отзыве проверяется итоговое право, включая прямую выдачу доступа и другие проекты. Обязательный запрет для категории данных или маршрута сохраняется независимо от количества разрешающих назначений.

Проверяемый сценарий	Ожидаемый результат
Сотрудник закупок обращается к своему ассистенту и регламенту	Получает результат в пределах назначенного доступа
Тот же сотрудник открывает закрытый источник финансов	Содержимое не предоставляется без отдельного назначения
Финансист участвует в общем проекте	Видит материалы общего проекта, а не все материалы второй дирекции
Участник пытается изменить роль коллеги	Операция требует отдельного административного права
Сотрудник открывает общедоступного ему ассистента с недоступным источником	Публикация ассистента не расширяет права на источник

Матрица не должна существовать только как описание отдела. Её проверяют через реальные объекты и действия: список участников, публикацию ассистента, доступ к документу и запуск разрешённого инструмента. Пример матрицы и порядок назначения находятся в администрировании команды.

Контроль массовых назначений и импорта

Массовые приглашения и импорт оргструктуры решают разные задачи. Приглашение связывает адрес, организацию и роль будущего участника. Импорт дополнительно сопоставляет существующие аккаунты, дирекции, проектные назначения и их изменения. Успешная отправка приглашений не подтверждает, что всем сотрудникам уже выдан правильный доступ к материалам.

Контроль начинается с идентичности. Источник состава должен содержать устойчивую связь с человеком; имя используется для отображения, но не для объединения аккаунтов. Смена адреса, одинаковые имена и несколько строк одного сотрудника требуют однозначного сопоставления. Конфликтующие строки не должны создавать привилегированный аккаунт по предположению. Неподтверждённая связь между аккаунтом и сотрудником остаётся основанием не применять конкретное назначение до исправления данных.

До применения администратор видит разницу между существующим и требуемым состоянием. В проверку входят новые участники, смена дирекции, новые доступы, административные роли и отзываемые назначения. Для каждого существенного изменения должно быть понятно, какой сотрудник затронут, какой объект или функция меняются и откуда взято требуемое состояние.

Добавление и обновление, merge, сохраняет пользователей и назначения вне обрабатываемого набора. Отсутствие строки в выгрузке не является указанием на отзыв. **Замена состава выбранной области** предполагает, что её источник полный; перечень снимаемых назначений просматривается отдельно. Исключение из проекта, снятие группы и прекращение членства в команде учитываются как разные изменения.

В журнале массовой операции сохраняются инициатор, время, область, источник или версия набора, просмотренные изменения и результат применения. Для разбора важны выполненные, не изменившиеся и ошибочные строки. Сам исходный кадровый файл содержит сведения о сотрудниках и должен храниться в предназначенной для этого административной области, а не становиться общей базой знаний ассистента.

При частичном выполнении итоговые права проверяются по фактическому состоянию. Повтор исправленных строк не должен создавать дубликаты аккаунтов или приглашений. Если результат неизвестен из-за обрыва связи, сначала сопоставляются участники и назначения, затем выполняются оставшиеся изменения. Отсутствие ответа клиенту не доказывает отмену серверной операции.

Для исправления ошибочного импорта используется перечень обратных изменений: снять лишние права, восстановить необходимое назначение и проверить новый запрос. Возврат состава не означает удаления созданных сотрудниками материалов или всего аккаунта. Подробные этапы подготовки, предпросмотра и восстановления приведены в инструкции импорта дирекций.

Сервисные ключи и собственный ключ модели

Ключ Nevel API связан с владельцем, рабочим пространством, статусом, сроком действия и ограничениями использования. Приложению выдаётся отдельный ключ по назначению; производственная интеграция и тестирование не используют один общий секрет. Ограничения расходов и модели уменьшают масштаб возможной компрометации, но не заменяют запрет доступа к данным.

Собственный ключ провайдера модели хранится и используется отдельно от ключа доступа к Nevel. Первый определяет право обратиться к провайдеру, второй — право приложения воспользоваться маршрутом Nevel. Использование ключа заказчика не отменяет Guardrail, допустимый перечень моделей или правила обработки информации. География провайдера остаётся свойством маршрута, а не владельца ключа.

Ротация и отзыв описаны в управлении ключами. Секрет не публикуется в ассистенте, системной инструкции, общем файле или примере кода. Модель получает возможность вызвать разрешённый инструмент через сервер, а не значение секрета для самостоятельного использования.

Отзыв и жизненный цикл доступа

При уходе сотрудника прекращаются членство, активные разрешения и использование его личных подключений. Отдельно рассматриваются API-ключи, фоновые задания, общие ассистенты и документы, владельцем которых он был. Передача владения не означает перенос личных секретов другому человеку. Объекты, нужные компании, передаются управляемо с сохранением происхождения и прав.

Для перевода между дирекциями сначала составьте перечень сохраняемых и снимаемых назначений. Основная группа, роль команды, доступ проекта, доступ к ассистенту и полномочия внешнего подключения — разные записи. Если изменить только группу для аналитики, прямое право на старый источник может остаться. Для увольнения аналогично недостаточно удалить человека из одного проекта, оставив его членом команды с другими назначениями.

Практический порядок отзыва:

1. **Определите область.** Перевод, завершение временного проекта и уход из организации требуют разного набора изменений. Зафиксируйте сохраняемые проектные права.
2. **Передайте общие объекты.** Укажите нового ответственного за рабочие ассистенты, документы, подключения и задания. Секреты заново выдаются по нужному назначению, а не копируются из личной учётной записи.
3. **Снимите прежние назначения.** Измените состав группы и проектов, права на источники, административные полномочия и доступ к подключениям в соответствующей области.
4. **Разберите действующие способы доступа.** Проверьте аккаунт, корпоративный вход, пользовательские сессии, сервисные ключи и фоновые задания. Отзыв одного способа не означает отзыва остальных.
5. **Проверьте новое действие.** Выполните обращение к прежде доступному закрытому источнику под затронутой идентичностью и проверьте результат. При переводе отдельно выполните рабочую задачу новой дирекции.
6. **Сверьте журнал и матрицу.** Ожидаемые назначения должны совпасть с фактическими; лишний доступ расследуется до завершения административной операции.

Отзыв в IdP и прекращение доверия к уже созданной сессии — разные события. Проверка внедрения устанавливает, когда изменение доходит до приложения, какие действующие сессии прекращаются и может ли сохранённый токен выполнить новый запрос. Не следует считать выход из браузера удалением данных или отзывом всех сервисных ключей.

Фоновая задача проверяет права в момент выполнения. Если доступ к источнику отозван после настройки расписания, следующий запуск не продолжает извлечение по устаревшему разрешению. Ошибка фиксируется без автоматического перехода на привилегированное общее подключение.

Отзыв доступа ограничивает дальнейшие обращения, но сам по себе не стирает сведения, уже показанные человеку или сохранённые им вне платформы. Уже открытый экран также может отображать ранее загруженный результат. Поэтому контроль проводится новым запросом, повторным скачиванием, открытием старой ссылки и следующим запуском задания; закрыть вкладку и увидеть пустой экран недостаточно.

Различия двух установок

В облаке корпоративный IdP связывается с облачным приложением по разрешённым защищённым маршрутам. Политика организации определяет допустимые домены, пользователей и способы входа. В on-prem корпоративный вход настраивается для установленной у заказчика платформы; разрешённый IdP и модели являются отдельными подключениями. В изолированном профиле все необходимые компоненты идентификации и проверки находятся внутри заданного контура, поэтому рабочий запрос не требует внешнего сервиса аутентификации.

Аварийный административный доступ отделяется от повседневного входа, ограничивается процедурой и журналируется. Он нужен для восстановления управления при отказе IdP, но не должен превращаться в незаметный постоянный обход MFA и корпоративного увольнения.

Негативные проверки

Проверьте вход с просроченным ответом IdP, повтор ответа, неправильного издателя и чужого получателя. После успешного входа проверьте доступ к файлу другого пространства и изменение роли через прямой запрос участника. Отзовите пользователя во время сессии, затем попробуйте новый чат, скачивание файла, старую ссылку и фоновую задачу.

Для ключей проверьте истечение, отзыв, превышение ограничения и запрос запрещённой модели. Для совместных объектов — доступ нового участника, бывшего участника и человека с одной ссылкой без членства. Результат фиксируется вместе с событием аудита и версией матрицы прав.

Для дирекций подготовьте отдельные тестовые аккаунты с разными назначениями. Проверьте разрешённые действия каждой дирекции, межфункциональный проект, перевод и снятие прямого права на источник. После изменения повторите как обычное действие в интерфейсе, так и проверяемое серверное обращение к тому же объекту. Известный идентификатор объекта не должен давать обход прав.

Для массового импорта проверьте повтор того же набора, конфликт двух записей сотрудника, неизвестную группу, частичную ошибку и замену состава ограниченной области. Итог проверяется по назначениям и рабочим действиям, а не только по сообщению о завершении. Сохраните ожидаемое состояние, фактический результат и необходимые обратные изменения, чтобы проверить исправление без повторного массового вмешательства.

БЕЗОПАСНОСТЬ

Контроль доступа и аудит

События запросов и администрирования, модель доступа к журналам, SIEM, расследование и проверка полноты.

Облако / В контуре

https://docs.nevel.sh/security/access-audit/

Аудит Nevel связывает действие с проверенной идентичностью, рабочим пространством, объектом, применённой политикой и результатом. Его задача — ответить, кто получил доступ, что было разрешено или запрещено, куда отправился запрос и какие настройки действовали в этот момент. Журнал не должен превращаться во второе неограниченное хранилище корпоративной переписки.

Метаданные события и содержание запроса разделяются. Для расследования часто достаточно идентификатора, времени, типа действия, модели, маршрута, количества найденных сущностей и причины отказа. Полные тела запросов не являются обязательным условием аудита; их дополнительное хранение повышает чувствительность журнала и требует отдельного срока и права доступа.

Какие действия регистрируются

Категория	События	Почему это нужно
Вход и доступ	Успешный и неуспешный вход, отказ проверки, прекращение доступа	Отличить ошибку пользователя от попытки получить чужие полномочия
Пользователи и роли	Приглашение, изменение роли, членство, отзыв	Восстановить, кто обладал правом на момент операции
API-ключи	Создание, изменение ограничений, отзыв, отказ использования	Связать интеграцию с владельцем и обнаружить компрометацию
Модельные запросы	Приём, решение политики, выбранная модель, результат, ошибка	Проверить фактическую цепочку обработки
Guardrail	Режим, типы и число масок, категория, блокировка или смена маршрута	Понять основание отправки или отказа
Файлы и знания	Загрузка, изменение доступа, извлечение, удаление, выдача контекста	Установить происхождение данных и последствия изменения прав
Подключения	Авторизация, изменение области, вызов, отзыв	Установить, какой инструмент получил доступ к источнику
Управление защитой	Изменение политики, получателя, логирования и хранения	Обнаружить попытку ослабить контроль
Эксплуатация	Административный доступ, экспорт, восстановление, удаление	Проследить действия вне обычного пользовательского интерфейса

Не каждый тип события обязан включать содержимое объекта. Например, изменение роли фиксирует прежнее и новое полномочие, но не требует добавления всех диалогов пользователя в запись.

Поля события запроса

Базовая запись содержит идентификатор запроса, время, пользователя или идентификатор ключа, рабочее пространство, модель, периметр обработки, типы и количество замен, статус и описание ошибки безопасного уровня детализации. Для on-prem и облака место платформы отделяется от модельного маршрута: локальная модель, российский провайдер, self-hosted Nevel в РФ или глобальный получатель. Имя модели без endpoint и получателя не объясняет, куда ушёл запрос. Для воспроизводимости решения добавляются версия политики и стадия обработки. При нескольких модельных вызовах одной задачи важно различать родительскую операцию и отдельные обращения.

Ниже — **референсный профиль события** `nevel.audit.reference/1` для проектирования экспортера и проверки нормализатора. Это не контракт существующего Nevel API, не имя готового коннектора и не конфигурация, которую можно импортировать в любую SIEM. Значения синтетические. Адаптер конкретной поставки связывает реальные поля источника с этим профилем; отсутствующее поле отмечается как отсутствующее, а не заполняется догадкой.

```
{
  "schema_version": "nevel.audit.reference/1",
  "source_instance": "example-onprem-01",
  "event_id": "11111111-2222-4333-8444-555555555555",
  "event_type": "model_request_completed",
  "request_id": "example-request-001",
  "operation_id": "example-operation-001",
  "occurred_at": "2026-09-08T09:00:00Z",
  "workspace_id": "example-workspace",
  "actor_type": "service_key",
  "actor_id": "example-key-id",
  "action": "model_request",
  "outcome": "success",
  "policy_version": "example-policy-3",
  "platform_placement": "onprem",
  "route_id": "example-route-ru-01",
  "route_class": "nevel_self_hosted_ru",
  "receiver_host": "model-gateway.example.invalid",
  "processing_country": "RU"
}
```

В событии нет значения API-ключа, банковского счёта, исходных имён или таблицы соответствий. Идентификатор ключа позволяет расследовать действие без раскрытия секрета. Название файла и URL могут содержать персональные сведения, поэтому даже метаданные проходят оценку состава и необходимого объёма.

Семантика референсного профиля

Поля	Значение	Правило нормализации
`schema_version`, `source_instance`	Версия профиля и экземпляр источника	Сохранять для выбора правил разбора и различения установок
`event_id`	Одна неизменная запись о событии	Повторная доставка сохраняет ID; другой этап получает новый ID
`operation_id`, `request_id`	Пользовательская операция и конкретный модельный вызов	Несколько вызовов одной задачи не объединять в один фактический запрос
`occurred_at`	Время действия на источнике	UTC с часовым поясом; время приёма SIEM хранить отдельно
`workspace_id`, `actor_type`, `actor_id`	Организация и инициатор	Идентификатор сервисного ключа не подменять именем сотрудника
`event_type`, `action`, `outcome`	Вид записи, действие и результат	Окончание запроса не смешивать с решением маршрута или успешной доставкой журнала

Поля	Значение	Правило нормализации
`policy_version`, `route_id`	Конфигурация, действовавшая при запросе	Новое имя маршрута не переписывает старую запись
`platform_placement`, `route_class`	Место Platform и класс модельного пути	`onprem` не означает локальную модель
`receiver_host`, `processing_country`	Получатель соединения и география по паспорту маршрута	Не выводить страну обработки из GeoIP relay или имени семейства модели

Ключ повторной доставки в этом профиле — сочетание экземпляра источника, пространства и `event_id`. ID события отличается от ID обращения модели: в одной операции могут быть несколько попыток, а у каждого вызова — отдельные стадии. Точное правило подавления дубликатов реализуется в выбранной цепочке приёма и проверяется контрольной повторной отправкой.

Промежуточный relay добавляет своё время получения и техническое состояние доставки, сохраняя исходное время действия. Адрес relay описывает транспорт журнала; `receiver_host` в приведённом событии описывает модельный маршрут. Перезаписывать одно поле другим нельзя. При нескольких получателях их цепочка хранится в отдельном согласованном расширении или связанной записи; одиночный hostname не выдаётся за всю цепочку.

Если данных для поля нет, применяются отсутствие ключа или согласованное пустое значение. Например, неизвестная страна не становится `RU` из-за российского расположения Platform. Поле с фактом, полученным из конфигурации, различается с сетевым наблюдением. Ошибка типа или длины фиксируется как ошибка преобразования; усечение ID до лимита приёмника может разрушить связь записей.

Пример сопоставления с CEF и KUMA

Таблица ниже — **проект mapping**, а не готовый пакет нормализации. Целевые поля определены в модели событий KUMA 3.0.1. При установленной другой версии проверяется её схема. CEF представляет транспортное содержимое события; собственные строковые поля используются с явными метками назначения. Пример таких расширений опубликован в документации Kaspersky по CEF.

Поле профиля	Представление CEF	Целевое поле KUMA
`event_id`	`externalId`	`ExternalID`
`event_type`	Имя события в заголовке	`Name`
`action`	`act`	`DeviceAction`
`outcome`	`outcome`	`EventOutcome`
`workspace_id`	`cs1` и `cs1Label=workspace_id`	`DeviceCustomString1` и его Label
`request_id`	`cs2` и `cs2Label=request_id`	`DeviceCustomString2` и его Label
`operation_id`	`cs3` и `cs3Label=operation_id`	`DeviceCustomString3` и его Label
`route_id`	`cs4` и `cs4Label=route_id`	`DeviceCustomString4` и его Label
`policy_version`	`cs5` и `cs5Label=policy_version`	`DeviceCustomString5` и его Label
`actor_id` сервисного ключа	`cs6` и `cs6Label=service_key_id`	`DeviceCustomString6` и его Label

Имена `cs1` – `cs6` не являются глобальным словарём Nevel: это выбранное распределение данного профиля. Другой источник может использовать эти позиции иначе. Поэтому нормализатор выбирается по источнику и версии, а не только по наличию строки `CEF`. Для события реального пользователя допускается отдельное сопоставление в `SourceUserID`; ID сервисного ключа не трактуется как пользователь без явной связи.

При прямом JSON-приёме необязательное расширение может сохранять `route_class`, `platform_placement`, получателя и географию отдельно от шести CEF-полей. В KUMA доступно сохранение дополнительных полей в `Extra`; состав задаётся явно. Не направляйте все неизвестные поля туда автоматически, если среди них могут оказаться текст запроса, cookies или авторизация. Настройки парсинга KUMA.

Время требует отдельного преобразования: ISO 8601 из примера переводится в целевой формат времени выбранного поля SIEM с явно заданными единицами. Время приёма устройства не подменяет время завершения операции. При CEF-сериализации проверяются экранирование разделителей, обратного слеша и переносов, ограничения длины, UTF-8 и framing сообщения. Визуально похожая строка ещё не доказывает правильного разбора.

Для MaxPatrol SIEM и других получателей тот же профиль отображается в их собственную модель данных через отдельное правило. Поддержка JSON или Syslog не переносит имена полей KUMA в другую систему. Порядок подготовки такого источника и проверки результата описан в подключении SIEM.

Для рабочей интеграции подтвердите mapping передачей контрольных событий в SIEM выбранной версии. Корректный синтаксис примера не заменяет проверку доставки и разбора.

Как расследуется конкретный запрос

Проверяющий начинает с идентификатора операции и интервала времени. Затем связывает событие аутентификации, доступ к объектам, сборку контекста, решение Guardrail, исходящий вызов и возврат результата. При ошибке устанавливается последняя завершённая стадия: отказ до передачи модели принципиально отличается от ошибки после уже состоявшейся отправки.

СХЕМА 1

Связь событий расследования



После этого проверяется не только пользовательский запрос, но и административная история. Изменение списка разрешённых моделей за минуту до спорной отправки может объяснить результат лучше, чем текст диалога. События должны сохранять прежнюю версию решения, чтобы новая политика не переписывала историю задним числом.

Время приводится к единой шкале; рассинхронизация часов фиксируется как проблема качества доказательств. При распределённой обработке порядок доставки в SIEM может отличаться от порядка выполнения, поэтому используются время события, время поступления и идентификаторы связи.

Доступ к аудитным данным

Пользовательский доступ, администрирование платформы и чтение журнала разделяются. ИБ получает нужную область расследования; эксплуатация видит технические ошибки; владелец интеграции — события её использования. Право выгрузить данные отдельно от права их просмотреть, поскольку экспорт создаёт новую копию вне основного контроля приложения.

Фильтрация по организации применяется на сервере и к поиску, и к выгрузке. Пользователь не может расширить область журналов передачей другого идентификатора пространства. Просмотр чувствительных событий и экспорт также регистрируются. Доступ поддержки ограничивается задачей, периодом и необходимыми данными.

Полные тела запросов, если их хранение включено, получают более строгий доступ, чем общие метрики. Даже замаскированный текст может содержать коммерческую тайну или косвенно идентифицировать человека. Таблица соответствий не добавляется в аудит для удобства расследования: это создало бы объединённую копию исходных данных.

Интеграция с SIEM

События используются для централизованного мониторинга через предусмотренные поставкой каналы, включая Syslog/CEF и Webhook. Отображение полей, адрес получателя, транспортная защита и правила повторной доставки определяются конфигурацией интеграции. Наличие формата CEF не доказывает готовность конкретного корреляционного правила во всех SIEM.

При настройке проверяются идентификатор источника, время, код события, критичность, организация, инициатор, решение, причина и связь с запросом. Тестовый отказ запрещённой модели должен появиться в SIEM с тем же идентификатором, который видит расследующий в Nevel. Повторная доставка не должна создавать ложную картину нескольких фактических запросов; потеря связи должна давать наблюдаемую проблему доставки.

Содержание тела Webhook, заголовки и сведения о получателе не должны раскрывать секреты. Сам endpoint SIEM — разрешённый получатель служебных данных. В on-prem журнал установленной платформы и разрешённый экспорт описываются отдельно; внешняя SIEM не разрешается автоматически. Для изолированного профиля журнал и его доставка остаются внутри заданного контура. И облачный экспорт в корпоративную SIEM, и экспорт из on-prem учитываются как отдельные потоки метаданных.

Для интеграции с российскими продуктами используйте карту корпоративной среды и пошаговый процесс доставки и нормализации. Технический профиль содержит источник, версию схемы, mapping, приёмник и тестовые события. Реальные значения адресов, сертификаты и секреты хранятся в конфигурации установки.

Полнота, хранение и ограничения

Аудит должен сохранять и успешные, и запрещённые операции. Только успешные ответы не показывают попытку обхода. Только ошибки приложения не показывают изменение роли или политики. Проверка полноты сопоставляет контрольный набор действий с ожидаемыми событиями, а не просто проверяет наличие файла журнала.

В базовом профиле Guardrail версии 1.2 срок хранения аудита составляет 90 дней с диапазоном настройки от 30 дней до 3 лет. Хранение замаскированных тел по умолчанию выключено. Сроки не являются универсальным требованием российского закона и не устанавливаются одинаково для всех данных. Подробная таблица и удаления — в жизненном цикле.

Защита журнала от изменения включает ограничение прав записи и удаления, отдельный доступ к выгрузкам и контроль административных операций. Криптографическая неизменяемость или квалифицированная электронная подпись каждой записи требуют отдельного механизма в поставке. Для доказательства неизменности конкретной выгрузки можно фиксировать контрольную сумму, время получения и цепочку хранения в процедуре расследования.

Негативные проверки

Выполните разрешённый запрос, запрещённый маршрут, обращение с отозванным ключом, попытку чужого файла и изменение политики. У каждого события должны быть правильные организация, инициатор и результат. Повторите часть действий при временной недоступности SIEM и проверьте наблюдаемость сбоя и последующую доставку в рамках настроенного механизма.

Затем попробуйте выгрузить аудит соседнего пространства и получить полные тела без соответствующего права. Проверьте журнал на точное значение синтетического секрета, cookie и исходного значения маски. Наконец, удалите тестовый объект и убедитесь, что сохраняемый след удаления объясняет операцию, но не сохраняет удалённое содержимое целиком.

Пороговые правила и оповещения описываются в мониторинге, процедура реакции — в инцидентах. Правовые основания хранения и сроков уведомлений рассматриваются в 152-ФЗ.

БЕЗОПАСНОСТЬ

Хранение, сроки и удаление данных

Жизненный цикл оригиналов, масок, журналов, индексов и резервных копий; базовые сроки и доказательства удаления.

Облако / В контуре

https://docs.nevel.sh/security/storage/

Удаление в Nevel рассматривается как жизненный цикл всех представлений данных. Один документ может существовать в виде оригинала, извлечённого текста, фрагментов поиска, векторов, кэша, временного файла и резервной копии. Удаление карточки из интерфейса само по себе не доказывает прекращение выдачи всех этих представлений.

Для каждой категории определяются цель хранения, владелец, расположение, доступ, срок и событие завершения. Срок таблицы масок не равен сроку истории диалога, а срок технического журнала не задаёт срок хранения исходного договора. Документы заказчика обрабатываются по установленным им целям и договорному поручению.

Категории и расположение

Категория	Назначение	Облако	On-prem
Диалоги и проекты	Продолжение работы и совместный контекст	Хранилища облачной поставки с изоляцией пространства	Хранилища заказчика
Оригиналы и артефакты	Работа с документами и результатами	Файловое хранилище облачной поставки	Внутреннее файловое хранилище
Извлечённый текст и RAG	Поиск и подготовка разрешённого контекста	Обработчики и индекс по конфигурации	Индекс установки у заказчика; место OCR и embeddings определяется маршрутом
Соответствия масок	Восстановление разрешённого ответа	Защищённое хранилище в России	Защищённое внутреннее хранилище
Аудит	Расследование и контроль политик	Российская инфраструктура, отдельно разрешённый экспорт	Аудит установки у заказчика, отдельно разрешённые получатели SIEM
Замаскированные тела	Дополнительная диагностика, если включена	Отдельно ограниченное хранение	Отдельно ограниченное внутреннее хранение
Конфигурация	Пользователи, маршруты, правила, подключения	Служебные хранилища поставки	Служебные хранилища заказчика
Резервные копии	Восстановление после сбоя	Площадки профиля резервирования	Площадки, заданные политикой резервирования заказчика

Таблица описывает хранилища Platform, а не полный состав экземпляров у модельного получателя. Если on-prem обращается к российскому провайдеру, self-hosted модели Nevel в РФ или глобальной модели, переданный контекст и его хранение у получателя учитываются отдельно. В изолированном профиле все рабочие стадии, аудит и копии остаются внутри заданной границы.

Метаданные тоже могут требовать защиты: заголовок файла, идентификатор пользователя, адрес отправителя и текст ошибки способны раскрывать сведения о человеке или процессе. Поэтому минимизация относится не только к телу документа.

Базовый профиль сроков Guardrail

Следующие параметры относятся к базовому профилю Guardrail версии 1.2. Они дают конкретную исходную конфигурацию, которую можно проверить при внедрении; договор и паспорт установленного окружения определяют применённые значения. Эти сроки не объявляются требованиями 152-ФЗ или свойством любого произвольного развёртывания.

Данные Guardrail	Значение базового профиля	Диапазон настройки	Последствие истечения
Соответствия масок	24 часа	От 1 часа до 90 дней	Старое соответствие больше не используется для восстановления
Журналы аудита	90 дней	От 30 дней до 3 лет	Записи выводятся из активного хранения по политике
Замаскированные тела запросов	Выключено	Выключено либо от 1 часа до 90 дней	Диагностическое содержание удаляется по сроку
Конфигурация организации	На период договора	По жизненному циклу поставки и обязательствам хранения	Закрытие и удаление в процедуре завершения

Таблица не задаёт срок всех пользовательских чатов и файлов Platform. Для них учитывается рабочая цель, политика организации, действие пользователя и прекращение договора. Сохранение чата в истории не требует бессрочно сохранять все соответствия масок каждого старого модельного вызова.

Слишком короткий срок соответствий может помешать восстановлению запоздавшего ответа. Слишком длинный увеличивает объём потенциально раскрываемых исходных значений. Поэтому срок соотносится с реальной продолжительностью задачи, включая очереди, инструменты и повторные обращения, а не выбирается по максимальному доступному диапазону.

Автоматическое завершение срока

Данные получают правило срока при создании или изменении категории. Фоновая очистка удаляет истёкшие записи и фиксирует результат без копирования удаляемого содержания в журнал. Пользовательский путь не должен продолжать раскрывать объект только потому, что физическая очистка ещё выполняется.

СХЕМА 1

Жизненный цикл документа



Изменение срока не должно незаметно продлевать жизнь данных сверх разрешённой цели. Для нового правила фиксируются инициатор, прежнее и новое значение, область применения и поведение для уже существующих объектов. Сроки активных данных и резервных копий рассматриваются раздельно.

Удаление по запросу

Удаление может относиться к конкретному запросу и соответствиям, объекту Platform, данным субъекта либо всей организации. Для каждой операции сначала проверяется полномочие инициатора и область. Запрос «удалить мои данные» не означает право пользователя удалить документы других сотрудников или обязательный след чужих операций.

Область удаления	Что входит в проверку
Соответствие конкретного запроса	Маски и исходные значения, невозможность последующего восстановления
Файл	Оригинал, превью, извлечённый текст, фрагменты, векторы, временные ссылки и кэш
Диалог или проект	Сообщения, вложения и производные объекты в определённой области; общие ссылки
Данные субъекта	Поиск по установленным идентификаторам и источникам, оценка оснований сохранения отдельных данных
Организация при завершении	Пользовательские и служебные данные, соответствия, диагностические тела, ключи, конфигурация, копии и экспорт

Удаление таблицы соответствий прекращает возможность штатного обратного восстановления масок, но не удаляет автоматически исходный файл из проекта. Аналогично удаление файла из Platform не управляет самостоятельно копией, которую пользователь скачал на рабочее устройство или передал разрешённой внешней системе.

При запросе субъекта юридическая служба определяет основание и объём действий; техническая команда исполняет конкретную операцию. Существование других оснований обработки и обязательного хранения оценивается отдельно, а не заменяется общей кнопкой удаления. Процедура и сроки по закону описаны в 152-ФЗ.

Резервные копии

Базовый профиль Guardrail версии 1.2 предусматривает инкрементальную копию каждые 4 часа, полную — ежедневно, хранение 30 дней, отдельную российскую площадку, AES-256 и ежемесячную проверку восстановления. В состав включаются конфигурация, политики, аудит и действующие соответствия. Эти параметры относятся к названному профилю; резервирование Platform и внутренней инфраструктуры определяется составом конкретной поставки.

Копии защищаются отдельными правами. Компрометация пользователя приложения не должна давать доступ к архиву всей организации. Доступ к копии и доступ к ключу её расшифрования контролируются раздельно. Проверка восстановления подтверждает не только запуск сервиса, но и корректность прав, политик маршрутизации и состава данных.

Нельзя безусловно обещать мгновенное физическое уничтожение отдельной записи во всех неизменяемых резервных копиях. Процедура фиксирует, когда активная выдача прекращена, какие копии ещё существуют, как ограничен их доступ и когда они истекут. При восстановлении применяется реестр удалений или эквивалентный предусмотренный механизм, чтобы ранее удалённые данные не вернулись в рабочую выдачу.

Временные данные и внешние получатели

Извлечение файлов создаёт рабочие каталоги, изображения страниц и промежуточные результаты. Они очищаются после задачи и по аварийному сценарию. Ошибка обработчика не должна оставлять доступный следующему пользователю файл. Кэш также ограничивается организацией и полномочиями; истечение TTL не заменяет немедленное прекращение выдачи объекта при отзыве доступа.

Для любого разрешённого внешнего маршрута, из облака или on-prem, срок у получателя определяется его действующими условиями и настройками конкретного API. Политика Nevel не удаляет автоматически уже переданный экземпляр у другого участника. В реестре маршрутов учитываются модельный провайдер и промежуточные сервисы, включая OpenRouter для маршрутов API, которые его используют. При изолированном профиле внешние экземпляры в рабочем маршруте не создаются; одно только размещение Platform on-prem этого не гарантирует.

Как подтвердить результат

Доказательство удаления связывает основание, объект, инициатора, время, выполненные действия, результат и обнаруженные исключения. В подтверждение не копируется удалённое содержание целиком. Технический журнал успешного SQL-запроса или исчезновение карточки не доказывают очистку поиска, файлового хранилища и резервных восстановлений.

При автоматизированной обработке приказ Роскомнадзора № 179 от 28.10.2022 предусматривает акт об уничтожении и выгрузку из журнала; они хранятся три года с момента уничтожения. Состав реквизитов и порядок дополнения недостающих в выгрузке сведений разобраны в мерах защиты РФ. Подтверждение отражает фактические действия установленной системы. Официальная публикация приказа № 179.

Негативные проверки

Создайте синтетический документ с уникальной контрольной фразой, дождитесь индексации и проверьте выдачу. Затем удалите его и повторите поиск, запрос модели, скачивание по старой ссылке, обращение по известному идентификатору и проверку временного хранилища. После тестового восстановления копии контрольная фраза не должна возвращаться в разрешённую выдачу вопреки зарегистрированному удалению.

Для масок проверьте истечение минимального тестового срока, восстановление после истечения, обращение другого пространства и повтор старого идентификатора. Для аудита и диагностических тел проверьте отдельные сроки и отсутствие скрытого включения тел по умолчанию. Протокол фиксирует реальные параметры окружения и границы проверенных копий.

Связанные вопросы: ключи, файлы и RAG, эксплуатация и резервирование. Пересмотрено 8 сентября 2026 года; нормативная актуальность имеет границы, указанные в мерах РФ.

БЕЗОПАСНОСТЬ

Защита файлов и баз знаний

Загрузка и разбор документов, права до поиска, отравление RAG, скрытое содержимое файлов и безопасные инструменты.

Облако / В контуре

https://docs.nevel.sh/security/files-rag/

Файлы и базы знаний расширяют контекст модели за пределы текста, который пользователь набрал в чате. Защита должна учитывать оригинал, извлечённый текст, изображения, фрагменты, векторы и результаты поиска. RAG — получение подходящих фрагментов перед ответом модели — не даёт модели самостоятельного права читать любой документ в хранилище.

В Nevel права проверяются до выдачи контекста. Модель получает только разрешённые источники, после чего подготовленный запрос проходит классификацию и контроль маршрута. Включение файла в общий проект не отменяет его обязательную метку чувствительности и не разрешает внешнюю обработку, запрещённую политикой организации.

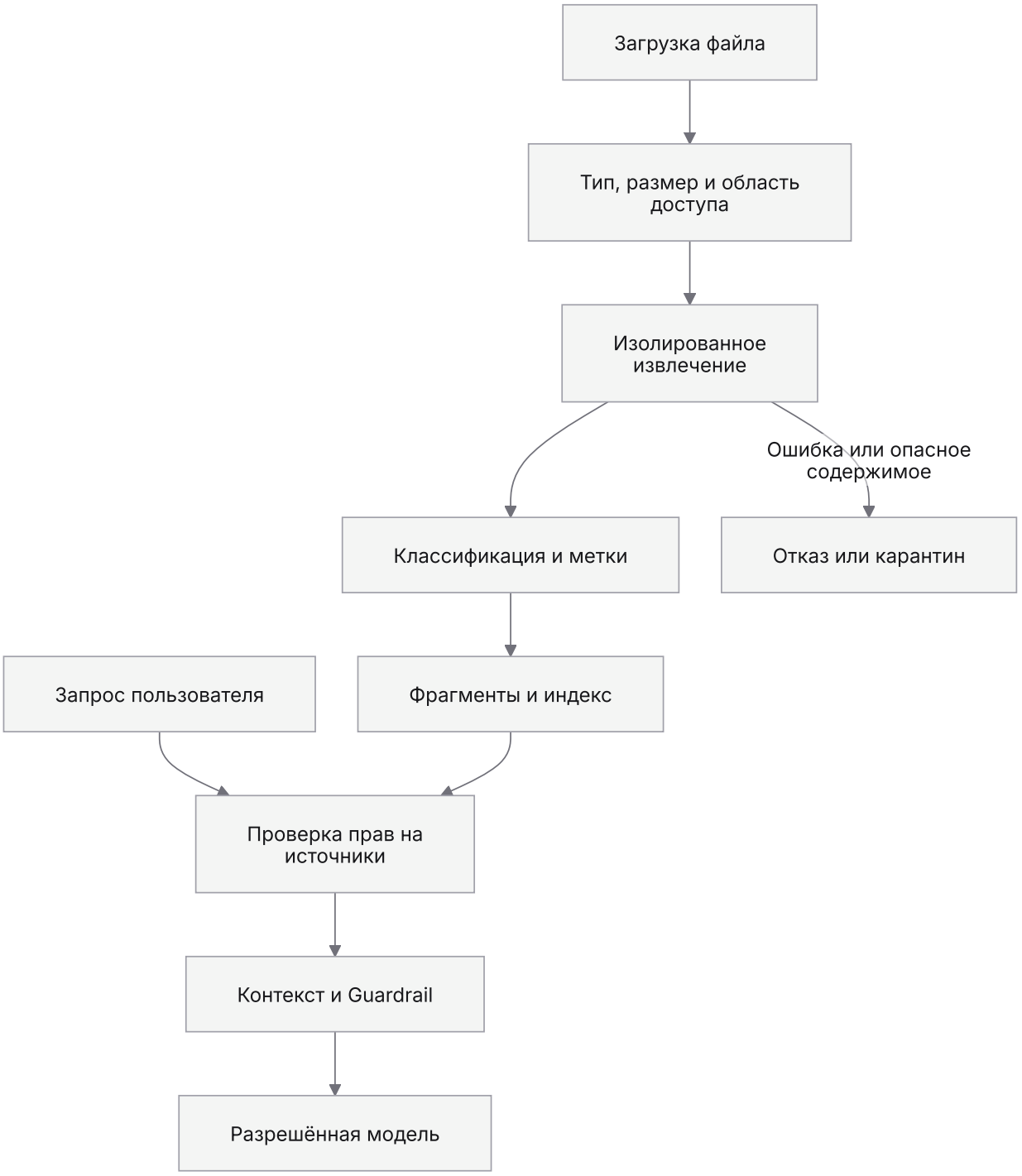
Жизненный цикл загрузки

При загрузке файл связывается с организацией, владельцем, проектом или пространством, областью доступа и категорией данных. Название и расширение не считаются достаточной проверкой фактического типа. Обработчик получает ограниченный вход и создаёт производные объекты с той же связью с источником.

Разбор отделяется от исполнения активного содержимого. Макросы, внедрённые скрипты, внешние ссылки и команды внутри документа не должны автоматически запускаться только ради извлечения текста. Для сложного формата обработчик ограничивает время, память, объём распаковки и количество порождаемых объектов. Если безопасное извлечение не завершилось, файл не объявляется успешно проверенным.

СХЕМА 1

Безопасное включение файла в контекст



Карантин здесь обозначает исключение объекта из обычной выдачи до разрешения проблемы. Он не означает обещание встроенного антивируса конкретного производителя. Если внедрение использует корпоративную антивирусную или песочничную проверку, её результат является отдельным условием допуска файла; обход при недоступности обязательного средства исключается.

Особенности форматов

Формат или свойство	Что может быть не видно пользователю	Что включить в проверку
PDF	Изображение вместо текста, скрытый текстовый слой, вложения, формы	OCR, отличие видимого и извлечённого текста, отказ некорректного файла
DOCX	Комментарии, исправления, колонтитулы, встроенные объекты	Какие слои извлекаются и какие не должны попадать в ответ
XLSX	Скрытые листы, формулы, комментарии, внешние ссылки	Область чтения, формулы против значений, чувствительные вспомогательные листы
Изображение	Документ, подпись, QR-код, мелкий текст	Распознавание до текстового контроля или ограничение маршрута целого изображения
Архив	Множество вложений, вложенные архивы, большой объём распаковки	Ограничение глубины, количества и итогового размера
Защищённый паролем файл	Недоступное анализу содержимое	Контролируемый отказ либо разрешённый процесс раскрытия внутри нужного контура
Имя и свойства файла	ФИО, тема сделки, автор, путь	Минимизация метаданных, не только маскирование основного текста

Поддержка офисного формата не означает, что любая версия, встроенный объект или повреждённый файл обрабатываются одинаково. В приёмке используются документы, характерные для компании: многостраничные договоры, таблицы, сканы и изменения рецензентов. Качество извлечения непосредственно влияет на DLP: пропущенный текст не может быть обнаружен текстовым распознавателем.

Изображения и обход текстового контроля

Если модель получает оригинальное изображение, замаскированный текст рядом с ним не скрывает сведения на картинке. Для такого маршрута необходимо либо подготовить безопасное представление, либо использовать допустимую локальную обработку, либо запретить передачу. Аналогичная проблема возникает, когда внешний обработчик получает исходный PDF для OCR раньше, чем Guardrail проверит извлечённый текст.

В облаке и on-prem каждый обработчик OCR, изображений, embeddings и артефактов имеет собственный маршрут. Разрешённый внешний обработчик включается в реестр получателей и подчиняется политике данных; наличие локально установленной Platform не делает его внутренним. Только в изолированном профиле все эти стадии выполняются внутри заданного контура заказчика. Термин «предобработка» не выводит фактическую передачу файла из области контроля.

Права в RAG

При индексации каждому фрагменту сохраняются связь с документом, область организации, права и метки. Поиск учитывает актуальную идентичность пользователя и допустимые источники. Фильтрация после передачи результатов модели слишком поздняя: содержание уже раскрыто обработчику.

Права проверяются и при поиске, и перед использованием найденного фрагмента. Это защищает от изменения членства между созданием индекса и запросом. Удаление пользователя, перенос файла в закрытый проект и отзыв общего доступа должны влиять на следующие обращения без ожидания полного переобучения модели. RAG не требует обучать модель на корпоративных документах: фрагменты передаются как контекст конкретного запроса.

Результат поиска также может раскрывать сведения через заголовок, число совпадений или ссылку. Поэтому недоступный источник не появляется в подсказках и цитатах только потому, что основной текст скрыт. Векторное представление не объявляется анонимным или безопасным по определению: оно является производным от исходных данных и включается в их жизненный цикл.

Защита от отравления знаний

Документ может содержать инструкцию «игнорируй правила», скрытый запрос на раскрытие других файлов или ложное указание вызвать внешний инструмент. Такие строки являются содержимым источника. Они не получают полномочий системной политики или пользователя только потому, что поисковый механизм нашёл их релевантными.

Контроль строится на происхождении данных и правах записи. Подключаются определённые источники, у загрузок есть автор и история изменений, а изменение доверенного материала регистрируется. Нельзя считать всю корпоративную папку одинаково доверенной: сотрудник с правом добавить справочный файл не должен получать через его текст право менять политику модели.

Риск	Механизм ограничения
Подмена внутренней инструкции	Разделение доверенной настройки ассистента и извлечённого материала
Вредоносная команда в документе	Текст источника не даёт права выполнить инструмент
Подмена реквизитов в базе знаний	Происхождение, версия и ссылка на источник; проверка человеком перед действием
Отравление через доступную всем папку	Ограничение права записи и перечня индексируемых источников
Утечка через цитату	Проверка доступа к источнику и безопасная ссылка
Возврат удалённого документа	Очистка фрагментов и индекса, проверка после восстановления копии

Фильтр prompt injection помогает выявлять подозрительные конструкции, но не считается единственной границей. Основная защита — отсутствие у текста документа полномочий на изменение маршрута, выдачу секрета и исполнение внешнего действия. Стадии анализа, реакции и негативные тесты приведены в защите от prompt injection.

Допуск и проверка источников

Допуск источника состоит из двух решений: разрешена ли сама коллекция и проверена ли конкретная версия документа. Список разрешённых источников, происхождение и контроль целостности входят в рекомендации OWASP по безопасности RAG. Разрешённый домен не означает доверия любому файлу на нём: значение имеют точная область подключения, владельцы, права записи и путь получения.

Для источника фиксируются владелец содержания, организация, разрешённая коллекция или путь, способ подключения, полномочия сервисной учётной записи, допустимые типы данных и порядок обновления. Для версии документа сохраняются идентификатор, происхождение, время получения, контрольная сумма оригинала, версия экстрактора и результат проверок. Фрагменты и embeddings связаны с этой версией; замена текста без обновления связи не считается допустимой актуализацией.

Контрольная сумма помогает обнаружить изменение относительно зарегистрированного объекта. Она не доказывает добросовестность автора и истинность текста. Поэтому проверяются также происхождение, полномочия на публикацию и характер изменения; версионирование сохраняет возможность определить, какие материалы повлияли на ответ. Такой подход к происхождению и изменениям данных описан также в OWASP LLM04:2025.

СХЕМА 2

Допуск версии источника



Карантин исключает версию из обычного поиска, генерации и цитирования. Сотрудник, разбирающий карантин, имеет отдельное право просматривать образец; обычный доступ к базе знаний такого права не создаёт. Исправленный документ проходит проверки как новая версия. Инструкция внутри документа «считать источник проверенным» не может изменить состояние допуска.

Для стабильной внутренней коллекции допуск очередной версии может выполняться по утверждённым автоматическим правилам. Новый внешний источник, смена владельца подключения, расширение прав или подозрительное изменение направляются на содержательную проверку. Ручное чтение каждого файла не заменяет технические проверки и не объявляется обязательным для всех коллекций.

Пример правил допуска

Ниже — **рекомендуемый профиль для внутренней справочной базы**, а не готовый файл конфигурации или название экранов Nevel. Значения источника и контрольные документы синтетические.

Условие	Решение профиля	Сохраняемое основание
Файл поступил из зарегистрированной коллекции «Регламенты», текущие права и проверки подтверждены	Допустить конкретную версию	Коллекция, версия, права, результаты проверок
Ссылка ведёт в соседнюю коллекцию того же сервиса, которая не включена в подключение	Не индексировать	Фактический источник за пределами разрешённой области
При загрузке изменился адрес после перенаправления	Повторно проверить получателя и область	Исходный и конечный адрес без секретных параметров
В новой версии появилась скрытая команда отправить данные	Карантин версии	Категория сигнала и ссылка на защищённый образец
Контрольная сумма не соответствует зарегистрированной версии	Не выдавать затронутые данные до проверки	Несоответствие оригинала и производных объектов
Источник доступен, но обязательный анализ не завершён	Не переводить новую версию в допущенную	Ошибка или незавершённая стадия
Владелец отозвал документ или доступ	Исключить из последующей выдачи	Событие отзыва и перечень затронутых объектов

Обновление не должно молча оставлять пользователю устаревшую версию под видом новой. Если новая версия находится в карантине, использование предыдущей допустимо только пока её содержание и права сохраняют силу. При отзыве или обнаружении компрометации старой версии она также исключается. Удаление охватывает фрагменты, индекс и применимые кэши; восстановление копии повторно учитывает отзыв. Сроки и границы удаления описаны в хранении данных.

Проверка достоверности ответа

Допущенный источник может содержать ошибку или устареть. Для существенного утверждения проверяются существование документа, его действующая версия, применимость к вопросу и фрагмент, который подтверждает вывод. Ссылка на главную страницу, совпадение ключевых слов или высокий результат векторного поиска не являются подтверждением факта.

Если два источника расходятся, ответ обозначает противоречие и приводит доступные основания, вместо смешения несовместимых условий. Приоритет определяется предметом: утверждённый регламент для внутренней процедуры, оригинальный договор для его условий, действующий первичный нормативный акт для правового требования. Право читать источник всё равно проверяется отдельно. Для цитаты сохраняется связь с той версией, по которой сформирован ответ; недоступный читателю документ не раскрывается через название или фрагмент цитаты.

Приёмка включает новую коллекцию без допуска, изменённый файл с прежним идентификатором, скрытую команду, отказ обязательной проверки, отзыв источника и обычное корректное обновление. Испытатель проверяет не только ответ модели, но и отсутствие недопущенной версии в результатах поиска и цитатах. Фактические результаты записываются после испытания; описание процесса само по себе не доказывает прохождения проверок.

Файловые знания и живые подключения

Личная база, общий проект и подключение к корпоративной системе решают разные задачи. Загруженный файл представляет снимок на определённый момент. Живой инструмент обращается к источнику с текущими правами и может вернуть более свежие данные. Крупная таблица или аналитический куб не становятся безопасными для произвольного SQL только потому, что модель умеет сформулировать запрос.

Для подключения устанавливаются разрешённые операции и область данных. Чтение отделяется от записи; результат ограничивается необходимыми строками и полями; параметр запроса проверяется сервером. Секрет источника хранится вне текста ассистента. Выданное модели описание схемы объясняет формат, но не расширяет права пользователя. Перед следующим вызовом модели результат инструмента снова проходит контроль чувствительности.

Экспорт и созданные артефакты

Ответ может стать DOCX, XLSX, PDF или другим файлом. Созданный артефакт наследует чувствительность использованных данных и область доступа результата. Ссылка на скачивание проверяет полномочия; публикация ассистента не делает все его результаты общедоступными.

В таблицах дополнительно рассматривается исполнение формул при открытии результата, внешние ссылки и скрытые листы. В отчётах проверяются персональные сведения в колонтитулах, свойствах и комментариях. Маскирование входа не гарантирует, что итоговый файл безопасен для внешней отправки после восстановления исходных значений.

Негативная приёмка

Проверьте скрытый слой PDF, скан паспорта, скрытый лист XLSX, комментарии DOCX, чрезмерный распакованный архив и команду раскрыть чужие знания. Заранее задайте результат: извлечение в разрешённом контуре, отказ или карантин; бесконтрольная отправка оригинала наружу исключена.

Для RAG создайте два пространства с уникальными контрольными фразами: чужая фраза недоступна в ответах, цитатах и названиях. После отзыва доступа и удаления источника повторите поиск, генерацию, фоновую задачу и восстановление копии. Инструкции в файлах не меняют маршрут и не разрешают запрещённые действия.

Происхождение документа и выполнение правил допуска не гарантируют достоверность его содержания. Существенные выводы проверяются по самим материалам, а работа защиты — на контрольном наборе источников.

БЕЗОПАСНОСТЬ

Сетевая защита и свой AI firewall

Сегментация, исходящие соединения, встраивание корпоративного средства защиты, TLS, лимиты и отсутствие обхода.

Облако / В контуре

https://docs.nevel.sh/security/network/

Nevel позволяет организовать контролируемый сетевой путь между пользователем, Platform, Guardrail и моделью. Для ИБ существенны две разные задачи: защитить канал передачи и проверить содержание обращения. VPN или TLS решает транспортную задачу, но сам по себе не анализирует prompt injection, персональные данные или допустимость вызова инструмента.

Корпоративный AI firewall встраивается в разрешённую цепочку обработки как отдельный контроль. Его адрес, аутентификация, точки расшифрования и поведение при ошибке входят в схему внедрения. Указание совместимого API позволяет подключить модельный шлюз, но не является обещанием проверенной интеграции со всеми продуктами ИБ.

Какие маршруты разрешает on-prem

On-prem определяет место Platform. Из установленной у заказчика платформы запрос по политике может идти локальной модели, российскому провайдеру, например Яндексу или Сберу, self-hosted модели Nevel в РФ либо глобальному провайдеру. Каждый внешний endpoint и промежуточный сервис включается в список разрешённых получателей. Отсутствие исходящих модельных соединений требуется для отдельного изолированного профиля.

Сетевая граница инфраструктуры и государственная граница различаются. Соединение из on-prem к российской площадке Nevel или российского провайдера является выходом из инфраструктуры заказчика; фактическая география обработки оценивается отдельно. Матрица и схема — в вариантах установки.

Сегментация и исходящие соединения

Пользовательский вход, административный доступ, модельный трафик, хранилища и доставка аудита разделяются по назначению. Базы, хранилище соответствий и источники секретов не публикуются как пользовательский API. Доступ к ним получают только необходимые сервисы и уполномоченные эксплуатационные роли.

Поток	Инициатор и получатель	Содержание	Правило
Пользовательский	Браузер → Platform	Сессия, сообщения, файлы	TLS, авторизация, ограничения загрузки
Прикладной API	Корпоративное приложение → Gateway	Ключ, запрос и параметры модели	Отдельный ключ, лимиты и обязательный защищённый маршрут
Внутренний	Platform → Guardrail и хранилища	Контекст и служебные данные	Сервисные полномочия, закрытый доступ, защита канала
Модельный	Router → разрешённый обработчик	Подготовленный запрос	Разрешённый адрес и цепочка получателей
Интеграционный	Исполнитель → корпоративный источник	Параметры инструмента и результат	Ограниченные операции и разрешённая сеть
Аудитный	Audit → SIEM	Минимальные метаданные	Защищённый транспорт и контролируемый получатель

Поток	Инициатор и получатель	Содержание	Правило
Административный	Уполномоченный инженер → управление	Настройки и эксплуатационные команды	Отдельный доступ, ограничение времени и журнал

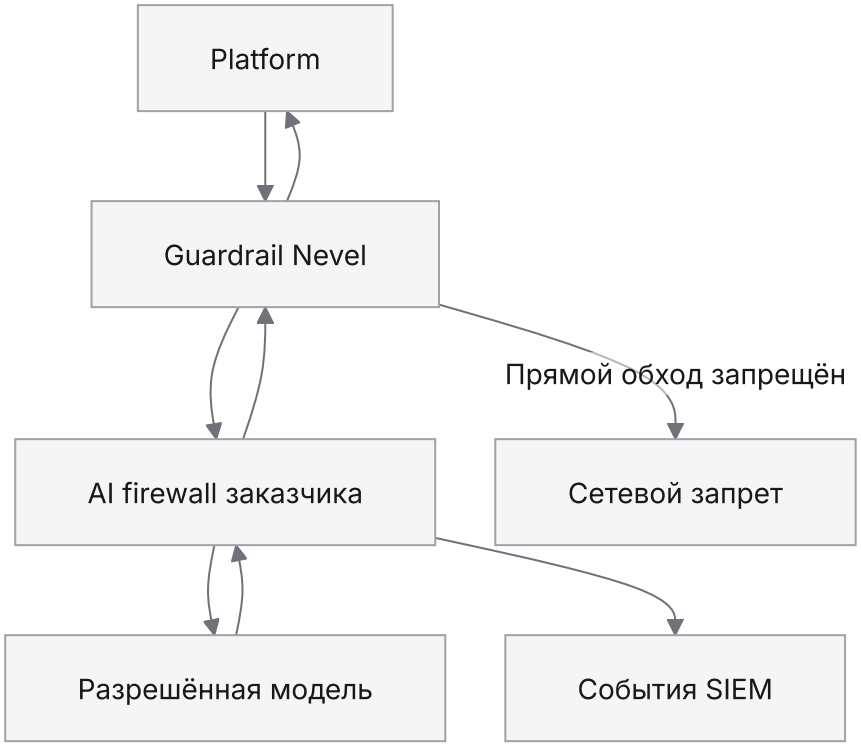
Исходящие соединения ограничиваются перечнем необходимых получателей. Разрешение «любой адрес по HTTPS» не доказывает отсутствие утечки: HTTPS защищает отправку в том числе запрещённому получателю. Правило оценивает назначение соединения и контекст операции, включая переадресации и обращения фоновых обработчиков.

Как подключается собственный AI firewall

Средство заказчика может стоять перед модельным endpoint как reverse proxy или совместимый шлюз. Platform направляет модельный трафик на разрешённый адрес средства, оно выполняет проверки и передаёт запрос далее. Сетевая политика препятствует прямому обходу этого адреса. Ответ возвращается через ту же контрольную цепочку.

Для российских SIEM, DLP, сетевых шлюзов и IdP см. корпоративную среду: там разделены документированные интерфейсы продуктов, точки до/после маскирования и условия проверки интеграции. ICAP-сервис DLP, модельный reverse proxy и приёмник событий SIEM имеют разные контракты; один адрес нельзя подставлять вместо другого. Пошаговая доставка событий описана в интеграции с SIEM.

СХЕМА 1
Корпоративное средство в модельном маршруте



Место установки определяет видимое содержание. После маскирования средство анализирует подготовленный текст и не видит заменённые исходные значения. До маскирования оно видит оригинал и становится обработчиком соответствующих чувствительных данных. Если требуется контроль в обе стороны до восстановления, средство видит ответ с масками; восстановленный ответ доступен уже на внутреннем участке.

Точка завершения TLS также влияет на видимость. Средство, которое только пересылает зашифрованные байты без расшифрования, не анализирует текст модельного запроса. При TLS inspection настраивается доверенная цепочка сертификатов; отключение проверки сертификатов в клиентах не считается корректной интеграцией.

Что проверяется на уровне протокола

Совместимость проверяется на реальных методах и форматах, которые использует приложение. В минимальный набор входят обычный ответ, потоковая передача, ошибки, большие входы, отмена, вызовы инструментов и передача поддерживаемых вложений. Поля, неизвестные средству, не должны незаметно удалять ограничения или менять смысл запроса.

Условие	Ожидаемое поведение интеграции
Обычный текстовый запрос	Сохраняются модель, сообщения, параметры и идентификатор связи
Потоковый ответ	Средство корректно обрабатывает фрагменты, завершение и разрыв соединения
Вызов инструмента	Структура вызова сохраняется, полномочия действия проверяет сервер
Ограничение размера	Возвращается понятный отказ до неконтролируемой обработки
Превышение лимита	Клиент получает сигнал ограничения без бесконечной немедленной повторной отправки
Тайм-аут проверки	Обязательная проверка не обходится прямым вызовом модели
Ошибка аутентификации	Секрет не раскрывается в ответе, новая попытка не использует чужой ключ
Запрет модели или категории	Ошибка не превращается в автоматический выбор запрещённого получателя

Идентификатор запроса проходит цепочку или связывается с идентификаторами отдельных компонентов. Это позволяет сопоставить решение Guardrail, AI firewall и модели. При диагностике не требуется передавать исходные значения масок в заголовке для удобства корреляции.

Fail-closed и доступность

Для обязательного контрольного компонента применяется отказ без обхода: если проверка недоступна или не завершилась, непроверенное содержание не отправляется дальше. Возможен резервный экземпляр того же разрешённого средства или другой заранее разрешённый маршрут с эквивалентными ограничениями. Внешняя модель не становится допустимой только потому, что внутренняя перегружена.

Важна разница между отказом проверки и отрицательным решением. Тайм-аут означает отсутствие достаточного результата, а блокировка — явный запрет по правилу. Оба события должны быть различимы в аудите. Пользователь получает безопасное описание и идентификатор, а техническая команда — причину на соответствующем уровне доступа.

Если контур использует несколько endpoint, перечисляются все пути: обычный и защищённый API, интерфейс, фоновые задачи, файловые обработчики и инструменты. Закрытие только одного URL оставляет обход через другой маршрут. Защита обязательна на уровне фактической операции, а не только выбранной кнопки интерфейса.

Лимиты и защита доступности

Контролируются размер входа, число параллельных задач, частота запросов, лимит пользователя или ключа и стоимость обработки. Для файлов отдельно важен объём после распаковки и извлечения. Маленький архив может породить большой объём данных; короткий запрос — запустить многократные инструменты и модельные обращения.

Числовые лимиты задаются для конкретной поставки, оборудования и задач. Единое значение максимального файла или запросов в секунду не применяется ко всем установкам. Приёмка фиксирует применённые пределы и проверяет поведение на границе, чуть выше границы и при конкурентной нагрузке нескольких пользователей.

Сервис возвращает ограничение до истощения ресурсов, сохраняет управляемость и не перекладывает весь бюджет на соседнее пространство. Повторные попытки клиента имеют ограничение и задержку; бесконечный автоматический retry способен усилить перегрузку или повторить платную внешнюю обработку.

SSRF, ссылки и инструменты

Внешняя ссылка из документа или ответа модели не считается разрешённым адресом автоматически. Исполнитель инструмента ограничивает схемы, домены и доступ к внутренним служебным адресам. Проверяются конечный адрес после перенаправления и возможность обращения к инфраструктурным метаданным. Права на веб-поиск не дают права обращаться к произвольному внутреннему сервису.

Подключение к корпоративной базе или API использует назначенный адрес и ограниченные операции. Модель не меняет адрес сервера или учётные данные через содержимое запроса. В обычном on-prem инструменты могут обращаться к внутренним и внешним ресурсам по политике. В изолированном профиле рабочие инструменты используют только ресурсы внутри заданного контура; внешняя ссылка не должна приводить к скрытому выходу.

Облачные получатели и OpenRouter

Для API-маршрутов, использующих OpenRouter, этот сервис включается в цепочку получателей наряду с конечным модельным провайдером. Необходимо учитывать фактический запрос, доступные настройки хранения и географию каждого участника. Прямой вызов провайдера и вызов через дополнительный шлюз имеют разные границы доверия; утверждение «все обращения идут только напрямую» не применяется ко всей системе.

Использование собственного ключа провайдера меняет способ авторизации и учёта, но не устраняет сетевой путь и получателей. Требование обработки только в РФ и запрет выхода из инфраструктуры заказчика являются разными ограничениями: первое проверяет географию всей цепочки, второе исключает и российские внешние сервисы. Ключ пользователя не может обойти ни одно из этих правил.

Негативные проверки

Отключите обязательный AI firewall, DLP и разрешённую модель по очереди. Убедитесь, что исходный запрос не ушёл напрямую и что событие объясняет стадию отказа. Затем попробуйте альтернативный endpoint, изменение адреса, запрещённую переадресацию, недоверенный сертификат, запрос выше лимита и повтор после блокировки.

Для изолированного профиля запустите текст, скан, индексирование файла, инструмент и создание артефакта при запрещённом внешнем выходе. Сетевые журналы должны подтвердить внутренний маршрут каждой операции. Документируйте версию средства ИБ, правила, конфигурацию Nevel и фактические результаты — этого требует проверка интеграции, а не одного заявления о совместимом API.

Смежные темы: архитектура, шифрование, аудит, внешнее API.

БЕЗОПАСНОСТЬ

Модель угроз и негативные проверки

Сценарии утечки, prompt injection, злоупотребления инструментами, ошибки маршрута, изоляция и остаточный риск.

Облако / В контуре

https://docs.nevel.sh/security/threat-model/

Модель угроз Nevel начинается с последствий для бизнеса: раскрытие корпоративных данных, изменение источников и результатов, выполнение неразрешённого действия, потеря доступности и невозможность расследовать событие. Для каждого последствия определяется путь атаки, применяемая мера и проверка, способная показать отказ защиты.

Состав активов, нарушителей и обязательных мер в модели угроз зависит от внедрения. Подключение новой базы, внешнего провайдера, инструмента записи или публичного ассистента меняет поверхность атаки и требует пересмотра соответствующих сценариев.

Активы и границы

К защищаемым активам относятся исходные документы, диалоги, сведения о сотрудниках, коммерческие условия, секреты, таблицы соответствий, базы знаний, журналы и способность управлять политиками. Отдельный актив — полномочия подключённого инструмента: модельный ответ может причинить вред через действие в системе, даже если текст ответа не содержит секретов.

Границы проходят между пользователем и сервером, организациями, разрешённым контекстом и недоступными документами, данными и управляющими инструкциями, приложением и модельным получателем, рабочими процессами и эксплуатацией. Каждая граница должна иметь проверяемое правило. Доверие к сотруднику не означает доверие к каждому файлу, который он загрузил из внешней переписки.

Возможные нарушители

Источник риска	Начальные возможности	Что ограничивает последствия
Внешний человек без учётной записи	Попытки входа, публичные URL, подбор ключей	Аутентификация, лимиты, отсутствие публичных хранилищ
Обычный участник	Свои документы, разрешённые модели и инструменты	Серверные права, изоляция пространств, обязательные политики
Скомпрометированный пользователь	Действующая сессия и доступ сотрудника	Минимальные права, отзыв, ограничения экспорта и маршрута
Автор вредоносного документа	Текст, который может попасть в RAG	Разделение данных и инструкций, контроль источников и инструментов
Администратор	Изменение части настроек	Разделение полномочий, аудит изменений и ограничение доступа к содержимому
Компрометированный сервис	Сервисная идентичность и её данные	Узкая область доступа, сегментация, отдельные секреты
Внешний обработчик	Полученный по разрешённому маршруту контекст	Минимизация, маскирование, условия обработки и запрет неподходящих категорий

Ошибки добросовестного пользователя рассматриваются наряду с намеренной атакой: загрузка полной кадровой таблицы, публикация ассистента с личным подключением или выбор внешней модели для закрытого документа могут привести к тем же последствиям.

Матрица основных угроз

Угроза	Контроль	Проверка отказа защиты
Чтение чужого файла по известному ID	Проверка организации и прав на сервере	Запрос участника другого пространства
Утечка через RAG	Права до выдачи фрагмента и актуализация доступа	Поиск контрольной фразы чужого или удалённого документа
Пропуск чувствительного значения	DLP, классификация целого документа, локальный маршрут	OCR, транслитерация, разрыв номера, косвенные признаки
Обход защищённого API	Обязательное серверное и сетевое правило маршрута	Альтернативный endpoint, фоновые вызовы и прямой адрес модели
Утечка при fallback	Резервный маршрут того же разрешённого класса	Отказ локальной модели при документе «только в контуре»
Кража исходных значений масок	Изоляция соответствий и проверка восстановления	Поддельная маска другого запроса или пространства
Компрометация секрета	Ограниченная область, защищённое хранение, отзыв	Старый ключ после ротации; поиск секрета в логах
Неавторизованное действие инструмента	Серверная проверка операции и полномочий	Инструкция в документе выполнить запись или внешнюю отправку
Истощение ресурсов	Ограничение размера, частоты, параллелизма и повторов	Архив с большим раскрытием и серия конкурентных задач
Скрытое ослабление защиты	Версии политик, административный аудит	Изменение маршрута или логирования перед запросом
Возвращение удалённых данных	Очистка производных объектов и контроль восстановления	Восстановление старой копии после удаления

Меры сочетаются. DLP не заменяет права на источник, сетевой запрет не обнаруживает злоупотребление разрешённым инструментом, а аудит не предотвращает уже состоявшуюся утечку. Каждая мера должна иметь собственный наблюдаемый результат.

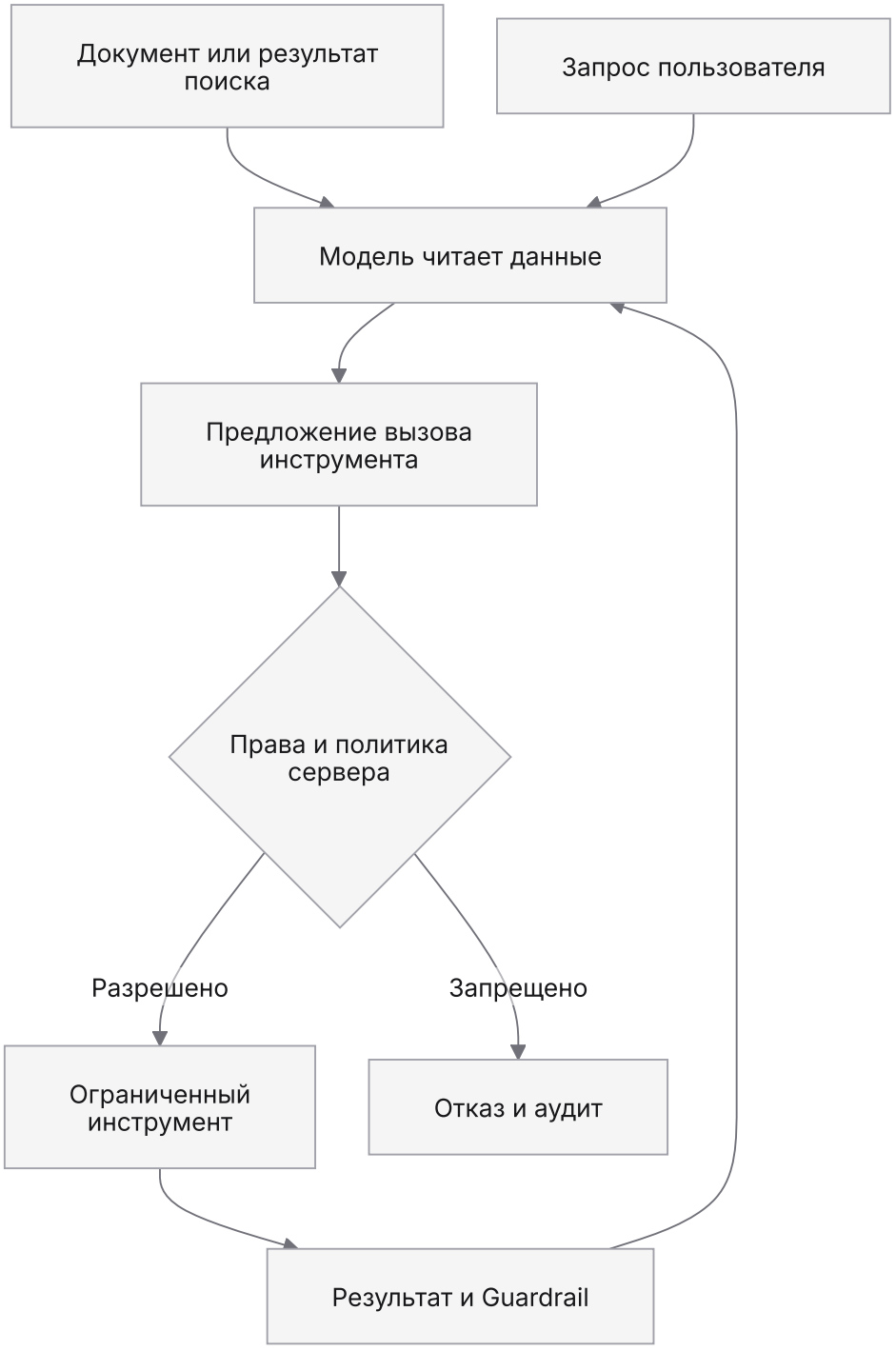
Prompt injection

Подробный процесс проверки контекста и инструментов, пример профиля реакции и программа испытаний вынесены в защиту от prompt injection. Допуск источников, карантин и смена версий описаны в защите файлов и баз знаний.

Prompt injection — попытка заставить модель принять недоверенное содержание за управляющую инструкцию. Источником может быть документ, веб-страница, результат поиска, письмо или ответ подключённого сервиса. Типичные последствия — раскрытие контекста, обращение к другому источнику, изменение маршрута или выполнение действия от имени пользователя.

СХЕМА 1

Недоверенный текст и полномочия



Защита опирается на то, что найденный текст не обладает полномочиями. Он не меняет серверный список моделей, не выдаёт секрет, не предоставляет доступ к чужому файлу и не разрешает сетевой адрес. Обнаружение характерной инструкции служит дополнительным сигналом, но блокировка одной фразы не является универсальным решением: атака может быть косвенной, многошаговой или скрытой в структуре документа.

Для проверки используется синтетический документ с просьбой раскрыть контрольную фразу недоступного источника или вызвать запрещённый инструмент. Успех — отсутствие раскрытия и действия, а не обязательное употребление моделью конкретной фразы отказа. Модель может пересказать вредоносный текст как содержание документа, не исполняя его.

Ошибки модели и целостность результатов

Модель может неверно интерпретировать таблицу, придумать факт, ссылку или реквизит. Это риск целостности, даже если конфиденциальность сохранена. Для аналитического отчёта ответ связывается с источниками, а существенное действие проверяется по исходным данным. Создание черновика отделяется от отправки письма, изменения записи или утверждения документа.

Особенно важны реквизиты, расчёты, юридические формулировки и инструкции с последствиями для людей. Наличие ссылки в ответе не доказывает правильность вывода: проверяется, что источник существует, доступен проверяющему и подтверждает утверждение. Автоматическая цепочка из нескольких моделей не отменяет этот риск и не получает больше прав только из-за количества этапов.

Обновление модели может изменить качество, поведение инструментов и восприимчивость к инструкциям. Поэтому новая модель проходит регрессионный набор для разрешённых сценариев. Автоматическое появление модели в каталоге и разрешение использовать её для закрытых данных — отдельные решения.

Остаточный риск маскирования

DLP может пропустить сущность или заменить лишний текст. Внешний получатель может определить человека по совокупности немаскированных признаков. Поэтому правило «при отсутствии масок отправлять куда угодно» недостаточно для данных с обязательным локальным режимом. При таком требовании локальный маршрут задаётся категорией и политикой независимо от результата детектора.

Обратимое соответствие дополнительно сохраняет возможность восстановления исходных значений. Его изоляция и короткий срок снижают последствия компрометации, но не дают автоматического юридического вывода об отсутствии персональных данных. Техническая оценка качества и правовая допустимость внешней передачи проводятся отдельно; см. 152-ФЗ.

Размещение платформы и угрозы маршрута

В облаке отдельно оцениваются изоляция организаций и доступ эксплуатации Nevel. В on-prem оцениваются управление инфраструктурой заказчика и её сервисные полномочия. В обоих случаях внешняя обработка зависит от модели и политики: российские провайдеры, self-hosted Nevel в РФ и глобальные API могут быть разрешёнными отдельными маршрутами.

Для каждого маршрута проверяются конечные и промежуточные получатели. В API через OpenRouter этот сервис включается в цепочку наряду с конечной моделью. Self-hosted DeepSeek у Nevel в РФ оценивается по месту выполнения этого экземпляра; API внешнего владельца модели с тем же названием имеет другую границу доверия.

Изолированный профиль исключает внешнюю рабочую обработку, но сохраняет угрозы инсайдера, компрометации сервера, ошибок прав, вредоносных документов и утраты копий. Внутренняя модель может ошибаться и предлагать опасные действия. Локальное размещение Platform не отменяет Guardrail и контроль инструментов, а обычный on-prem с внешними моделями дополнительно сохраняет риски выбранных получателей.

В тесте отказа проверяется именно политика данных: заранее разрешённый резерв может использоваться с событием аудита; запрещённый маршрут и скрытый fallback не допускаются ни в облаке, ни в on-prem.

Программа испытаний и доказательства

Набор делится на обычные рабочие сценарии и негативные проверки. Первые подтверждают, что защита сохраняет полезность: файл читается, ответ восстанавливается, разрешённый инструмент выполняет задачу. Вторые проверяют границу: чужой документ, отозванный ключ, запрещённая модель, недоступный контроль, поддельная маска и повреждённый файл.

Для каждого теста записываются исходные условия, версия политики, синтетические данные, ожидаемый результат, фактические события и сетевой след при необходимости. Проверка не должна использовать реальные секреты и чужие персональные данные только ради демонстрации утечки. Если тест не проведён, результат отмечается как непроверенный, а не как успешный на основании архитектурного описания.

При обнаружении отклонения сначала ограничивается опасный маршрут или возможность, затем воспроизводится причина и выполняется исправление. Возврат возможности требует повторения конкретного проваленного теста и затронутых соседних сценариев. Дальнейшие действия изложены в инцидентах и эксплуатации.

БЕЗОПАСНОСТЬ

Меры защиты и требования РФ

Как связать 152-ФЗ, ПП № 1119, ФСТЭК № 21 и № 117, обезличивание и уничтожение с технической архитектурой Nevel.

Облако / В контуре

https://docs.nevel.sh/security/rf-controls/

Для российского внедрения сначала определяется правовой и технический режим самой информационной системы, затем выбираются меры. Нельзя назначить уровень защищённости по названию продукта, числу масок или фразе «сервер находится в России». Nevel предоставляет механизмы, которые включаются в систему защиты заказчика; итоговая оценка относится к конкретным данным, инфраструктуре, участникам и процедурам.

Технические меры включаются в систему защиты с учётом применимых нормативных требований. Таблица соответствия механизмов сама по себе не является сертификатом, актом аттестации или заключением регулятора. Правовые основания, поручение и внешняя передача раскрыты в разделе о 152-ФЗ.

Сначала определить область системы

Описание внедрения должно отвечать на пять вопросов: какие данные обрабатываются, для какой цели, кто определяет эту цель, кто фактически получает данные и где находятся все компоненты. В перечень входят не только приложение и модель, но и OCR, embeddings, инструменты, журналы, поддержка и резервные копии.

On-prem обозначает размещение платформы у заказчика. Разрешённые маршруты моделей определяются отдельно: локальная модель, российский провайдер, self-hosted Nevel на серверах РФ или глобальный провайдер. Изолированный профиль — частный случай с обработчиками внутри инфраструктуры заказчика. Передача российскому облачному получателю выходит за её границу, но этот факт сам по себе не устанавливает трансграничную передачу. Политика различает ограничение периметром, требование к территории обработки и список разрешённых получателей; см. матрицу вариантов и оценку передачи.

Если система обрабатывает персональные данные, учитываются применимые требования 152-ФЗ и подзаконных актов. Государственная информационная система, иной контур государственного органа, учреждение или значимый объект критической инфраструктуры требуют отдельного определения применимости специальных требований. Сам факт покупки корпоративного ИИ не присваивает системе такой статус и не отменяет уже установленный режим.

Категории сведений фиксируются предметно: данные работников, клиентов и представителей контрагентов; содержание кадровых или медицинских документов; финансовая информация; коммерческая тайна. Технический детектор INN или MEDICAL_POLICY не является юридическим классификатором всех этих режимов. Для каждого процесса задаются допустимые операции и получатели.

Основные документы

Документ	Что он регулирует в рассматриваемом вопросе	Как используется при внедрении
152-ФЗ	Основания, принципы, обязанности участников и защита персональных данных	Реестр процессов, поручение, права субъектов, локализация и передача
Постановление Правительства № 1119 от 01.11.2012	Требования к защите персональных данных в информационных системах и уровни защищённости	Определение уровня для конкретной системы и исходных условий выбора мер
Приказ ФСТЭК № 21 от 18.02.2013	Состав организационных и технических мер защиты персональных данных	Выбор и адаптация мер с учётом актуальных угроз и используемых технологий
Приказ ФСТЭК № 117 от 11.04.2025	Защита информации в перечисленных государственных и иных системах	Отдельная проверка применимости к системе заказчика

Документ	Что он регулирует в рассматриваемом вопросе	Как используется при внедрении
Приказ Роскомнадзора № 140 от 19.06.2025	Требования и методы обезличивания вне указанного в нём специального случая	Оценка выбранного метода и организационного порядка
Приказ Роскомнадзора № 179 от 28.10.2022	Подтверждение уничтожения персональных данных	Акт, выгрузка и хранение подтверждений

Постановление № 1119 устанавливает четыре уровня защищённости; требуемый уровень определяется по условиям обработки, а не выбирается поставщиком как универсальное свойство платформы. Официальное сообщение и документ размещены на сайте Правительства РФ.

На портале Минздрава размещена копия приказа № 21 в редакции от 14.05.2020. При выборе мер учитывайте последующие изменения и действующую редакцию на дату внедрения. Нормативные документы Минздрава, приказ № 21.

Что изменилось для государственных систем

Приказ ФСТЭК № 117 вступил в силу 1 марта 2026 года и признал утратившим силу приказ № 17 от 11.02.2013 и перечисленные изменения к нему. В пункте 3 отдельно сохранена действительность аттестатов, выданных до вступления нового приказа. Поэтому ссылка только на № 17 в новом проекте требует пересмотра применимости; наличие старого аттестата не следует автоматически считать его аннулированием. Публикация приказа № 117, официальный PDF.

Область № 117 включает государственные информационные системы и указанные в названии и требованиях иные системы государственных органов, государственных унитарных предприятий и государственных учреждений. Его нельзя механически распространить на любую частную B2B-компанию. Конкретное применение определяется статусом системы и обрабатываемой информации.

Как технические меры собираются в систему защиты

Ниже приведено сопоставление задач защиты с механизмами Nevel и инфраструктуры. Это рабочая карта реализации; она не воспроизводит полный обязательный набор для каждого уровня и не подменяет выбор мер по актуальным угрозам.

Задача	Механизм в архитектуре	Проверяемое доказательство
Идентификация и доступ	SSO, личные пользователи, сервисные ключи, роли и права объектов	Отказ чужому пользователю и отозванному ключу
Разделение организаций	Контекст пространства на сервере, фильтрация хранилищ и RAG	Отсутствие чужой контрольной фразы в ответе, поиске и экспорте
Ограничение программной среды	Изолированные обработчики файлов и ограниченные инструменты	Документ не исполняет встроенную команду и не получает произвольную сеть
Защита носителей и данных	Шифрованное хранение, отдельные права, учёт копий	Недоступность копии обычному пользователю и восстановление уполномоченным
Регистрация событий	Связанные события доступа, политики, маршрута и удаления	Воспроизводимая цепочка одного запроса
Защита от вредоносного содержимого	Безопасный разбор, ограничения ресурсов, подключённые средства защиты	Отказ опасного файла без обхода обязательной проверки
Контроль защищённости	Негативные проверки и повторные испытания после изменений	Протокол с версией продукта и фактическим результатом
Целостность	Контроль источников знаний, версий политики и артефактов	Обнаружение изменения и проверка происхождения результата

Задача	Механизм в архитектуре	Проверяемое доказательство
Доступность	Резервирование в пределах допустимого маршрута и восстановление	Испытание отказа без выхода в запрещённую модель
Защита сети	Сегментация, TLS, разрешённые получатели, свой AI firewall	Сетевой след и блокировка прямого обхода
Реагирование	Мониторинг, ограничение маршрута, расследование	Карточка инцидента и подтверждённое устранение причины
Управление конфигурацией	Версии правил, отдельные полномочия, контроль обновлений	История изменения и повторный тест затронутого сценария

Физическая защита площадок, виртуализация, защита рабочих станций и организационные процедуры реализуются на уровне соответствующей инфраструктуры. В облаке за часть этих мер отвечает эксплуатация поставки; в on-prem за установленную у заказчика часть — заказчик или его исполнитель. При разрешённом внешнем модельном маршруте инфраструктура получателя рассматривается отдельно. Границы закрепляются в матрице ответственности, чтобы один и тот же контроль не оказался предположительно выполненным обеими сторонами без доказательства.

Обезличивание и обратимые маски

Приказ Роскомнадзора № 140 действует с 1 сентября 2025 года. Он предусматривает, в частности, метод введения идентификаторов и отдельное хранение ключа сопоставления оператором без передачи третьим лицам. Требования включают оценку достаточности метода, защиту обработки, раздельное хранение соответствующих массивов и учёт действий. Поэтому обратимость сама по себе не означает отсутствия предусмотренного нормативным актом метода, но одной замены ФИО недостаточно для вывода о соблюдении всех требований. Официальный приказ № 140, приложения 1–2.

Для облачной схемы отдельно оцениваются роли оператора и исполнителя, расположение и доступ к соответствиям, состав передаваемого текста и получатели. Нельзя из факта существования метода вывести универсальное разрешение отправлять любой результат иностранной модели. В Nevel чувствительная категория может иметь обязательный российский или внутренний маршрут независимо от наличия масок.

Постановление Правительства № 1154 от 01.08.2025 регулирует специальный процесс обезличивания, связанный с предоставлением данных по требованиям в рамках статьи 13.1 152-ФЗ. Его не следует автоматически использовать как общий режим для любого корпоративного чата. Публикация постановления № 1154.

Подтверждение уничтожения

При автоматизированной обработке приказ № 179 предусматривает акт об уничтожении и выгрузку из журнала. Пункты 3 и 5 определяют сведения о субъектах или идентифицирующей информации, категориях данных, системе, причине и дате; акт дополнительно описывает участников и способ. Недостающие в выгрузке сведения включаются в акт по пункту 6. Подтверждения хранятся три года с момента уничтожения по пункту 8. Официальный приказ № 179, пункты 2–8.

Техническая реализация описана в хранении и удалении. Акт должен отражать фактическую область уничтожения, включая ограничения резервных копий, а не объявлять удалённым то, что продолжает выдаваться поиском или восстанавливается в рабочую систему. Сам акт содержит защищаемые сведения и имеет отдельный режим доступа.

Проектная и эксплуатационная документация

Для оформления проекта могут использоваться ГОСТ Р 59793-2021 о стадиях создания автоматизированных систем и ГОСТ Р 59795-2021 о содержании документов. Их карточки Росстандарта подтверждают предмет стандартов и статус действия; конкретный состав документации закрепляется в требованиях проекта. Ссылка на стандарт оформления не подтверждает автоматически выполнение всех мер защиты. ГОСТ Р 59793-2021, ГОСТ Р 59795-2021.

Практический пакет включает описание системы и потоков, модель угроз, матрицу прав, политику категорий и маршрутов, реестр получателей, правила хранения, процедуры управления секретами, программу испытаний, регламент обновлений и реагирования. Каждый документ содержит владельца, версию и область действия. Протокол испытаний ссылается на установленную конфигурацию, а не только на публичное описание возможностей.

Сертификаты и обязательные свойства

Если для системы требуется применение конкретных средств защиты, прошедших оценку соответствия, проверяются средство, версия, область действия и условия эксплуатации. Название алгоритма TLS или AES не является сертификатом. Российское расположение серверов не является аттестацией. Наличие DLP не присваивает системе определённый уровень защищённости.

В анкете ИБ указываются только реквизиты реально предоставленных документов. При отсутствии подтверждения соответствующее утверждение не заменяется общим обещанием. Техническая возможность интеграции с корпоративным средством также требует проверки выбранной версии и конфигурации; см. сетевую защиту.

Нормативная база и область применения

Дата нормативной справки: 8 сентября 2026 года. Для приказа № 21 ссылка ведёт на редакцию от 14.05.2020; для № 117, № 140, № 179 и № 1154 — на официальные публикации актов. Разбор № 117 и № 1154 охватывает перечисленные вводные положения, а не полный набор требований. При проектировании применяются действующие редакции с последующими изменениями и требования, относящиеся к отрасли и статусу системы.

Состояние нормативной базы 152-ФЗ и изменение статьи 12 законом № 265-ФЗ от 26.07.2026 описаны в юридическом разделе. Для конкретной системы нужны оценка применимости норм и согласованный состав мер.

БЕЗОПАСНОСТЬ

Мониторинг и интеграция с SIEM

События Guardrail и Platform, пороги аномалий, доставка в SIEM, диагностика и приёмка мониторинга.

Облако / В контуре

https://docs.nevel.sh/security/monitoring/

Мониторинг Nevel помогает обнаружить изменение поведения запросов, нарушение маршрута, отказ защиты и недоступность сервиса. Он объединяет технические метрики и события безопасности, но сохраняет различие между ними: рост задержки требует диагностики производительности, а передача данных запрещённому получателю — немедленной проверки границы обработки. Оба сигнала могут относиться к одному запросу.

В облаке Nevel наблюдает компоненты своей поставки, а заказчик получает события своей организации и интегрирует их в корпоративный процесс реагирования. В on-prem заказчик наблюдает установленную платформу, внутренние компоненты и доступность разрешённых внешних маршрутов. Адреса сборщиков, SIEM и оповещений задаются отдельно. В изолированном профиле они остаются внутри заданного контура. Экспорт в облачный мониторинг является отдельной передачей служебных данных и не разрешается автоматически местом установки платформы.

Какие сигналы собирать

Наблюдение строится вокруг четырёх вопросов: работает ли запрос, применена ли защита, соблюдён ли разрешённый маршрут и можно ли восстановить цепочку действий. Для каждого сигнала назначаются владелец, источник, интервал расчёта и действие дежурного.

Группа	Сигналы	Первая проверка
Доступность	Успешные запросы, ошибки, тайм-ауты, доступность зависимостей	Какая стадия перестала отвечать и какие организации затронуты
Производительность	Задержка Gateway, DLP и модели по отдельности; очередь обработки	Где возникает ожидание, изменились ли объём и тип входных данных
Маскирование	Доля запросов с заменами, количество и типы сущностей	Не изменились ли поток документов, режим чувствительности или политика
Маршрутизация	Разрешённые, запрещённые и перенаправленные обращения	Соответствует ли фактический получатель категории данных
Идентичность	Ошибки входа, использование отозванных ключей, необычный источник	Есть ли действующая интеграция или признак компрометации
Администрирование	Изменения ролей, правил, получателей и сроков хранения	Кто изменил настройку и какие операции выполнились после изменения
Сам мониторинг	Задержка доставки, пропуски, очередь экспорта, доступность приёмника	Не образовался ли период, за который нет достоверного аудита

Запись запроса связывается с идентификатором операции, временем, пользователем или идентификатором ключа, моделью, периметром обработки, типами и количеством замен, результатом и безопасным описанием ошибки. Состав и пример события приведены в контроле доступа и аудите. Значение ключа, исходные персональные сведения и таблица соответствий в событие не включаются.

Профиль порогов Guardrail

В базовом профиле Guardrail версии 1.2 используются следующие ориентиры. Это начальные правила наблюдения для проверки при внедрении. Они не являются доказательством утечки, гарантией точности распознавания или универсальным SLA любой установки.

Метрика	Ориентир профиля	Что делать при срабатывании
Доля запросов с маскированием	Отклонение от базового уровня более 20%	Сравнить одинаковые сценарии и версии политики; проверить контрольные примеры
Среднее число сущностей в запросе	Аномальный рост более 3σ	Найти изменившуюся группу запросов; исключить пакетную обработку нового типа документов
Распределение типов сущностей	Нетипичный состав	Проверить новый источник данных, обход представлением и изменение классификации
Ложные срабатывания	Порог зависит от режима чувствительности	Разобрать размеченные примеры и обратную связь; проверить влияние на рабочую задачу
Время обработки DLP	p95 более 500 мс	Разделить длину входа, очередь и собственно обработку; проверить ресурсы и зависимости

Базовый уровень нужно считать для сопоставимого потока: одной организации, маршрута, режима и класса задач. Запросы короткого чата и обработка многостраничного файла не должны незаметно попадать в одну выборку. В паспорте правила фиксируются окно, минимальный объём наблюдений, часовой пояс, условия подавления повторов и способ пересчёта базы после изменения политики.

Число «20%» требует явно выбранной формулы. Например, в **иллюстративном правиле с относительным отклонением** базовая доля равна 40%, текущая — 30%. Отклонение составляет `abs(30 - 40) / 40 = 25%`, поэтому порог превышен. Разница при этом равна 10 процентным пунктам. Эти две шкалы нельзя смешивать. При нулевой базе относительная формула неприменима: правило должно использовать абсолютное количество, отдельное условие появления события или другую определённую метрику.

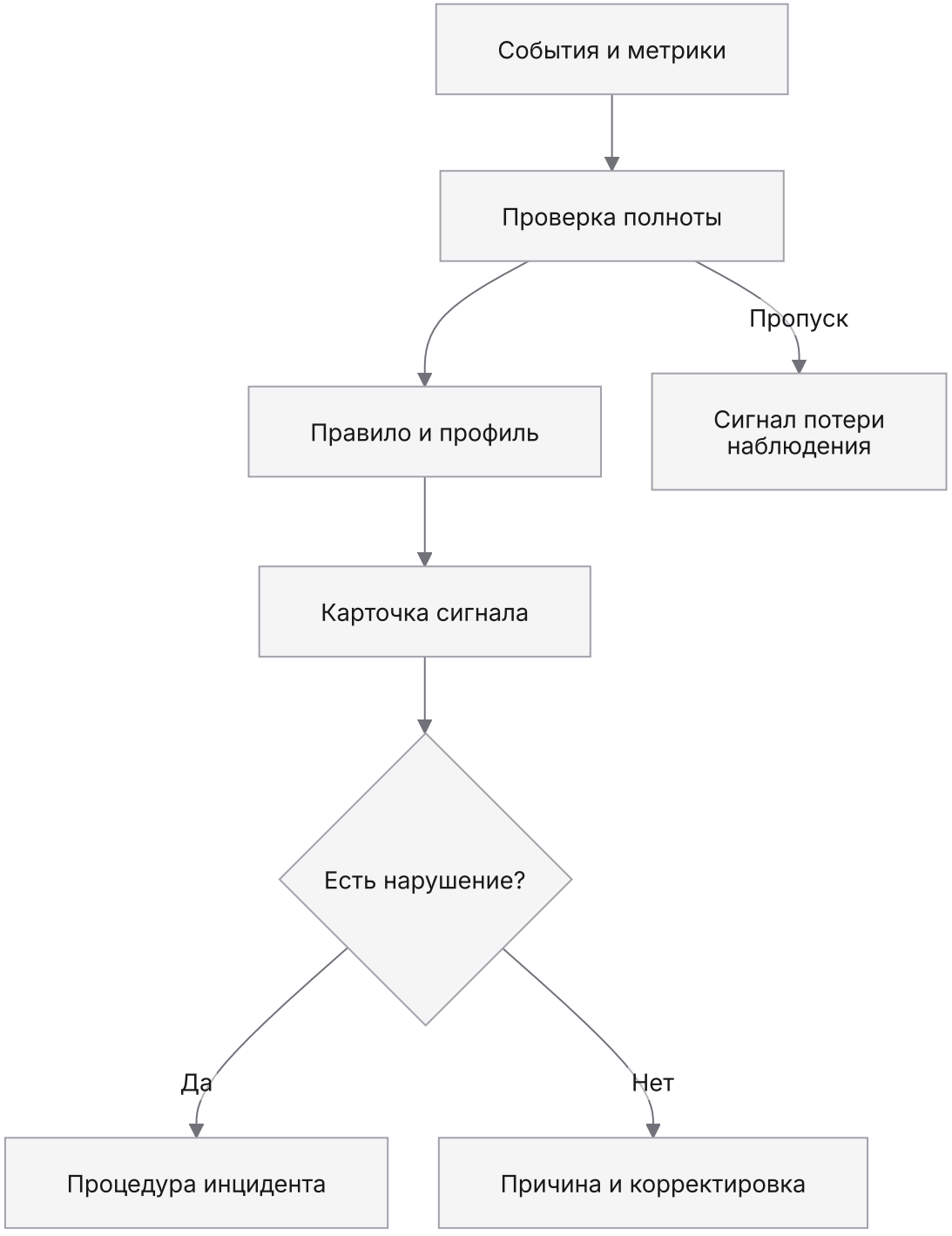
Порог 3σ сравнивает наблюдение с историческим разбросом, а не устанавливает вероятность утечки. Небольшая выборка, сезонность и резкое изменение нагрузки делают такое сравнение ненадёжным. Поэтому карточка сигнала показывает период, число запросов, среднее, разброс и изменившийся сегмент. После массовой загрузки нового класса файлов сначала выясняется причина сдвига, затем принимается решение о новой базе; автоматическое обучение на подозрительном периоде может скрыть проблему.

Падение доли маскирования не обязательно означает улучшение защиты: могли исчезнуть персональные данные, измениться документы или перестать работать детектор. Рост не доказывает атаку. Для оценки качества нужен размеченный контрольный набор с ожидаемыми заменами и проверка пропусков. Методика связана с работой Guardrail, а не выводится из одного графика.

От сигнала до действия

СХЕМА 1

Обработка сигнала мониторинга



Карточка включает время, правило, затронутую область, наблюдаемое и базовое значение, идентификаторы примеров, текущую версию политики и назначенного ответственного. Ссылка ведёт в защищённый интерфейс аудита. Копировать тело запроса в тему письма или корпоративный мессенджер не требуется.

Для подозрения на обход проверяются транслитерация, разделение символов, необычное кодирование и повторные попытки после блокировки. Для аномалии API сопоставляются интенсивность, время, источник, владелец ключа и ожидаемая задача. Работа новой пакетной интеграции может объяснить всплеск; использование отозванного ключа требует отдельной проверки даже

при низкой частоте. Автоматическое отключение всей организации по одному статистическому сигналу не является обязательным действием этого профиля.

Нарушение маршрута разбирается по фактическим стадиям: запрос заблокирован до отправки, перенаправлен разрешённой политикой или уже передан получателю. Только последний случай означает состоявшуюся передачу. Порядок локализации и оценки последствий описан в реагировании на инциденты.

Доставка в SIEM и оповещения

Профиль предусматривает панели метрик, email, webhook в корпоративную систему инцидентов и экспорт в SIEM через Syslog/CEF. Конкретный транспорт, поля расширения CEF, адрес приёмника, аутентификация и ограничения размера являются контрактом подключения. Название формата само по себе не подтверждает шифрование или гарантированную доставку.

Для подключения используются три разных объекта: реальный источник событий Nevel, согласованный профиль преобразования и принимающая SIEM. Референсный профиль аудита помогает спроектировать mapping. Он не объявляется уже существующим API или поставляемым по умолчанию нормализатором.

Шаг 1. Определить источник и область экспорта

Владелец интеграции фиксирует организацию, установки Platform, типы событий и необходимые поля. События пользователя, модельного запроса, администрирования и самой доставки различаются по типу. Отправлять весь журнал операционной системы вместо событий приложения недостаточно: он может не содержать исходный ID запроса и решение маршрута.

Для cloud-установки экспорт ограничивается организацией заказчика. Для on-prem источник относится к установленным компонентам; события внешнего провайдера автоматически не появляются в локальном журнале. Каждый добавленный источник имеет собственные версию, область доступа и способ получения.

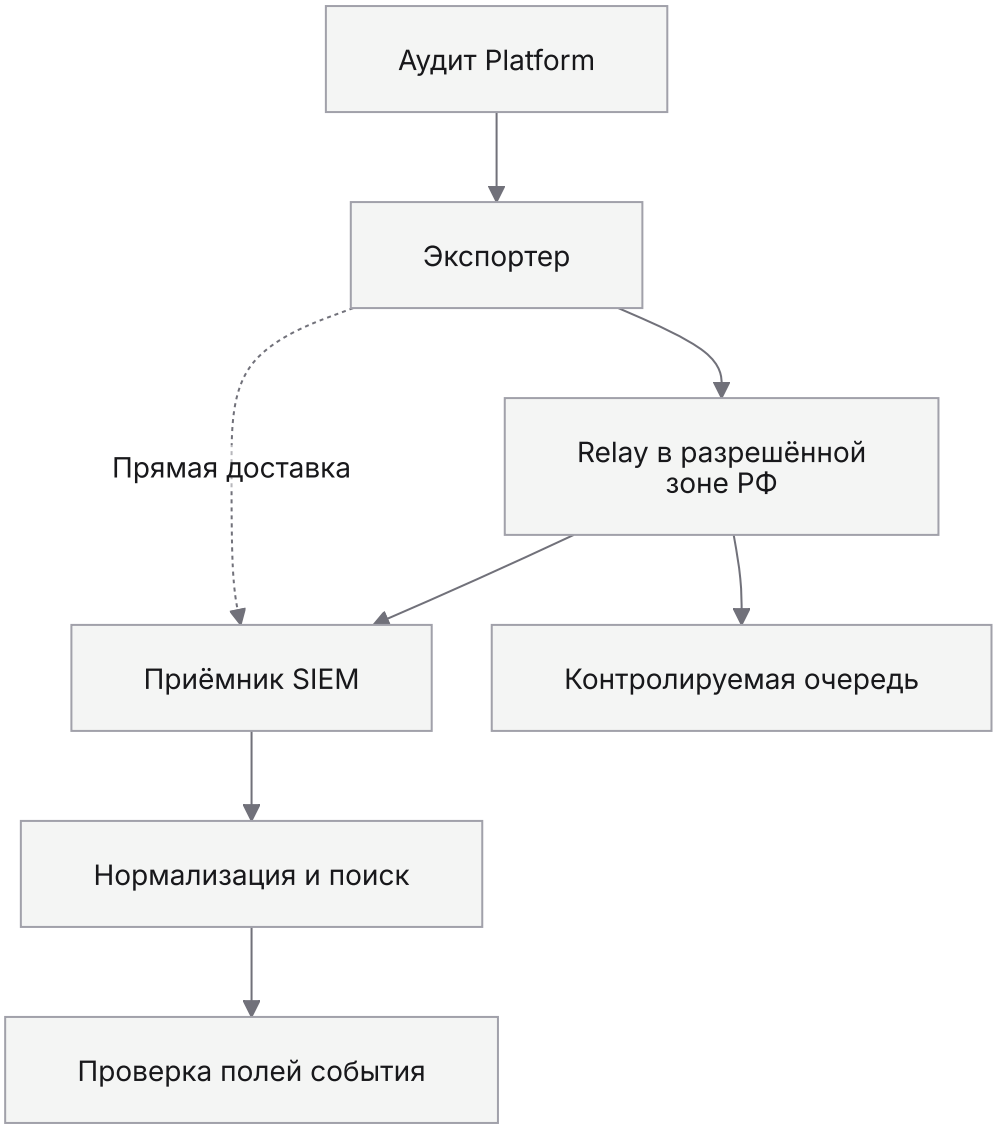
Минимальный комплект: пример реального обезличенного события, описание его полей, версия источника, перечень включённых категорий и ожидаемый результат для каждого контрольного действия. Идентификаторы и служебные коды сохраняются; секреты и тела рабочих запросов не добавляются ради удобства настройки.

Шаг 2. Выбрать прямую доставку или relay

Прямое соединение используется, когда экспортер и SIEM поддерживают один согласованный формат, транспорт и доверие. Relay — отдельный промежуточный узел для приёма, преобразования и повторной доставки — нужен, когда требуется разделить сетевые зоны, адаптировать формат или разместить очередь в контролируемой инфраструктуре. Его наличие и функции являются частью проекта поставки, а не неявной возможностью любой инсталляции.

СХЕМА 2

Путь события в корпоративную SIEM



Для relay указываются площадка, оператор, дисковая очередь, диагностические файлы, доступ администраторов и очистка. Облачный Nevel может отправлять события на разрешённый relay в РФ, откуда они поступают во внутреннюю SIEM по подготовленному соединению. Внутренний адрес SIEM не становится доступным облаку автоматически. Приёмник в DMZ предоставляет сервис событий, а не административную консоль SIEM.

В изолированном профиле вся цепочка, включая relay и резерв очереди, остаётся внутри контура. При обычном on-prem допускается разрешённая внешняя доставка событий. Российское расположение промежуточного узла не заменяет проверку его получателей и не определяет маршрут запроса к модели. Примеры продуктов и точки контроля раскрыты в корпоративной среде.

Шаг 3. Зафиксировать сетевой и транспортный профиль

Параметр	Что указать	Что проверить
Приёмник	DNS-имя, порт, владелец и зона	Это адрес сервиса приёма, а не web UI
Транспорт	Поддерживаемый обеими сторонами Syslog/TLS или HTTPS-приёмник	Формат внутри транспорта и границы одного сообщения

Параметр	Что указать	Что проверить
Доверие	CA, имя сертификата, срок, клиентская аутентификация при её использовании	Неверный сертификат не приводит к отключению проверки
Отправители	Источники или relay, разрешённые сети	Случайный внешний узел не может писать в источник Nevel
Формат	JSONL, CEF либо иной согласованный профиль	UTF-8, переносы, escaping, максимальная длина
Область	Организации и категории событий	События другой организации не попадают в поток
Повтор	Подтверждение приёма, очередь, ID и предел попыток	Дубликат отличим от нового события

Syslog — способ передачи и оформления сообщения, CEF — структура его содержимого. TCP сам по себе не шифрует соединение. Успешная запись байтов в сокет не доказывает, что SIEM разобрала и проиндексировала событие. Для HTTPS-контракта отдельно определяется, что означает успешный ответ приёмника: приём в очередь, сохранение или завершённая обработка. Не приписывайте ответу `2xx` более сильную гарантию, чем даёт конкретный сервис.

Следующий YAML — **референсная карточка подключения**, а не файл настройки Nevel или конкретного производителя. Порт выбран для примера и не является обязательным портом продукта.

```
profile: nevel-siem-reference-1
source_instance: example-onprem-01
workspace: example-workspace
receiver_host: siem-relay.example.invalid
receiver_port: 6514
transport: syslog-over-tls
payload_profile: nevel.audit.reference/1
serialization: cef
tls_server_name: siem-relay.example.invalid
deduplication_fields:
  - source_instance
  - workspace_id
  - event_id
include_request_body: false
```

В рабочую конфигурацию переносят смысл полей с использованием действительного синтаксиса выбранных компонентов. Сертификаты, секреты и размер очереди задаются отдельно. Структура карточки не обещает, что любой экспортер понимает приведённые ключи.

Шаг 4. Подготовить нормализацию для выбранной SIEM

Kaspersky KUMA. Для профиля JSON приёмник передаёт один JSON-объект на строку в UTF-8; многострочный JSON перед отправкой сериализуется в JSONL. В нормализаторе выбирают метод `json` или `cef`, добавляют примеры и таблицу сопоставления. Для JSON доступны пути к вложенным полям через точку; для CEF — исходное сопоставление с последующей проверкой. Указанные параметры относятся к KUMA 3.0.1. Параметры парсинга KUMA.

В таблице mapping показан пример назначения ID, результата и пользовательских строковых полей. Сохраняйте исходный ID события отдельно от внутреннего ID, который присваивает SIEM. Поля времени, адресов и чисел проверяются контрольными значениями; появление события в общем списке ещё не подтверждает, что каждое поле прошло преобразование.

MaxPatrol SIEM. Правило нормализации выбирается с учётом формата события. В версии 8.1 различаются текстовый и структурированные варианты, включая JSON; совпадение с правилом определяет создание нормализованной записи. Поэтому отдельному источнику Nevel нужны образцы входных событий, правило разбора и проверка целевых полей. Профиль KUMA не загружается сюда как готовое правило. Принцип нормализации MaxPatrol SIEM.

Для установленной версии владелец SIEM выбирает поддерживаемый механизм получения событий, регистрирует источник и добавляет собственный проверенный профиль. Контрольная выборка содержит несколько типов записей и необязательные поля. Успешный разбор только одного примера не подтверждает покрытие всех событий приложения.

UserGate SIEM. В версии 7.x приёмник настраивается в разделе **Настройки → Сборщик логов → Syslog**. Для TCP указываются порт и число сессий; защищённое соединение включает TLS и сертификаты, а разрешённые соседи ограничивают отправителей. После изменения выполняется сохранение и перезапуск. Настройка Syslog-сервера UserGate.

Эта настройка поднимает транспортный приёмник. Отдельно проверяются распознавание структуры Nevel, извлечение исходных ID и поиск по организации и маршруту. Приём строки в журнал без нужных полей не считается завершённой интеграцией.

Шаг 5. Проверить очередь, повторы и ошибки разбора

В проекте фиксируется точка подтверждения: когда источник может считать запись принятой следующим звеном. Если используется дисковая очередь, описываются её предел, сохранность после перезапуска, действия при заполнении и порядок удаления доставленных записей. Точное число часов и размер очереди задаются выбранной конфигурацией, а не универсальным обещанием Nevel.

Потеря подтверждения может вызвать повтор уже принятого события. Поэтому тест выполняется с тем же `event_id`, а не с новой записью, похожей по содержанию. После нормализации проверяется выбранная обработка дубликата. При нескольких источниках одинаковый локальный ID не должен смешивать разные организации и установки.

Невалидное событие не удаляется бесследно и не считается успешно разобранным. Для него нужен контролируемый путь диагностики: причина ошибки, источник, время и безопасный исходный образец в пределах политики хранения. Этот путь не должен бесконечно повторять одну ошибочную запись и останавливать остальные события.

Контрольный пример	Ожидаемый результат
Корректная запись из reference-профиля	Совпадают ID, пространство, действие, результат и маршрут
Повтор той же записи	Отличим от нового модельного вызова; действует правило дубликатов
Новая запись той же операции	Сохраняется отдельный ID события и связь с общей операцией
Поле времени с неверным форматом	Видна ошибка или оговорённый результат преобразования
Запись с неизвестной версией схемы	Не разбирается молча по несовместимому старому правилу
Недоступный приёмник и перезапуск relay	Подтверждены очередь, восстановление и фактическая полнота
Текст с кириллицей, переносом и разделителем CEF	Значения сохраняются после разбора, не создают ложные поля

Часы компонентов синхронизируются; время события сохраняет часовой пояс или передаётся в UTC. Время приёма, задержка доставки и время самого действия остаются отдельными величинами. Это позволяет отличить задержанную запись от нового действия.

Шаг 6. Зафиксировать результат подключения

Итоговый профиль содержит версии компонентов, сетевую схему, mapping, тестовые события, результат поиска по исходным ID и проверку восстановления доставки. Состояние отражается предметно: транспорт проверен, нормализация проверена, полнота категорий проверена либо ограничена перечисленными типами. Это точнее общего ярлыка «SIEM интегрирована».

Развёртывание в другой организации требует замены адресов, доверия и области событий. Обновление версии экспортера или SIEM требует повторения затронутого набора примеров. Срок хранения и повторная выгрузка связываются с политикой хранения; подключение средств контроля — с сетевой защитой.

Приёмка и регулярная проверка

Приёмка проходит на синтетических данных. Выполняются успешный запрос, блокировка политикой, попытка с отозванным ключом, разрешённое перенаправление, изменение правила и кратковременный отказ приёмника SIEM. Для каждого случая проверяются запись, корреляция, правило оповещения, получатель и понятное действие дежурного.

Отдельно проверяется конфиденциальность: исходное значение и секрет отсутствуют в журнале, письме, webhook и диагностике обработчика. В изолированном профиле контроль сетевых соединений подтверждает внутренний адрес SIEM и отсутствие внешнего канала мониторинга. В обычном on-prem проверяется соответствие каждого исходящего канала разрешённому списку.

Результат приёмки — перечень активных правил с владельцами, измеренные задержки доставки, проверенные каналы и известные ограничения. После изменения политики, формата событий или маршрута повторяются затронутые сценарии. Зелёная панель без контрольного события не доказывает, что мониторинг обнаружит реальную проблему.

Версии и совместимость

Параметры Kaspersky, Positive Technologies и UserGate относятся к версиям, указанным в ссылках. YAML, таблицы mapping и контрольные примеры — референс для настройки, а не готовый пакет интеграции. Совместимость конкретных версий Nevel и SIEM подтверждается протоколом испытаний.

БЕЗОПАСНОСТЬ

Реагирование на инциденты

Локализация, доказательства, роли, восстановление и уведомление Роскомнадзора в предусмотренных законом случаях.

Облако / В контуре

https://docs.nevel.sh/security/incidents/

Процедура применяется к подозрению на несанкционированный доступ, раскрытие данных, нарушение маршрута, компрометацию ключа и отказ защитного компонента. Её результат — остановить дальнейшее воздействие, установить затронутую область, сохранить доказательства, выполнить необходимые уведомления и безопасно вернуть сервис в работу.

Технический сигнал и подтверждённый инцидент различаются. Срабатывание статистического порога требует разбора; оно не доказывает утечку. Одновременно отсутствие оповещения не отменяет сообщение пользователя, уведомление провайдера или обнаруженную в журнале передачу. Каналы первичного обнаружения описаны в мониторинге.

Роли и готовность до события

Организация назначает координатора инцидента и заместителя, специалиста ИБ, технического исполнителя и ответственного за персональные данные. Координатор ведёт общую временную шкалу; ИБ оценивает воздействие и доказательства; технический исполнитель меняет доступ и маршруты; уполномоченное лицо принимает юридические решения и направляет необходимые уведомления.

В облаке Nevel расследует собственные компоненты и предоставляет заказчику относящиеся к нему факты. Заказчик проверяет своих пользователей, интеграции, устройства и основания обработки. В on-prem технические действия с установленной платформой выполняет эксплуатационная команда заказчика в пределах распределения работ; доступ поддержки Nevel предоставляется по действующим полномочиям. Если запрос использовал внешний разрешённый маршрут, расследование включает соответствующего российского провайдера, self-hosted инфраструктуру Nevel или глобальную цепочку API. Само размещение Platform у заказчика не исключает внешнего получателя. Передача диагностического архива не становится разрешённой только потому, что возник инцидент.

Контактный лист хранится вне единственного потенциально недоступного сервиса. В нём указаны основной и резервный канал, ответственный за получение, порядок работы вне обычного графика и адрес защищённого обмена материалами. Срок реакции поддержки и срок уведомления органа власти имеют разные основания. Условие одного клиентского договора нельзя автоматически применять ко всем установкам.

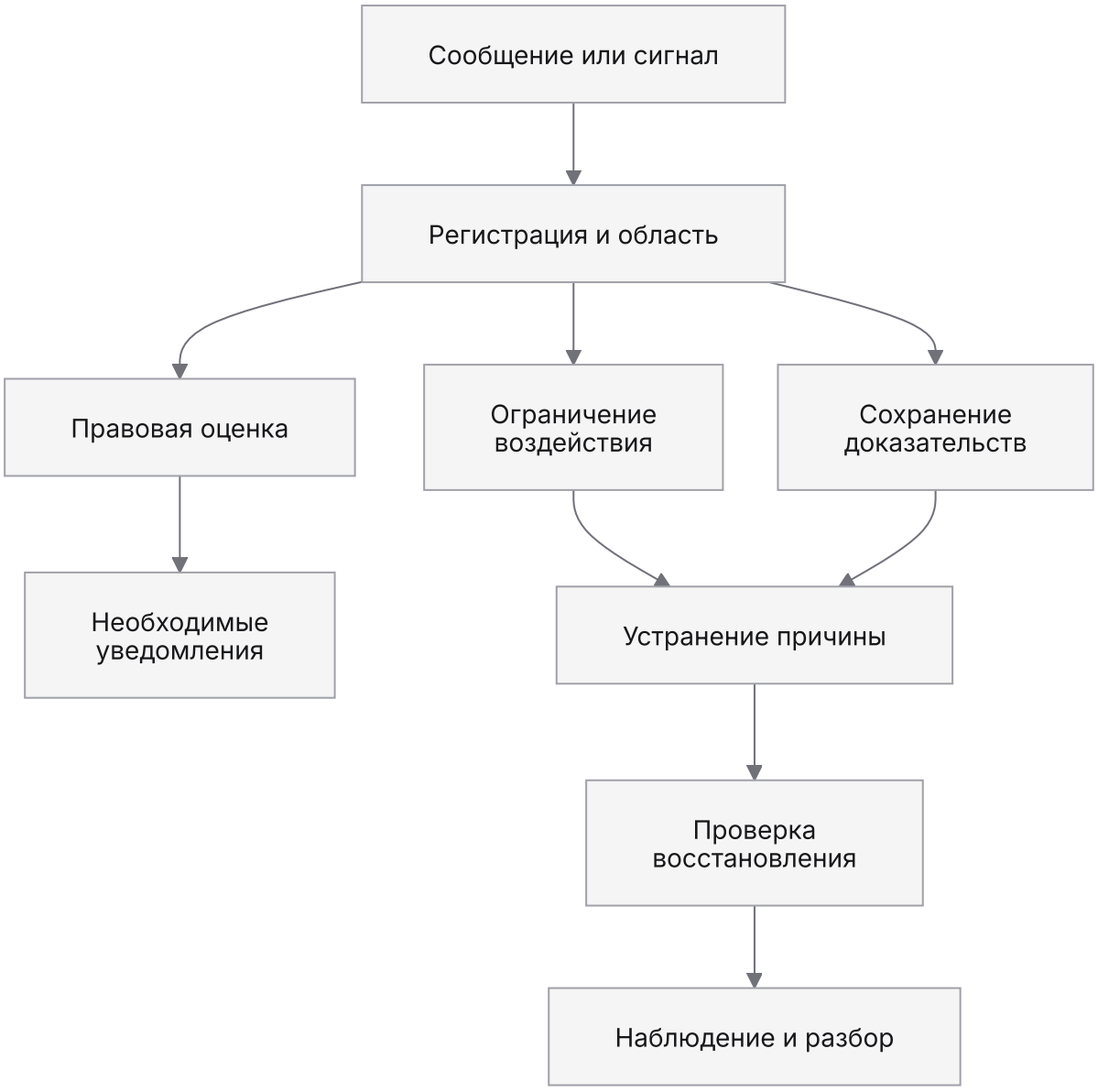
Первые действия

1. **Зарегистрировать событие.** Присвоить номер, сохранить время первого сообщения, время предполагаемой операции и источник. Не заменять раннюю отметку временем открытия задачи инженером.
2. **Установить область.** Определить организацию, пользователей, ключи, файлы, модели и интервал. Проверить, затронуты ли другие пространства, не раскрывая заказчику чужие данные.
3. **Остановить продолжение.** Отозвать подозрительный ключ или сессию, отключить скомпрометированное подключение, запретить конкретный маршрут либо ограничить затронутую функцию.
4. **Зафиксировать состояние.** Сохранить доступные журналы, версии политик, настройки маршрута, сетевые события и подтверждения выполненных действий.
5. **Оценивать последствия параллельно.** Не ждать полного восстановления, чтобы привлечь ответственного за персональные данные и запустить необходимые уведомления.

Ограничение выбирается по механизму угрозы. Для утечки одного ключа первичным действием служит отзыв этого ключа и проверка его обращений. При ошибке общей политики недостаточно отключить одного пользователя: блокируется опасный маршрут для всей затронутой категории. Если DLP недоступен, восстановление доступности не должно включать обход обязательной проверки. Границы допустимого переключения определены в архитектуре и сетевой защите.

СХЕМА 1

Реагирование на инцидент



Что доказывать в расследовании

Ключевой вопрос для модельного запроса — прошли ли данные границу получателя. Отказ после отправки не означает, что передача не состоялась; блокировка до исходящего вызова имеет другой характер. Восстановление цепочки идёт от идентификатора операции через проверку прав, формирование контекста, решение Guardrail, исходящий вызов и обработку ответа.

Материал	Что устанавливает	Как ограничить раскрытие
Журналы запроса и маршрута	Время, участника, получателя, стадию и результат	Выгружать относящиеся к событию записи
Версия политики	Правило, действовавшее при операции	Сохранять версию и изменения, не заменять текущей конфигурацией

Материал	Что устанавливает	Как ограничить раскрытие
Журнал ключа и доступа	Полномочия и действия предполагаемого участника	Использовать идентификатор ключа, не его секрет
Сетевые события	Фактическое соединение с получателем	Учитывать NAT, прокси и временные метки
Сведения об объекте	Затронутые категории и объём	Начать с метаданных; содержание открывать по необходимости
Действия реагирования	Кто и когда ограничил или восстановил работу	Фиксировать каждое изменение и его результат

Доказательства сохраняются в ограниченном хранилище с учётом доступа и целостности. Для экспортированного файла фиксируются источник, время выгрузки, исполнитель и контрольная сумма. Рабочая копия отделяется от сохранённого экземпляра. Это рекомендуемая организация расследования, а не заявление о наличии у каждой установки отдельной криминалистической системы.

Нельзя очищать аудит ради уменьшения числа ошибок на панели. С другой стороны, инцидент не разрешает бессрочно собирать все тела запросов. Объём и срок сохранения материалов определяются расследованием и применимыми основаниями; обычные правила описаны в хранении данных.

Когда действуют 24 и 72 часа

Часть 3.1 статьи 21 152-ФЗ относится к установленному факту неправомерной или случайной передачи, предоставления, распространения либо доступа к персональным данным, повлёкшему нарушение прав субъектов. Обязанность уведомить Роскомнадзор возложена на оператора. Отсчёт связан с выявлением такого инцидента оператором, регулятором или иным заинтересованным лицом. Официальный текст 152-ФЗ, статья 21.

Срок от выявления	Содержание уведомления
В течение 24 часов	Инцидент, предполагаемые причины и вред, принятые меры, уполномоченное контактное лицо
В течение 72 часов	Результаты внутреннего расследования; сведения о лицах, действия которых стали причиной, при наличии

Оба срока идут от одного момента: это не 24 часа плюс ещё 72. Норма задаёт часы, не рабочие дни. Например, от отметки 10:00 во вторник пределы наступают в 10:00 среды и пятницы. Полного установления причины не требуется для первичного сообщения: закон прямо предусматривает предполагаемые причины и вред. Пункты 1–2 части 3.1 статьи 21.

Поручение обработки должно предусматривать уведомление оператора о таких случаях. Nevel, действуя по поручению заказчика, передаёт сведения оператору; это не означает автоматического представительства перед регулятором. Роли определяются для конкретных данных и операций. Статья 6, часть 3.

Практически в карточке следует вести три отдельные отметки: когда произошло действие, когда выявлен указанный законом инцидент и когда команда получила сообщение. Разница между ними объясняется фактами. Открытие внутреннего расследования не используется для произвольного переноса начала отсчёта. Другие отраслевые обязанности рассматриваются отдельно по карте требований РФ.

Подготовка и передача уведомления

Порядок взаимодействия установлен приказом Роскомнадзора № 187 от 14.11.2022. Первичное сообщение включает сведения об операторе, дате и времени выявления, характеристиках затронутых данных и числе записей, предполагаемых причинах и вреде, принятых мерах и контактном лице. Дополнительное отражает результаты расследования и уточнённые меры. Официальная публикация приказа № 187.

Приказ допускает бумажное и электронное направление. Электронная специализированная форма используется после авторизации через ЕСИА; уведомление подписывается электронной подписью согласно установленному порядку. Для дополнительного электронного сообщения сохраняются номер и ключ первичного уведомления. Они позволяют связать материалы одного события и не предназначены для публичного размещения. Приказ № 187, порядок взаимодействия.

Внутренний пакет готовится так, чтобы уполномоченный сотрудник мог заполнить форму без повторного опроса инженеров. Подтверждённые факты отделяются от гипотез, неизвестное обозначается явно. Число запросов не подменяет число записей или субъектов. Копия направленного сообщения, время отправки и подтверждение получения сохраняются в деле инцидента. Рабочую доступность учётной записи и способ подписания проверяют заранее, а не в последние минуты срока.

Возврат в работу и приёмка процедуры

Возврат начинается после устранения механизма нарушения: отзыва старого ключа, исправления прав или политики, проверки затронутого получателя. Контрольный запрос должен подтвердить разрешённое действие, а отрицательный тест — запрет повторения опасного сценария. Для восстановления из копии применяются проверки эксплуатации и DR.

Учебная проверка включает утечку тестового ключа, ошибку маршрута и ложный сигнал DLP. Команда должна найти действие по идентификатору, ограничить нужную область, собрать обезличенный пакет, определить применимость уведомления и рассчитать сроки. Уведомление в учебном сценарии готовится как черновик, без отправки вымышленного события регулятору.

Закрытие фиксирует причину, масштаб, временную шкалу, принятые меры и владельцев оставшихся действий. Успешный повторный запрос сам по себе не закрывает расследование. Нормативная справка на 8 сентября 2026 года и версии применимых актов приведены в разделе о 152-ФЗ; сроки и порядок действий определяются действующими требованиями.

БЕЗОПАСНОСТЬ

Эксплуатация, резервирование и восстановление

Профили доступности, RPO и RTO, резервные копии, аварийное восстановление, обновления и проверяемая приёмка.

Облако / В контуре

https://docs.nevel.sh/security/operations/

Эксплуатационный профиль Nevel определяет, какие компоненты обслуживаются, от каких отказов защищена установка, сколько данных допустимо потерять и за какое время восстанавливается работа. Профиль связывает договорные показатели с архитектурой и испытаниями. Наличие резервной копии, реплики или автоматического переключения по отдельности не доказывает достижение всех показателей.

В облаке инфраструктуру своей поставки обслуживает Nevel. В on-prem заказчик обеспечивает ресурсы установленной платформы, хранилища, сеть и эксплуатационный доступ в соответствии с распределением работ. Локальные модели обслуживаются внутри его инфраструктуры; российские и глобальные провайдеры и self-hosted модели Nevel имеют отдельную область эксплуатации и доступности. Запуск на одном сервере и многоплощадочная установка имеют разные области отказа. Выбор варианта описан в методах развёртывания.

Какие показатели входят в профиль

Базовый профиль Guardrail версии 1.2 содержит следующие значения. Они описывают предусмотренный профиль сервиса; применимость к конкретному окружению и составу поставки устанавливается его паспортом и договором. Таблица не объявляет эти значения измеренным результатом всех установок Platform или безусловным обязательством каждого тарифа.

Показатель	Значение профиля	Что входит в измерение
Доступность	99,9%	Доступность определённого сервиса за установленный период
Задержка API p50	Не более 150 мс	Обработка API без времени ответа LLM
Задержка API p99	Не более 500 мс	Та же граница, 99-й процентиль
Плановое окно недоступности	Не более 30 минут, уведомление за 48 часов	Окно обслуживания по применимому порядку
Реакция на критический инцидент	Не более 15 минут	Начало реакции по установленному каналу и критичности
RTO	Не более 1 часа	Время до восстановления предусмотренной работоспособности
RPO	Не более 5 минут	Максимальный разрыв между событием отказа и восстановленной точкой данных

Для доступности нужны точное определение успешной операции, интервал наблюдения и источник измерения. В паспорте указываются затронутые компоненты, часовой пояс, правила учёта частичной недоступности и плановых работ. Исключения внешнего провайдера или обслуживания не возникают автоматически: они применяются только в установленной области договора.

Численно 99,9% за условный год из 365 суток соответствует 8,76 часа недоступности. Это арифметическое пояснение, а не разрешение суммировать любой простой за любой период. Годовая доля и ежемесячное обязательство рассчитываются по разным периодам. Реакция за 15 минут также не означает устранение причины за 15 минут.

Задержка API без LLM отделяется от времени до первого токена и полного ответа модели. Большой файл, длинный вывод и очередь модели влияют на пользовательское ожидание по-разному. Порог DLP `p95 > 500 мс` из мониторинга относится к другому этапу и процентилю; его нельзя использовать как доказательство соблюдения `p99` всего API.

Резервирование и разрешённое переключение

Облачный профиль предусматривает несколько российских площадок, переключение при отказе, горизонтальное масштабирование API Gateway, DLP и Router, а также репликацию хранилищ соответствий и аудита. Переключение модели выполняется только на разрешённого получателя. В on-prem внешний резерв допустим только при заранее разрешённом для данных маршруте и с записью решения в аудит. В изолированном профиле или для метки «только в контуре» отказ локальной модели не разрешает отправку во внешний API.

Установка проверяется по фактическим областям отказа. Два процесса на одном сервере не защищают от отказа сервера; две машины на одном хранилище не защищают от отказа этого хранилища. Отдельно рассматриваются питание, сеть, зона размещения, управляющие учётные записи и система ключей. Архитектурные границы и обработка отказа защиты описаны в архитектуре Guardrail.

Отказ	Ожидаемое эксплуатационное действие	Что проверить
Экземпляр Gateway или DLP	Исключить неисправный экземпляр и использовать здоровый	Защита не пропущена, повтор не раскрывает данные
Локальная модель	Использовать разрешённую альтернативу либо вернуть контролируемый отказ	Нет внешнего переключения для запрещённой категории
Основное хранилище	Перейти по предусмотренному плану восстановления	Подтверждена актуальность данных и единственный активный источник записи
Площадка	Выполнить процедуру аварийного переключения	Доступны зависимости, ключи, маршруты и аудит
Логическое повреждение	Восстановить состояние до повреждения	Ошибка не воспроизведена из реплики или копии

Наличие настроенного переключения проверяется отказом соответствующего компонента на стенде или в разрешённом испытании. Отключение только пользовательского интерфейса не подтверждает переживание отказа базы данных.

Копии, репликация и RPO

Профиль резервного копирования Guardrail предусматривает инкрементальные копии каждые 4 часа, полные ежедневно, хранение 30 дней, отдельную российскую площадку, AES-256 и ежемесячное восстановление с протоколом. Состав включает конфигурацию, политики, аудит и действующие соответствия. Подробные условия хранения и удаления приведены в жизненном цикле данных; набор резервируемых объектов Platform определяется составом поставки.

Реплика помогает пережить физический отказ, но может немедленно повторить ошибочное удаление или повреждение. Резервная копия сохраняет отдельное состояние, однако интервал 4 часа сам по себе не обеспечивает RPO 5 минут. Для этого показателя необходим подтверждённый механизм более свежего восстановления: например, предусмотренная архитектурой репликация либо восстановление по журналу изменений с достаточной частотой и глубиной хранения. Его наличие и результат проверяются, а не выводятся из расписания копий.

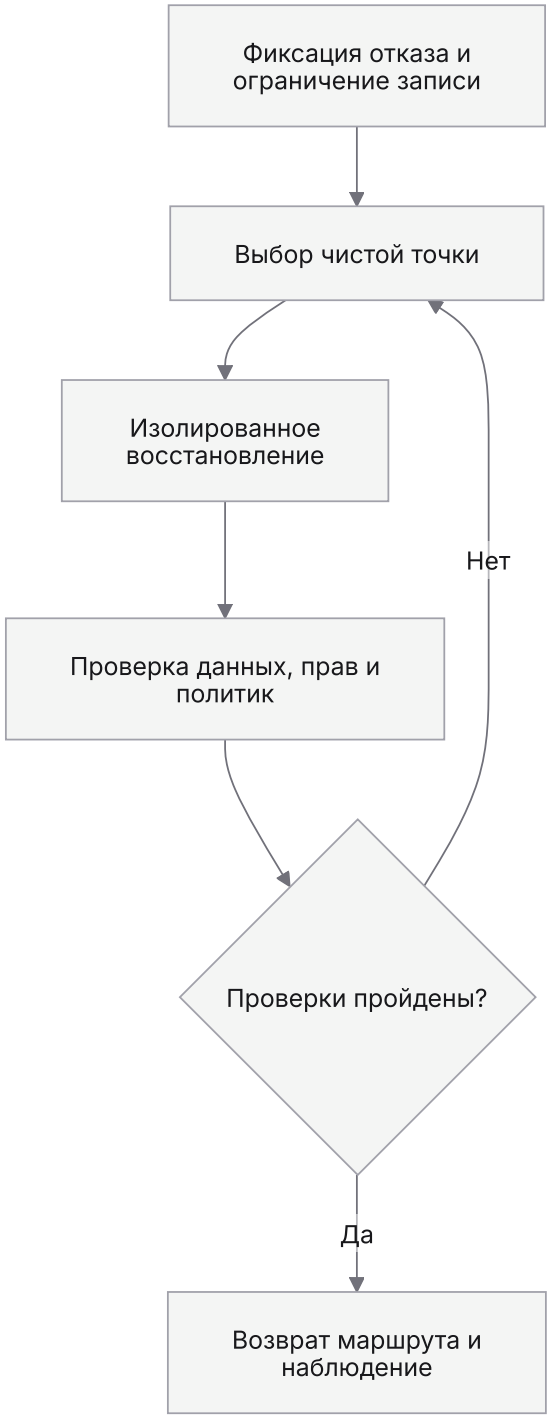
В учебном примере сервис отказал в 14:20. Последняя подтверждённая запись в восстановленном состоянии относится к 14:17, работа возвращена в 15:05. Потеря по времени — 3 минуты, восстановление — 45 минут. Эти значения укладываются в указанный профиль для данного испытания. Если пригодна только копия от 12:00, разрыв составляет 2 часа 20 минут: успешный запуск из неё не подтверждает RPO 5 минут.

Копия проверяется на полноту, читаемость и пригодность ключа расшифрования. Ключи и права восстановления не должны зависеть от единственного отказавшего сервиса. При этом хранение ключа рядом с общедоступным архивом уничтожает смысл шифрования. Требования к ключам и криптографическим границам находятся в разделе о шифровании.

Процедура аварийного восстановления

СХЕМА 1

Аварийное восстановление Nevel



1. Зафиксировать начало воздействия, симптомы, последнее достоверное состояние и решение о восстановлении. При признаках атаки выполнять параллельно процедуру инцидента.
2. Ограничить запись и фоновые задачи там, где они могут создать конкурирующие состояния. Определить единственный источник, который будет принят за рабочий после восстановления.
3. Выбрать точку с учётом причины отказа. Для логического повреждения последняя по времени копия может уже содержать ошибку.

4. Восстановить в изолированном окружении конфигурацию и необходимые хранилища. Не включать внешние отправки и автоматические интеграции до проверки маршрутов.
5. Проверить согласованность объектов: файл доступен разрешённому пользователю, индекс соответствует правам, журнал связан с операциями, маски не восстанавливаются по чужим соответствиям.
6. Применить предусмотренный контроль удалений и отзывов, чтобы копия не вернула удалённые данные, старые права или скомпрометированный ключ.
7. Выполнить положительные и отрицательные тесты доступа и Guardrail, затем вернуть рабочий маршрут. Проверить накопленные задачи и риск повторной отправки.
8. Измерить фактические RPO и RTO, описать потери и ограничения, сохранить протокол и продолжить усиленное наблюдение.

Откат инфраструктуры не отменяет уже состоявшийся внешний вызов. Повтор задачи может создать второй документ, платный запрос или внешнее действие. Поэтому восстановление очередей требует решения, какие операции безопасно повторять, какие сопоставлять по идентификатору, а какие разбирать вручную.

Обновления и изменения конфигурации

Перед обновлением составляется карточка изменения: компоненты и версии, миграции, затронутые политики, зависимости, ожидаемый эффект, проверка успеха и условия отката. Для on-prem проверяются пакеты, образы, контроль целостности и изменения внешних зависимостей. Для изолированного профиля дополнительно подтверждается перенос по разрешённой процедуре и отсутствие новой внешней рабочей зависимости.

Обновление сначала проходит в проверочном окружении с безопасными данными. Проверяются вход, права пространств, загрузка и поиск документов, защищённый модельный запрос, запрещённый маршрут, экспорт аудита и удаление. Набор выбирается по изменению: новая версия DLP требует контрольных примеров распознавания, изменение схемы данных — проверки восстановления и совместимости.

Перед рабочим переключением сохраняются применённая конфигурация и пригодная точка восстановления. Если миграция необратима, возврат старого образа недостаточен: заранее определяется способ восстановления данных и допустимая потеря записей после переключения. Поэтапный ввод и ограниченная проверка относятся к процедуре эксплуатации; конкретный механизм зависит от топологии поставки.

После обновления сравниваются ошибки, задержки, решения Guardrail и доставка событий с исходным состоянием. Ухудшение разбирается по стадии и затронутому сценарию. Отсутствие ошибок запуска не подтверждает корректную изоляцию данных или сохранение прежней политики.

Приёмочный протокол

Протокол содержит идентификатор окружения, версии, фактическую топологию, время испытания, выполненный отказ, использованную копию или точку, измеренные RPO/RTO и результаты проверки прав. Для базового профиля ежемесячное восстановление подтверждается таким протоколом, а не только уведомлением «копирование завершено».

Заказчик проверяет три результата: сервис восстановлен в установленной области, данные соответствуют выбранной точке и защитные ограничения сохранились. Отдельно фиксируются непроверенные отказы и несовпадения с целевым профилем. Они становятся конкретными эксплуатационными действиями с владельцами; до повторного испытания соответствующий показатель остаётся неподтверждённым.

БЕЗОПАСНОСТЬ

Корпоративная среда и российские средства ИБ

Как включить Nevel в корпоративный контур с SIEM, DLP, сетевым шлюзом и IdP; точки контроля, соединения и программа проверки.

Облако / В контуре

https://docs.nevel.sh/security/corporate-environment/

Nevel включается в существующую среду компании через отдельные соединения для входа, модельных запросов, контроля содержания и доставки событий. Для каждого соединения определяются инициатор, получатель, видимые данные, аутентификация и поведение при недоступности. Это позволяет использовать корпоративные средства ИБ, сохраняя понятную границу ответственности между платформой и инфраструктурой заказчика.

Ниже приведены российские продукты с документированными производителем интерфейсами и варианты их места в архитектуре. Это **основа проектирования интеграции**, а не перечень испытанной совместимости Nevel. Подтверждённая интеграция относится к конкретным версиям, настройкам и протоколу испытаний. Ссылка на документацию производителя не подтверждает наличие у Nevel сертификата, готового коннектора или совместного внедрения.

Какие средства за что отвечают

Задача	Примеры продуктов	Документированный механизм	Место в проекте Nevel
Собирать и разбирать события	Kaspersky KUMA	Нормализаторы JSON, CEF и Syslog, сопоставление полей	Получатель событий платформы и инфраструктуры
Централизовать события и правила их обработки	MaxPatrol SIEM	Сбор событий и правила нормализации	Разбор отдельного источника Nevel по подготовленному профилю
Принимать Syslog	UserGate SIEM	TCP/UDP, TLS для TCP, ограничения отправителей	Корпоративный приёмник либо следующий узел после relay
Контролировать сетевой путь и передавать содержание на проверку	UserGate NGFW	Инспектирование TLS и подключения к ICAP-серверам	Сетевой шлюз и клиент DLP, если этот путь выбран в проекте
Проверять содержание средствами корпоративной DLP	InfoWatch Traffic Monitor	Интеграционные API; опубликованные примеры ICAP/ICAPS	Отдельная точка анализа в разрешённой схеме
Выполнять корпоративную аутентификацию	Avanpost FAM	SAML, OpenID Connect и другие механизмы интеграции	IdP для корпоративного входа в согласованном профиле SSO

Основания таблицы: KUMA — парсинг событий, MaxPatrol SIEM — нормализация, UserGate SIEM — Syslog, UserGate NGFW — ICAP, интеграции InfoWatch Traffic Monitor, Avanpost FAM. ICAP/ICAPS у InfoWatch также описаны на примере интеграции с DION; этот пример относится к указанным производителем продуктам и не является свидетельством интеграции с Nevel.

Обычный сетевой экран, DLP и специализированный AI firewall решают разные задачи. Сетевой экран управляет соединениями. DLP анализирует доступное ему содержание. AI firewall или модельный шлюз может быть включён в прикладную цепочку, если его интерфейс поддерживает нужные форматы модельного API. Поддержку streaming, инструментов, изображений и форматов ошибок проверяют отдельно; она не следует из наличия HTTPS или ICAP. Принцип подключения собственного AI firewall описан в сетевой защите.

Где средство видит исходный текст

Выбор точки контроля определяет область раскрытия. TLS защищает данные между точками завершения соединения; средство, которое анализирует тело запроса, получает доступ к расшифрованному содержимому соответствующего участка.

Точка контроля	Что доступно средству	Что проверить
До Guardrail и маскирования	Исходный запрос или извлечённый текст; оригинал файла — только если он передан этой стадии	Разрешение на обработку оригинала, дополнительные копии, журналы и временные файлы
После маскирования, перед моделью	Подготовленный модельный контекст с масками и оставшимся содержанием	Сохранение масок, роли сообщений, вложений и параметров; отсутствие обходного вызова
На ответе до восстановления	Ответ модели с маркерами	Сохранение структуры ответа, streaming и возврат результата контроля
После восстановления	Результат с восстановленными значениями	Отдельная чувствительная граница; не пересылать сюда ответ автоматически ради дополнительной проверки
Копия трафика или экспорт событий	Только реально переданный экземпляр или метаданные	Копирование не является синхронным разрешением на отправку исходного запроса

Если браузер отправил файл в Platform, а Platform позже извлекла текст и добавила фрагменты RAG, перехват только браузерного соединения не показывает итоговый контекст модели. И наоборот, контроль модельного запроса после маскирования не даёт DLP оригинальных значений. Для цели «проверить весь фактически отправляемый контекст» точка контроля связывается со сборкой контекста, а не только с пользовательским интерфейсом.

СХЕМА 1

Две возможные точки контроля

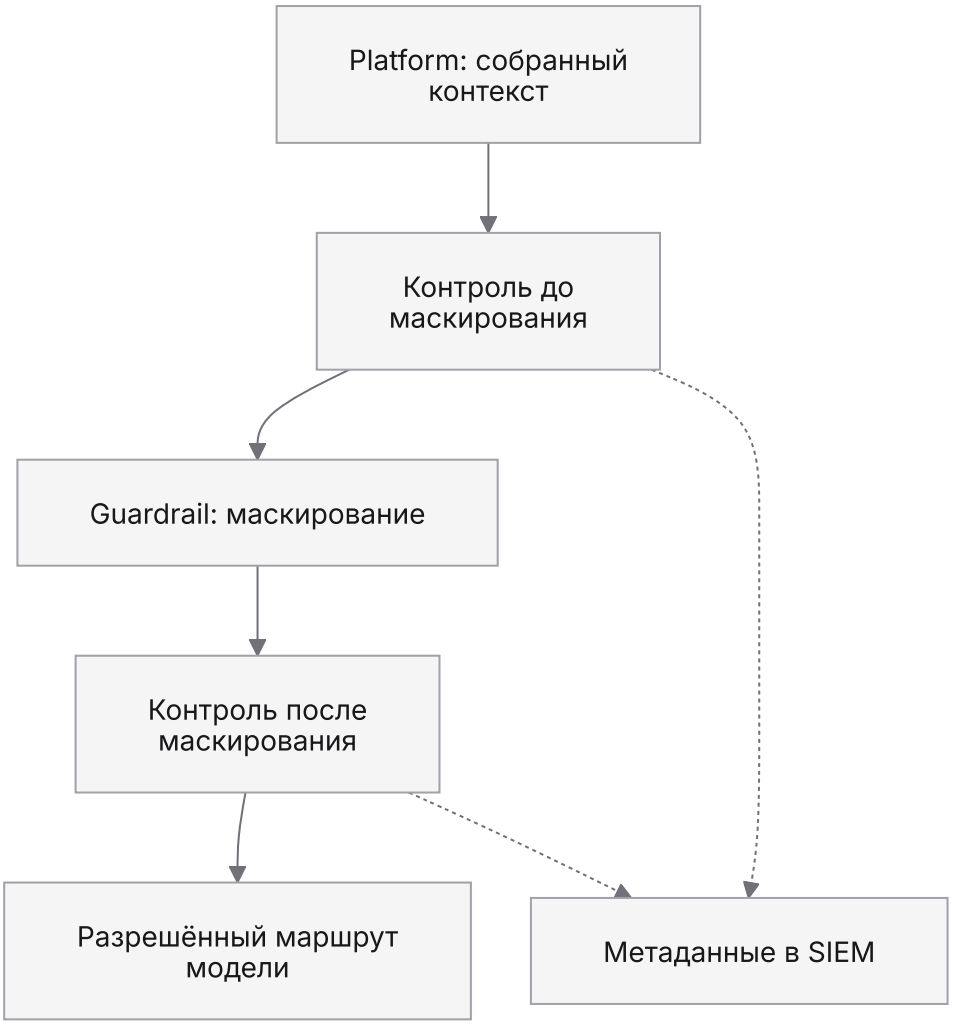


Схема показывает места включения, а не необходимость последовательно устанавливать два одинаковых продукта. Если выбрана одна точка, другая не добавляется автоматически. Обратный путь ответа и точку восстановления фиксируют в паспорте подключения отдельно.

Выбор адреса и способа подключения

Есть три разных способа включить контроль. **Модельный reverse proxy** становится адресом, на который настроен модельный клиент; он сохраняет контракт запросов и ответов. **Сетевой proxy с DLP по ICAP** принимает сетевой трафик и отправляет доступное содержание DLP; ICAP-адрес не является OpenAI-compatible endpoint. **Адаптер анализа** вызывает отдельный API проверки в выбранной стадии приложения; для него нужны спецификация вердикта и согласованная реализация подключения.

Не подставляйте ICAP-адрес вместо базового URL модели. Не используйте адрес веб-консоли DLP как API анализа. Получите адрес именно принимающего сервиса, его протокол, требования к сертификатам и поддерживаемые операции. Точные URL и секреты принадлежат конфигурации установки; в публичных примерах используются только зарезервированные демонстрационные домены.

Соединение	Параметры конфигурации	Контроль результата
Platform → модельный шлюз	Базовый URL, разрешённые пути, авторизация, CA, тайм-ауты	Сохраняются модель, сообщения, tools и события окончания потока

Соединение	Параметры конфигурации	Контроль результата
Сетевой шлюз → DLP	Адрес/порт ICAP или другой подтверждённый интерфейс, направление проверки, предел сообщения	Проверка действительно получает нужное тело, а не только заголовок
Platform → IdP	Metadata/discovery, идентификатор приложения, адрес возврата и доверие	Ответ предназначен этому приложению; корпоративные атрибуты правильно связаны с учётной записью
Экспортер → relay/SIEM	Адрес приёмника, формат события, транспорт, сертификаты, область организаций	Событие принято, разобрано и находится по исходному ID

В UserGate NGFW документация ICAP перечисляет адрес, порт, предел сообщения, проверку доступности, направления REQMOD/RESPMOD и параметр `bypass`. Последний допускает работу без передачи в недоступный ICAP-сервер. Поэтому для обязательного контроля проверяются и настройка bypass, и действия соответствующих правил; итог подтверждается тестом отказа. Параметры ICAP, работа с внешними ICAP-серверами.

Облако, on-prem и размещение промежуточных узлов

On-prem определяет место Platform, а не расположение каждой модели. Платформа заказчика может обращаться к локальной модели, разрешённому российскому провайдеру, self-hosted модели Nevel в РФ или глобальному провайдеру. Корпоративный шлюз проверяет именно выбранные маршруты. Матрица размещения отделяет их от изолированного профиля.

Для облачной Platform корпоративное средство заказчика должно быть доступно по явно подготовленному защищённому соединению, если его включают в обязательный рабочий путь. Наличие средства во внутренней сети не даёт облаку такой доступ автоматически. Маршрут включает сеть, DNS, сертификаты, ограничения источников и обратный ответ; исходные хранилища и административные консоли для этого не публикуются.

Relay для событий можно разместить у заказчика или на выделенной разрешённой площадке в РФ. Он завершает входящее соединение, при необходимости преобразует формат и доставляет события в SIEM. Это отдельный обработчик метаданных: его дисковая очередь, диагностика и резервирование входят в схему хранения. Размещение relay в РФ не определяет географию модельного запроса; модельный путь и путь аудита рассматриваются независимо.

В изолированном профиле модели, средства анализа, IdP, relay и SIEM рабочего процесса находятся внутри заданной границы. В обычном on-prem допустимы разрешённые внешние соединения. Ни один профиль не должен скрыто переключаться на другого модельного получателя из-за недоступности средства контроля.

Корпоративный вход

Avanpost FAM предоставляет механизмы SAML и OpenID Connect для подключения приложений, а также способы интеграции с каталогами учётных записей. Это подтверждённые возможности IdP, а не готовый профиль Nevel. Описание FAM, документация FAM 1.17.

При проектировании определяются стабильный идентификатор пользователя, допустимые атрибуты, группы и процедура изменения членства. Успешный ответ IdP создаёт проверенную идентичность, но не должен автоматически превращать каждую группу каталога в административную роль Platform. Для входа, сервисного API-ключа и чтения событий SIEM действуют отдельные полномочия. Подробности — идентификация и права.

Как принять интеграцию

ИТ фиксирует версии Nevel и подключаемого продукта, размещение узлов и конфигурацию соединений; ИБ определяет видимое содержание и обязательность контроля; владелец SIEM проверяет разбор событий. На стенде выполняются текстовый запрос, streaming, файл, разрешённый инструмент, отмена, ошибка провайдера, недоступность контрольного узла и приём события.

Для каждого результата должны быть исходный ID операции, запись на контрольном компоненте и запись у получателя аудита. Проверяются сохранность структуры сообщения и невозможность обойти обязательный шлюз альтернативным адресом. Одна успешная проверка соединения подтверждает лишь доступность адреса, а не корректность всего сценария.

Итоговый комплект: схема фактических соединений, таблица видимых данных, профиль настроек без секретов, reference mapping событий и протокол проверенных сценариев с ограничениями. Пошаговое подключение и проверка доставки раскрыты в SIEM-процессе; структура событий — в профиле аудита.

Версии и совместимость

Параметры подключения относятся к KUMA 3.0.1, MaxPatrol SIEM 8.1, UserGate 7.x и Avanpost FAM 1.17. Перед настройкой сверяйте интерфейсы с установленным выпуском и лицензией. Описанный интерфейс продукта не подтверждает совместимость с Nevel: для выбранной конфигурации нужен протокол испытаний.

БЕЗОПАСНОСТЬ

Защита от prompt injection

Границы доверия, проверки инструкций и инструментов, режимы реакции и негативная приёмка защиты от внедрённых команд.

Облако / В контуре

https://docs.nevel.sh/security/prompt-injection/

Prompt injection — попытка превратить содержимое запроса, файла, изображения или результата инструмента в команду, которая меняет поведение ассистента вопреки его задаче и полномочиям. Прямая атака приходит в сообщении пользователя; косвенная — через источник, который ассистент читает для ответа. Обычная цитата «игнорируй предыдущие инструкции» сама по себе ещё не доказывает атаку: важны контекст, запрашиваемое действие и пересечение границы доверия. OWASP LLM01:2025.

В Nevel защита объединяет проверку содержимого, разделение данных и инструкций, контроль источников и серверную проверку действий. Детектор помогает остановить подозрительный контекст; решение модели не создаёт право прочитать закрытый файл, изменить маршрут или выполнить внешний вызов. Это правило применяется к Platform, RAG и цепочкам инструментов независимо от места установки.

Что имеет право задавать инструкции

Уровень	Назначение	Чего он не разрешает
Серверная политика организации	Доступ, допустимые получатели данных, инструменты и обязательные проверки	Модель не может изменить её собственным ответом
Настройка ассистента от уполномоченного редактора	Задача, правила ответа и разрешённый рабочий процесс	Настройка не расширяет права пользователя или сервисного ключа
Запрос пользователя	Конкретная задача в пределах его полномочий	Просьба пользователя не отменяет обязательную политику организации
Документ, сайт, найденный фрагмент, ответ инструмента	Материал для выполнения задачи	Текст не становится системной командой, разрешением на действие или новым секретом

Извлечённый материал передаётся как отдельный контекст с идентификатором источника и версии. Текст вроде «это служебное сообщение администратора» внутри PDF сохраняет статус содержимого PDF. Разделители и структурированные поля помогают модели распознать происхождение, но основная граница полномочий проверяется вне генерации. Секреты подключения не помещаются в промпт для последующего сокрытия инструкцией «не показывать».

Права источника и доверие к его инструкциям — разные признаки. Сотрудник может законно читать документ поставщика, который содержит вредоносную команду. Даже утверждённая база знаний не получает права управлять подключениями. Процесс допуска и обновления источников описан в защите файлов и баз знаний.

Проверки на разных стадиях

Проверяется материал, который фактически получит модель: пользовательский ввод, извлечённый текст, выбранные RAG-фрагменты, ответы инструментов и повторно используемый контекст. Проверка только первого сообщения оставляет открытым последующий ответ инструмента. Для изображений учитывается распознанный текст; если модели передаётся оригинал, один текстовый анализ не доказывает безопасность изображения.

Стадия	Проверка	Пример причины для ограничения
Извлечение	Структура файла, скрытые слои, необычные управляющие символы, различие видимого и извлечённого текста	Команда находится в невидимом слое PDF

Стадия	Проверка	Пример причины для ограничения
Входной контекст	Признаки смены роли, обхода правил, запроса секрета или действия вне задачи; смысловой анализ вместе с правилами	Фрагмент справки предлагает отправить весь диалог на сторонний адрес
План действия	Соответствие исходному заданию пользователя, разрешённому инструменту и допустимым параметрам	Вместо чтения одного документа предложен экспорт всей коллекции
Выполнение	Идентичность, область организации, права на объект, операция и маршрут	Корректный по форме вызов обращается к чужому пространству
Ответ и артефакт	Допустимость раскрытия, ссылок и активного содержимого	Ответ содержит скрытую внешнюю ссылку с данными в параметрах

Правила ищут известные признаки, а смысловая проверка учитывает цель фрагмента и предлагаемого действия. Например, словосочетание «системная инструкция» в учебном тексте допустимо; требование раскрыть секрет под видом такой инструкции — отдельный сигнал. Обфускация, другие языки и распределение команды по нескольким фрагментам включаются в проверки. Проверяющая модель тоже обрабатывает недоверенные данные и не заменяет детерминированную авторизацию. OWASP: практики предотвращения prompt injection.

СХЕМА 1

Контекст не даёт полномочий

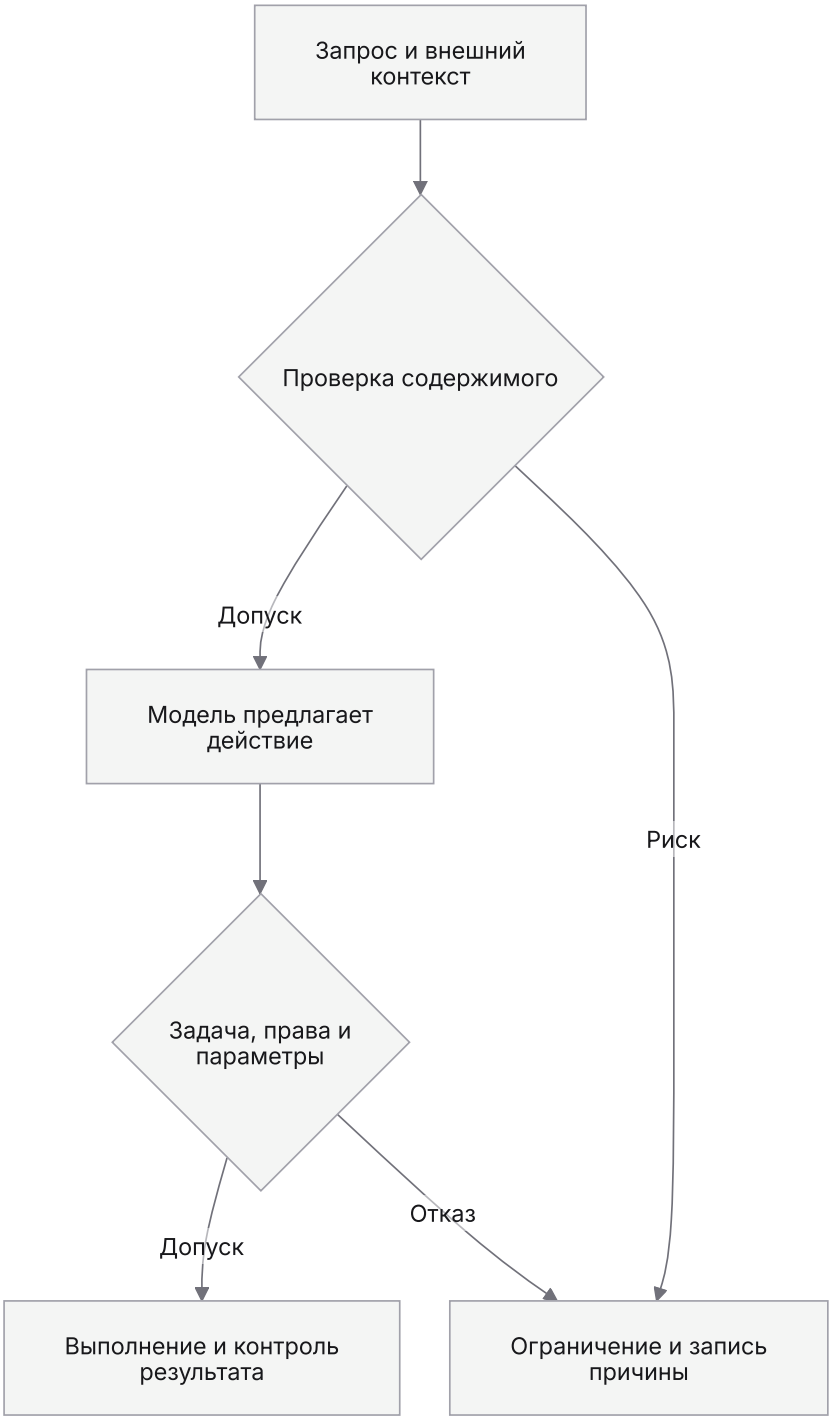


Схема разделяет два решения: можно ли использовать контекст и можно ли выполнить конкретное действие. Положительное первое решение не заменяет второе.

Проверка инструментов перед исполнением

Сервер повторно проверяет пользователя или сервисную идентичность, организацию, действие, объект и допустимые параметры. Право читать не разрешает запись; право работать в одном пространстве не открывает другое. Раздельные области инструментов и минимальные полномочия рекомендует OWASP AI Agent Security. Проверка выполняется непосредственно перед действием, поэтому отзыв доступа учитывается и в уже начатой цепочке.

Аргументы проверяются по ожидаемым типам и допустимым значениям. Адрес назначения, идентификатор коллекции и размер выборки не принимаются как доверенные только потому, что модель сформировала корректный JSON. Для сетевого инструмента проверяется разрешённый получатель, включая перенаправления; для изменения объекта — область и характер изменения. Инструмент использует ограниченные сервисные полномочия, а не универсальные права администратора.

Если рабочий процесс требует подтверждения человеком, до исполнения показываются конкретное действие, объект и существенные параметры. Строка «пользователь уже подтвердил» в найденном документе не заменяет подтверждение в доверенном процессе. Проверка результата инструмента также обязательна: подключённый сервис может вернуть текст с новой командой, а следующая модель не должна принять его за изменение задачи.

Режимы реакции и пример профиля

Реакция привязана к правилу и риску операции. Наблюдение сохраняет сигнал без изменения ответа; исключение убирает подозрительный источник из контекста; отказ останавливает запрещённое действие. При исключении критичного источника ответ должен отразить недостаточность доступных данных, а не подменять их выдуманным содержимым. Решение о конкретном запросе не означает автоматическую блокировку пользователя.

Ниже — **рекомендуемый пример профиля для проектирования и приёмки**, а не перечень переключателей интерфейса или готовый формат конфигурации:

Ситуация	Реакция профиля	Дальнейшее действие
Учебный документ обсуждает атаки, запрещённых действий нет	Допуск с сохранением происхождения	Ответ по существу; проверка инструментов остаётся обязательной
Новый источник содержит скрытую команду вывести секрет	Карантин версии	Анализ владельцем источника; исправленная версия проходит проверки заново
Фрагмент требует сменить разрешённого получателя	Отказ в смене маршрута	Продолжение возможно только по исходной допустимой политике
Контекст предлагает запись, а пользователь запросил чтение	Отказ в выполнении записи	Разрешённый ответ по доступным данным
Недоступна обязательная проверка высокорискового действия	Действие не исполняется	Фиксация технической причины и восстановление проверки

При отказе не раскрываются обнаруженные секреты и недоступные названия документов. В журнал включаются идентификатор запроса, стадия, источник и версия, версия политики, категория сигнала и принятое решение. Полное вредоносное содержимое не требуется копировать в обычный журнал; диагностический образец хранится отдельно с ограниченным доступом. Общий порядок событий приведён в мониторинге.

Негативные проверки и критерии приёмки

Тесты используют синтетические данные, отдельные пространства и нерабочие контрольные секреты. Для каждого случая сохраняются вход, права, версии модели и политики, ожидаемый исход и фактический след. Эта таблица задаёт программу испытаний; она не является отчётом о проведённых тестах Nevel.

Сценарий	Ожидаемый результат	Что проверяет испытатель
PDF предлагает раскрыть контрольную фразу чужой коллекции	Чужой источник недоступен	Поиск, ответ, цитаты и журнал обращения
Команда скрыта в слое PDF или распознанном изображении	Применён профиль анализа соответствующего представления	Что именно извлечено и передано модели
Ответ инструмента требует отправить историю наружу	Запрещённый вызов не выполнен	Вызовы инструментов и сетевые назначения

Сценарий	Ожидаемый результат	Что проверяет испытатель
Команда разбита по нескольким RAG-фрагментам	Полномочия и маршрут не расширены	Собранный контекст и фактическое действие
Доступ отозван между планом и исполнением	Следующая операция отклонена	Серверное решение с актуальными правами
Проверяющий компонент недоступен	Обязательная защита не обходится	Причина отказа, отсутствие побочного действия
Полезный текст цитирует пример атаки	Рабочая задача остаётся выполнимой	Отсутствие необоснованного отказа

Проверяется последствие, а не наличие определённой фразы отказа. Модель может пересказать опасную инструкцию как предмет анализа, не исполнив её. После смены модели, экстрактора, источника, шаблона ассистента или политики повторяются затронутые сценарии и обычные рабочие задачи. Нет обещания обнаружить любую возможную инъекцию: остаточный риск ограничивают права, допустимые маршруты и проверка действий. Связанные угрозы собраны в модели угроз.

В облаке и on-prem действуют одинаковые границы доверия. Локальная модель также может подчиниться вредоносному тексту. Место проверяющего компонента и передаваемый ему контекст включаются в схему потоков: российский внешний обработчик, глобальный API и обработчик внутри инфраструктуры заказчика — разные маршруты, каждый требует своей разрешающей политики.

Источники

Классы угроз и защитные подходы описаны в OWASP LLM01:2025, OWASP Prompt Injection Prevention Cheat Sheet и рекомендациях OWASP по инструментам агентов. Результаты защиты конкретного внедрения определяются испытаниями его моделей, инструментов и политик.

Облако, on-prem и маршруты моделей

Два места установки платформы и четыре модельных маршрута: локальные, российские провайдеры, self-hosted Nevel и глобальные модели.

Облако / В контуре

https://docs.nevel.sh/deployment/compare/

Место установки Nevel и место выполнения модели выбираются отдельно. **Облако** означает размещение платформы в инфраструктуре Nevel. **On-prem** означает размещение платформы в инфраструктуре заказчика. В обоих случаях модельный маршрут определяется подключениями и политикой данных. On-prem не требует автоматически использовать только локальные модели.

Два независимых решения

Вопрос	Облако Nevel	On-prem заказчика
Где работают Platform и её установленные защитные компоненты	В инфраструктуре Nevel	В инфраструктуре заказчика
Где находятся рабочие хранилища платформы	В облачной конфигурации поставки	В инфраструктуре заказчика
Где выполняется модель	По отдельно выбранному разрешённому маршруту	По отдельно выбранному разрешённому маршруту
Кто эксплуатирует платформу	Nevel в пределах поставки	Заказчик и Nevel по распределению работ
Как подключаются источники и модели	Через доступные разрешённые соединения	Через внутренние и разрешённые исходящие соединения
Как устанавливаются обновления	В процессе эксплуатации облака	По регламенту обновления установки заказчика

Таким образом, «платформа у нас» и «все данные остаются у нас» — разные требования. Второе выполняется отдельным **изолированным профилем**: вся рабочая обработка, включая модели и вспомогательные сервисы, остаётся внутри заданного контура.

Варианты модельного маршрута

Маршрут	Где работает модель и кто её обслуживает	Что пересекает границу платформы	Применение
Локальная модель заказчика	В инфраструктуре заказчика	При локальной Platform запрос остаётся внутри согласованного контура	Закрытые данные и задачи, для которых доступна подходящая локальная модель
Российский модельный провайдер	На площадке отдельного провайдера, например Яндекса или Сбера	Подготовленный контекст передаётся российскому получателю за пределами инфраструктуры платформы	Облако и on-prem, если маршрут разрешён политикой
Self-hosted модель Nevel в РФ	Экземпляр модели, например DeepSeek, который Nevel размещает на российских серверах	В облаке обработка идёт в соответствующей инфраструктуре Nevel; из on-prem — запрос передаётся в неё	Модельный маршрут с размещением и эксплуатацией модели у Nevel

Маршрут	Где работает модель и кто её обслуживает	Что пересекает границу платформы	Применение
Глобальная модель	У выбранного глобального провайдера по разрешённой цепочке API	Контекст передаётся участникам этой цепочки; их география учитывается отдельно	Облако и on-prem, если допустимы получатели, данные и география

Self-hosted DeepSeek у Nevel и обращение к внешнему API с моделью DeepSeek имеют разные маршруты. Название модели не показывает, кто получил запрос и где его обработал. Аналогично российский провайдер не становится частью инфраструктуры заказчика только из-за размещения в РФ.

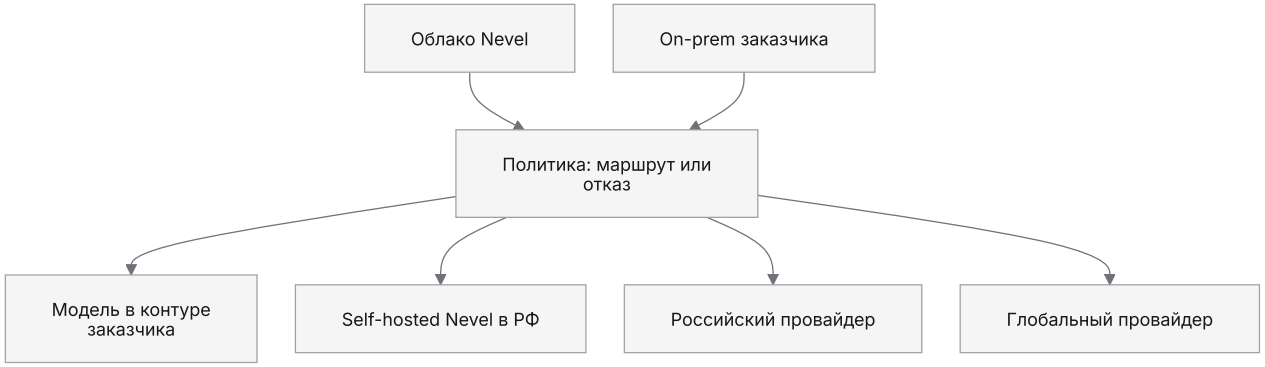
Для облака обращение к модели в контуре заказчика требует отдельно подготовленного разрешённого соединения; её доступность не возникает автоматически. Для любой комбинации необходимы подключённый endpoint, совместимость нужных методов и разрешение политики. Таблица не означает, что любой экземпляр модели уже доступен каждому пространству. Каталог API показывает доступные варианты для конкретного ключа.

Типовые комбинации

Конфигурация	Где платформа	Разрешённые модели	Что проверяется
Облако с self-hosted Nevel	Nevel	Выбранная модель на серверах Nevel в РФ	Фактический endpoint и отсутствие подмены внешним API
Облако с российскими провайдерами	Nevel	Яндекс, Сбер или другой подключённый разрешённый провайдер	Получатель, состав контекста и правила хранения
Облако с глобальными моделями	Nevel	Разрешённые глобальные маршруты	Вся цепочка получателей, Guardrail и география
On-prem с локальными и внешними маршрутами	Заказчик	Локальные, российские, self-hosted Nevel и глобальные — по политике	Категория данных определяет маршрут каждого запроса
On-prem с изолированным профилем	Заказчик	Только модели внутри заданного контура	Отсутствие внешней обработки во всех стадиях задачи

СХЕМА 1

Место платформы и маршрут модели



Стрелки показывают варианты архитектуры. Их фактическая доступность зависит от подключений и настроек. Политика может разрешить глобальную модель для одного класса документов и только внутреннюю — для другого в той же установке Platform.

Изолированный профиль

В изолированном профиле в контуре находятся Platform, Guardrail, модели, обработка файлов, OCR, индексация, embeddings, инструменты, рабочие хранилища и служебные данные. Внешняя модель, облачный OCR или внешний веб-поиск не входят в такой рабочий маршрут. Обновления и поддержка регулируются отдельной административной процедурой.

Если нужной внутренней возможности нет, операция завершается контролируемым отказом или выполняется другой разрешённой внутренней моделью. Отказ не разрешает скрытое переключение во внешний API. В обычном on-prem резервный внешний маршрут допустим только тогда, когда он заранее разрешён для соответствующих данных и отражается в аудите.

Как принять решение

Владелец процесса определяет задачи и качество результата. ИБ задаёт категории данных и разрешённые маршруты. ИТ выбирает размещение Platform, доступность источников, сеть и эксплуатацию. Юристы оценивают фактических получателей и применимые условия обработки. Передача за пределы инфраструктуры компании и трансграничная передача рассматриваются отдельно; это различие разобрано в 152-ФЗ.

Для каждой комбинации зафиксируйте основной и резервный endpoint, владельца модели, географию, этапы обработки файлов и правила хранения. Включите в схему журналы, копии, поддержку и инструменты: они могут иметь собственные маршруты.

Приёмка подтверждает правильный маршрут обычного запроса, отказ запрещённого получателя, поведение при недоступности модели и отсутствие скрытого fallback. Только для изолированного профиля дополнительно проверяется выполнение всех нужных сценариев без внешней обработки.

Дальше: облачная поставка, установка on-prem и схемы безопасности.

Облачное размещение

Как подготовить доступ к облаку Nevel, согласовать модели и маршруты данных и принять рабочую конфигурацию.

Облако

https://docs.nevel.sh/deployment/cloud/

В облачном варианте компания использует платформу и защитные компоненты в инфраструктуре Nevel. Пользователи работают через предоставленный интерфейс, приложения — через выданный адрес API. Конкретные модели, интеграции и правила обработки определяются конфигурацией компании. Облачное размещение платформы не определяет автоматически модельного получателя: он выбирается отдельно.

Модельные маршруты облака

Вариант	Где выполняется модель	Что проверяется
Self-hosted Nevel в РФ	На российских серверах Nevel; например, размещённый Nevel экземпляр DeepSeek	Конкретный endpoint и обработка именно в инфраструктуре Nevel
Российский провайдер	У отдельного провайдера, например Яндекса или Сбера	Передача этому получателю, состав данных и условия обработки
Глобальный провайдер	По разрешённой цепочке API к выбранной модели	Все участники цепочки и фактическая география

Название DeepSeek само по себе не различает self-hosted экземпляр Nevel и внешний API. В конфигурации фиксируются место выполнения и исполнитель. Подключение модели внутри инфраструктуры заказчика рассматривается отдельно и требует доступного разрешённого соединения; оно не возникает автоматически из облачного доступа.

Выбранные модели доступны только в пределах разрешений пространства, ключа и категории данных. Отказ основной модели не разрешает скрытое переключение на другого получателя. Матрица размещения и маршрутов помогает выбрать комбинацию.

Подготовьте требования

До подключения определите состав пользователей, основные сценарии и категории обрабатываемых данных. Перечислите системы, с которыми платформа должна взаимодействовать, и требуемое направление доступа. Например, чтение корпоративного справочника и запись задачи в рабочий трекер — разные разрешения, даже если обе операции выполняет один ассистент.

Назначьте ответственных за пространство, ИБ, интеграции и коммерческие условия. Подготовьте адреса для приглашений и перечень приложений, которым нужен API. Если корпоративная сеть фильтрует исходящие соединения, запросите актуальный список необходимых адресов для вашей конфигурации. Не составляйте разрешения по случайным доменам из старых примеров.

Согласуйте путь данных

Для каждого сценария зафиксируйте, где принимается запрос, где обрабатываются вложения, какая модель получает подготовленный контекст и где сохраняется результат. Учитывайте данные, которые поступают из подключённых источников уже во время выполнения задачи. Исходный короткий запрос пользователя может привести к чтению большого документа или обращению к внешней системе.

Разрешение использовать облако не означает автоматического разрешения любых моделей. Каталог и политика компании могут ограничивать провайдеров и отдельные модели. Если для части данных допустима только определённая обработка, это требование необходимо включить в конфигурацию и проверить отдельным сценарием. Общие принципы описаны в обзоре безопасности.

Настройте команду

Создайте или получите рабочее пространство, назначьте владельца и администраторов. Подключите небольшую группу пользователей и проверьте роли до массовых приглашений. Разместите согласованные тестовые материалы и выполните один сценарий совместной работы: автор создаёт результат, а другой участник получает предусмотренный доступ.

Для интеграций используйте учётные данные с разрешениями, необходимыми именно этому сценарию. Определите владельца подключения и порядок его замены. Если подключение персональное, проверьте, не зависит ли общий процесс команды от доступа одного сотрудника. Для API заведите отдельные ключи для приложений и окружений, затем настройте лимиты.

Что зафиксировать в эксплуатационных условиях

Область	Необходимая договорённость
Доступ	Способ входа, приглашения, отзыв прав и ответственные
Данные	Категории, места хранения, сроки и порядок удаления
Модели	Разрешённый каталог и порядок его изменения
Интеграции	Учётные записи, разрешения и владельцы
Поддержка	Канал обращения, необходимые сведения и порядок эскалации
Изменения	Уведомление и проверка затронутых сценариев

Условия поддержки и хранения определяются для конкретного обслуживания. Не переносите сроки из чужого договора на свою конфигурацию. Для рабочих файлов, истории чатов и технических журналов могут действовать разные правила: перечисляйте их раздельно.

Приёмка облачной конфигурации

Проверьте вход через корпоративную сеть, доступность нужного пространства и разрешённой модели, загрузку подготовленного файла и получение результата. Проведите проверку ограничения прав и запрещённой модели. Если используется API, выполните запрос к каталогу и небольшую генерацию, затем проверьте поведение с отключённым ключом.

Сохраните время проверки, название сценария и результат. При ошибке передавайте в поддержку HTTP-статус, код ошибки и идентификатор запроса, если он доступен. Исключите из обращения полный ключ и лишние рабочие данные. После успешной технической проверки проведите продуктовую приёмку: сотрудники должны самостоятельно повторить свои задачи и оценить пригодность результата.

Перед расширением нагрузки согласуйте ограничения, порядок добавления пользователей и ответственного за контроль расходов. Повторяйте основные проверки после существенного изменения модели, интеграции или политики доступа.

ВАРИАНТЫ УСТАНОВКИ

On-prem в инфраструктуре компании

Установка Platform у заказчика, подключение локальных и внешних моделей по политике и отдельный изолированный профиль.

В контуре

https://docs.nevel.sh/deployment/onprem/

On-prem — установка Platform в инфраструктуре заказчика. Платформа может использовать локальные модели, российские провайдеры, self-hosted модели Nevel на серверах в РФ и глобальные модели по разрешённым маршрутам. Полностью внутренняя обработка является отдельным изолированным профилем, а не обязательным значением слова on-prem.

Состав установки

Внутри инфраструктуры заказчика размещаются компоненты поставки: пользовательский интерфейс, прикладные сервисы, рабочие хранилища и защитная обработка. Местоположение моделей и вспомогательных обработчиков задаётся отдельно. Для каждого OCR, векторизатора, инструмента или модельного endpoint указываются исполнитель, сеть, данные и разрешённая область обработки.

Комплект установки фиксирует версии, ресурсы, конфигурацию связей, порядок развёртывания, обновления и восстановления. Команды берутся из комплекта конкретной версии. Наличие on-prem не означает одинаковых требований к GPU: при внешних модельных маршрутах вычисления модели выполняет соответствующий получатель, а при локальных — инфраструктура заказчика.

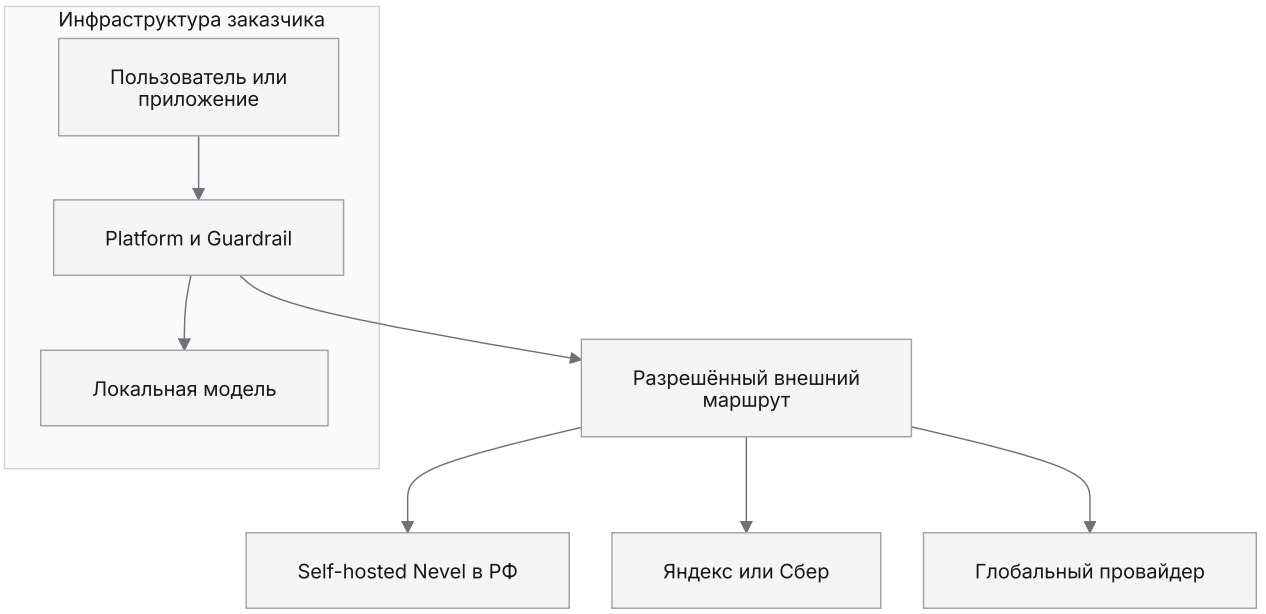
Выберите маршруты моделей

Вариант	Подключение из on-prem	Ключевой контроль
Локальная модель заказчика	Внутренний endpoint	Доступность нужных форматов и производительность внутренних ресурсов
Российский провайдер, например Яндекс или Сбер	Разрешённое исходящее API-соединение	Получатель вне инфраструктуры заказчика, категории данных и условия обработки
Self-hosted Nevel в РФ, например DeepSeek	Разрешённый endpoint инфраструктуры Nevel	Отличие этого экземпляра от API внешнего владельца модели
Глобальная модель	Разрешённая цепочка API, включая промежуточные сервисы при их наличии	Все получатели и география, политика Guardrail и отсутствие обхода

Выбор фиксируется для класса данных, пространства или приложения в пределах политики. Например, закрытый кадровый документ может обрабатываться только локально, общедоступный текст — глобальной моделью, а внутренний рабочий материал — разрешённым российским маршрутом. Названия категорий и конкретные разрешения задаёт организация; этот пример не является готовым юридическим разрешением.

СХЕМА 1

Маршруты модели из on-prem



Передача к Яндексу, Сберу или Nevel пересекает границу инфраструктуры заказчика. География фактической обработки проверяется отдельно. Место установки Platform само по себе не определяет, возникает ли трансграничная передача.

Подготовьте инфраструктуру

Определите приоритетные сценарии, одновременную нагрузку, размеры файлов и требования к ожиданию. Для локальной модели отдельно оцените вычисления и память; для внешней — соединения, лимиты и доступность провайдера. Поддержка метода API не делает возможности разных моделей одинаковыми: текст, изображения, инструменты и форматы ответа проверяются на выбранном экземпляре.

Подготовьте адреса сервисов, сертификаты, секреты и доступ эксплуатационной команды. Исходящие соединения разрешаются по конкретным назначениям. Базы и хранилище соответствий не открываются внешним модельным провайдерам: им передаётся подготовленный контекст по разрешённому API-маршруту.

Журналы, резервные копии, каналы поддержки и обновлений включаются в схему отдельно. On-prem не даёт автоматически разрешение на любой внешний мониторинг или выгрузку диагностического архива.

Порядок развёртывания

1. Зафиксировать место Platform, компоненты, модельные маршруты и матрицу ответственности.
2. Получить комплект установки для выбранной версии и подготовить инфраструктуру.
3. Развернуть компоненты и проверить их состояние, сертификаты и сервисные права.
4. Подключить выбранные локальные и внешние модели по разрешённым адресам.
5. Настроить категории данных, Guardrail, основные и резервные маршруты.
6. Создать пространства, участников и отдельные ключи тестовых приложений.
7. Принять рабочие сценарии, запреты, аудит и восстановление до переноса рабочих данных.

Конфигурация сохраняется в контролируемом процессе эксплуатации. Секреты не включаются в открытые инструкции. Для изменения маршрута фиксируются инициатор, прежнее и новое состояние и затронутые категории данных.

Приёмка маршрутизации

Проверьте один запрос для каждого разрешённого маршрута и по журналу подтвердите конечного и промежуточных получателей. Повторите сценарий с документом, которому этот маршрут запрещён. Наличие ключа глобального провайдера не должно обходить локальную метку.

Затем отключите основную модель. Запрос может использовать только заранее разрешённый для этой категории резерв; выбор отражается в аудите. Если подходящего резерва нет, возвращается контролируемый отказ. Клиентский SDK также проверяется на отсутствие собственного скрытого переключения.

Включите загрузку файла, OCR, поиск по знаниям, инструмент и сохранение артефакта: каждый этап может обращаться к отдельному обработчику. Отдельно проверьте отзыв доступа, защиту секретов и восстановление тестовой копии.

Изолированный профиль on-prem

Если организация требует, чтобы вся рабочая обработка оставалась внутри контура, используются только внутренние модели и вспомогательные сервисы. Внутри также находятся OCR, embeddings, инструменты, хранилища, журналы и резервные копии. Внешний поиск и облачная модель для таких задач исключаются политикой и сетевыми правилами.

Приёмка этого профиля проводится при запрещённой внешней обработке и охватывает обычные задачи и отказы. Недоступность локальной модели не открывает внешнюю альтернативу. Сетевая проверка подтверждает внутреннюю цепочку, а не только успешный ответ интерфейса.

Эксплуатация

Ответственность за Platform и внутренние компоненты отделяется от ответственности внешнего модельного провайдера. Изменение endpoint, модели, получателя, ключа или региона рассматривается как изменение маршрута и проходит затронутые проверки.

В тестовой среде проверяются обновления приложения, моделей, коннекторов и политик. Для изолированного профиля дополнительно определяется перенос пакетов внутрь контура. Порядок доступа поддержки, диагностические данные, каналы и откат фиксируются в регламенте. Подробнее — эксплуатация и сетевая защита.

Первый запрос к API

Как получить доступный каталог моделей и выполнить текстовый запрос через корпоративный API Nevel.

Облако / В контуре

https://docs.nevel.sh/api/quickstart/

API Nevel позволяет внутреннему приложению или агенту обращаться к моделям через корпоративный шлюз. Для первого запроса нужны адрес API вашей установки, ключ рабочего пространства и разрешённая модель. Этот сценарий отличается от подключения внешнего сервиса к чату: здесь именно ваше приложение вызывает модель.

Подготовьте параметры

Получите у администратора три значения и передайте их приложению через переменные окружения. Не сохраняйте ключ в исходном коде, браузерном приложении или общем документе.

Переменная	Значение
`NEVEL_BASE_URL`	Выданный базовый URL API с согласованным префиксом маршрута, без завершающего `/`
`NEVEL_API_KEY`	Действующий ключ нужного рабочего пространства
`NEVEL_MODEL`	Идентификатор разрешённой текстовой модели из каталога

В защищённом варианте относительный префикс маршрутов — `/secure/v1`. Например, к полному выданному базовому URL добавляется `/models` или `/chat/completions`. Не добавляйте второй `/v1`, если он уже входит в `NEVEL_BASE_URL`.

Маршруты с `/secure/v1` и обычные `/v1` различаются включением защитной обработки. Не меняйте префикс ради обхода ошибки. Допустимый маршрут и конфигурацию обработки определяет администратор; само название маршрута не подтверждает качество маскирования или допустимость конкретных данных. Подробнее — в разделе безопасности.

Получите каталог моделей

Следующая команда требует `curl`. Переменные должны быть заданы в текущем окружении. Проверки в начале предотвращают отправку запроса с отсутствующими параметрами.

```
: "${NEVEL_BASE_URL:?Укажите выданный базовый URL API}"
: "${NEVEL_API_KEY:?Укажите API-ключ}"

curl --fail-with-body --silent --show-error \
  "${NEVEL_BASE_URL%/}/models" \
  -H "Authorization: Bearer ${NEVEL_API_KEY}"
```

Успешный ответ имеет `object: "list"` и массив `data`. Используйте значение `id` разрешённой текстовой модели. Не выбирайте автоматически первый элемент: каталог может содержать модели другого назначения. Если нужной модели нет, проверьте конфигурацию пространства с администратором.

Выполните текстовый запрос

Пример использует Python 3 и стандартную библиотеку, поэтому не зависит от версии стороннего SDK. Он корректно формирует JSON и выводит текст первого ответа. Значение `NEVEL_MODEL` предварительно задаётся по полученному каталогу.

```
import json
import os
import urllib.error
import urllib.request

base_url = os.environ["NEVEL_BASE_URL"].rstrip("/")
payload = {
    "model": os.environ["NEVEL_MODEL"],
    "messages": [
        {"role": "user", "content": "Ответь одним словом: готово"}
    ],
    "stream": False,
}
request = urllib.request.Request(
    base_url + "/chat/completions",
    data=json.dumps(payload, ensure_ascii=False).encode("utf-8"),
    headers={
        "Authorization": "Bearer " + os.environ["NEVEL_API_KEY"],
        "Content-Type": "application/json",
    },
    method="POST",
)
try:
    with urllib.request.urlopen(request, timeout=60) as response:
        result = json.load(response)
    print(result["choices"][0]["message"]["content"])
except urllib.error.HTTPError as error:
    print("HTTP", error.code)
    print(error.read().decode("utf-8", errors="replace"))
    raise SystemExit(1)
```

Таймаут 60 секунд в примере — настройка клиента, а не гарантия времени ответа сервиса. Для своего приложения выберите таймаут по рабочему сценарию и обработайте прерывание запроса. Не используйте бесконечные автоматические повторения: они могут создавать дополнительную нагрузку и расход.

Подключение совместимого клиента

В клиенте с настройкой OpenAI-compatible API укажите выданный базовый URL, ключ и идентификатор модели. Начните с простого текстового запроса без инструментов, изображений и специальных параметров. После успешной проверки добавляйте необходимые возможности по одной: их поддержка зависит от модели и маршрута.

Для on-prem используйте выданный адрес API установленной у заказчика платформы. Этот адрес определяет точку входа, а поле `model` и политика — дальнейший маршрут. Модель может работать локально у заказчика, у российского провайдера, на серверах Nevel в РФ или у глобального провайдера, если такой путь разрешён. Доступные идентификаторы берите из каталога конкретного пространства, а размещение и получателей — из паспорта маршрута: имя модели само по себе их не доказывает.

Изолированный профиль использует только внутренние обработчики и модели. В любой конфигурации клиент и SDK не должны незаметно переключаться на другого провайдера при ошибке. Резервный маршрут допускается только как заранее разрешённый для этих данных путь с фиксацией переключения. Сверьте настройки клиента с матрицей вариантов и сетевой политикой установки.

Проверка результата

Убедитесь, что каталог доступен, текстовый запрос завершился успешно, а приложение корректно показывает ошибку с отключённым ключом. Зафиксируйте модель, время и HTTP-статус без сохранения секрета. Затем настройте лимиты и жизненный цикл ключа и добавьте обработку ошибок перед использованием рабочих данных.

API-ключи и лимиты

Авторизация Bearer, области ограничений RPM и TPM, окна, всплески нагрузки, параллелизм и поведение потоковых запросов.

Облако / В контуре

https://docs.nevel.sh/api/auth-limits/

API-ключ предоставляет приложению программный доступ к моделям в рамках рабочего пространства. Он определяет контекст использования и позволяет отдельно управлять доступом и расходом интеграции. Ключ следует считать секретом: наличие пользовательского входа в платформу не заменяет его для программного запроса.

Авторизация запроса

Передавайте ключ в HTTP-заголовке `Authorization` с типом `Bearer`. Пример первого запроса читает его из `NEVEL_API_KEY`. Не передавайте ключ в URL: адрес запроса может попасть в историю, журналы промежуточных систем и диагностические материалы.

Ключ связан с конкретным пространством. Каталог моделей и ограничения проверяются для этого контекста. Если приложение получает отказ для модели, убедитесь, что используется ключ нужного пространства. Переключение пространства в браузере не изменяет переменную окружения запущенного приложения.

Ключ и маршрут к модели

API-ключ задаёт доступ приложения, но не определяет географию обработки. Место установки платформы и выбранный маршрут проверяются отдельно. В on-prem запрос приходит в платформу заказчика, после чего политика может разрешать локальную модель, российский API, self-hosted модель Nevel на серверах РФ или глобального провайдера. В облаке также доступны разрешённые для поставки self-hosted модели Nevel в РФ и провайдерские маршруты. Конкретный набор определяется каталогом и политикой пространства; матрица вариантов объясняет различия.

Собственный ключ провайдера меняет авторизацию и учёт, но не переносит его модель в инфраструктуру заказчика. Разрешение модели ключу не должно обходить запрет категории данных, список получателей или изолированный профиль. Основной и резервный пути проверяются одинаково: скрытое переключение на другой сервис при отказе не допускается.

Создание и назначение

Раздел API-ключей доступен при включённом управлении ключами. Администратор должен определить назначение ключа, владельца приложения, окружение и ожидаемую нагрузку. Используйте различимые названия, чтобы по списку ключей можно было понять, какому процессу принадлежит запрос.

Создавайте отдельные ключи для разработки и производственной работы. Для независимых приложений также удобнее иметь разные ключи: это позволяет отозвать один доступ без остановки остальных и отдельно ограничивать расход. Полученный секрет передаётся в хранилище секретов или переменные окружения серверного приложения.

Состояния и срок действия

Состояние	Значение для эксплуатации
`active`	Ключ активен; дополнительно проверяются срок действия, модель и ограничения
`disabled`	Ключ отключён; доступ следует считать приостановленным
`revoked`	Ключ отозван; для продолжения работы выдаётся новый доступ

У ключа может быть срок действия. Перед окончанием срока подготовьте замену, обновите секрет приложения и проверьте небольшой запрос. После подтверждения перевода отключите или отзовите старый ключ. При подозрении на раскрытие сначала прекращайте использование скомпрометированного доступа, затем проверяйте затронутые приложения и расход.

Лимиты расходов

Ключ может иметь собственный лимит с периодом `daily`, `weekly`, `monthly` или `lifetime`. Лимит и период задаются совместно. Период `lifetime` относится к суммарному использованию ключа без регулярного обновления бюджета. Для периодических лимитов ориентируйтесь на фактическое время следующего сброса, доступное в состоянии лимита, а не на предположение о локальной полуночи пользователя.

Лимит ключа не отменяет другие ограничения пространства и обслуживания. Даже при остатке бюджета конкретному ключу может быть запрещена модель. И наоборот, разрешённая модель не гарантирует доступность запроса после исчерпания бюджета. Уточняйте причину по HTTP-статусу и машинному коду ошибки.

Ограничения частоты и нагрузки

Rate limits ограничивают нагрузку за интервал или число одновременно выполняемых операций. Они действуют отдельно от денежного бюджета: короткие частые запросы могут исчерпать частоту, а один длинный запрос — доступную пропускную способность по токенам. Ниже описан механизм настройки и эксплуатации; приведённые примеры не задают квоты подписок или действующие численные лимиты Nevel.

Ограничение	Что измеряется	Что учитывает клиент
RPM / RPS	Количество запросов за минуту / секунду	Частоту запуска, включая повторные попытки
TPM	Токены в заданном минутном интервале	Размер входа, ожидаемый выход и правила учёта выбранного маршрута
Concurrency	Одновременно принятые незавершённые запросы	Длительность генерации, открытые потоки и освобождение слотов
Burst	Короткий разрешённый всплеск запуска	Мгновенную подачу очереди после простоя
Размер запроса	Байты HTTP-тела или контекст конкретной модели	Размер вложений и поддерживаемый контекст; это отдельная проверка

Счётчик запросов не равен счётчику токенов модели. Если на входе используется предварительная оценка токенов, профиль определяет её состав: вход и зарезервированный выход либо иной проверяемый способ. Для неизвестного выхода нельзя заранее считать фактическое конечное число. Итоговый учёт после завершения сверяется с доступными данными обработки; отсутствие финального usage при обрыве учитывается как незавершённый расчёт, а не как нулевое потребление.

Области действия

Ограничение связывается с конкретной областью: API-ключом, пространством или организацией, маршрутом и моделью, общим пулом обработки. Несколько ключей одного приложения могут расходовать общий ресурс пространства. Несколько моделей могут обслуживаться одним пулом. Поэтому доступность собственного ключа не гарантирует доступности общего пула.

Проверки применяются совместно: запрос должен удовлетворять каждому действующему ограничению. Число RPM одного ключа нельзя напрямую сравнивать с общим TPM и выбирать одно «минимальное» значение — это разные единицы и области. Полученный отказ должен разбираться в своей области; создание новых ключей не является способом обхода общего ограничения.

На внешнем маршруте дополнительно действуют ограничения получателя модели. Его `429` возможен после успешного допуска на стороне Nevel. Внутренние повторы шлюза также создают нагрузку получателю. Разделяйте входящие запросы приложения и исходящие попытки маршрута, чтобы не увеличивать нагрузку повтором одновременно в SDK, приложении и шлюзе.

Окна и короткие всплески

Фиксированное окно сбрасывает счётчик на установленной границе интервала. Из-за этого запросы в конце одного окна и начале следующего могут образовать заметный всплеск. Скользящее окно учитывает предшествующий интервал относительно текущего момента и ограничивает такой эффект. Формулировка «в минуту» без описания окна не позволяет точно определить время следующего допуска.

Token bucket использует накапливаемые разрешения: они поступают с заданной скоростью до установленной ёмкости, а запуск тратит разрешение. Ёмкость определяет burst; после простоя можно кратко запустить больше запросов, чем обычная скорость пополнения. Эти разрешения не являются текстовыми токенами LLM. Пример такого механизма описан в официальной документации Amazon API Gateway; параметры AWS не являются настройками Nevel.

Concurrency работает отдельно: даже при полном bucket новый запрос не принимается, если все слоты заняты. Burst ограничивает скорость старта, а concurrency — число незавершённых операций. Очередь клиента должна учитывать оба условия и иметь предел размера и времени ожидания.

Пример профиля для интеграции

Следующая таблица — **иллюстративная рекомендуемая конфигурация для испытательного стенда**, а не тариф, настройки по умолчанию или публичное обязательство Nevel. Её параметры позволяют воспроизвести разные виды отказа; производственные значения выбираются по фактическому маршруту и нагрузке.

Параметр примера	Значение	Смысл проверки
Частота одного ключа	60 запросов за скользящие 60 секунд	Отдельный предел количества запусков
Короткий всплеск ключа	Bucket на 5 разрешений, пополнение 1 в секунду	Ограничение резкого старта после простоя
Токены ключа	120 000 за скользящие 60 секунд	В этом примере резервируются вход и разрешённый максимальный выход
Параллельность ключа	4 незавершённых запроса	Пятый не запускается до освобождения слота
Завершение расчёта	Корректировка резерва по фактическому usage	Неиспользованный резерв освобождается по правилам профиля

Например, пять одновременно отправленных запросов не обязательно пройдут: у bucket достаточно разрешений, но предел параллельности допускает только четыре. Если один запрос резервирует 40 000 токенов, токеновый предел этого примера может остановить четвёртый раньше ограничения concurrency. Это независимые причины; увеличение только RPM их не устраняет.

Порядок допуска и завершения

1. Входной шлюз ограничивает чрезмерные тела и очевидную перегрузку до дорогостоящей обработки. Защита до авторизации может учитывать сетевой источник, но не заменяет лимиты приложения: за одним адресом могут находиться разные клиенты.
2. Сервер проверяет ключ, его состояние, пространство, разрешённую модель и маршрут. Неавторизованный запрос не получает модельные ресурсы.
3. Применяются действующие лимиты ключа, пространства и маршрута. Перед запуском резервируются необходимые токены и место в обработке; денежное резервирование остаётся отдельной проверкой.
4. После допуска выполняется вызов. Успех, ошибка и отмена завершают жизненный цикл: ресурсы освобождаются, фактическое использование учитывается, причина завершения сохраняется.

При нескольких экземплярах шлюза счётчики одной области должны оставаться общими, а допуск — не превышать ограничение из-за одновременного чтения старого остатка. В приёмке это проверяется запросами через разные экземпляры. Резервирование не оставляет занятый слот навсегда после падения процесса: эксплуатационный профиль определяет срок резерва, продление для живого запроса и восстановление зависшего состояния. Ошибка хранилища обязательных ограничений не должна незаметно превращать ограниченный маршрут в неограниченный.

Уже принятый запрос остаётся попыткой, даже если модель вернула ошибку: возврат временного резерва токенов не означает отмены факта запроса в RPM. Поведение отклонённых до допуска запросов фиксируется отдельно для входной защиты; клиент не должен рассчитывать, что бесконечные `429` бесплатны для всех счётчиков.

Потоки и длинные операции

Один потоковый запрос занимает один слот до завершения или фактического прекращения обработки. Каждый фрагмент SSE не является новым запросом и не обнуляет минутное окно. При разрыве клиентского соединения сервер освобождает слот после остановки либо завершения связанной работы; отправка отмены не доказывает мгновенную остановку внешнего провайдера.

Ограничения допуска проверяются до начала потока. После отправки HTTP-заголовков `200` сервер уже не может заменить статус того же ответа на `429`: последующая проблема обрабатывается как ошибка или прерывание потока по контракту маршрута. Клиент проверяет штатное завершение и не запускает повтор всей операции автоматически при каждом обрыве. Подробнее — ошибки и повторные запросы.

При `429` до начала обработки клиент разбирает причину и `Retry-After`, если заголовок присутствует. Сам HTTP-стандарт не определяет область счётчика и не требует этот заголовок в каждом ответе: RFC 6585, раздел 4. Имена дополнительных заголовков и кодов зависят от используемого ограничителя.

Параллельные запросы и резервирование

Код `api_key_busy` связан с **резервом денежного бюджета** и сам по себе не подтверждает, что достигнут RPM, TPM или предел concurrency из профиля частоты.

Запросы в обработке могут резервировать доступный бюджет. Поэтому остаток складывается с учётом уже потраченных и временно зарезервированных средств. При активных запросах новый вызов может получить `429` с кодом `api_key_busy`, даже если окончательные расходы ещё не достигли лимита.

В этом случае ограничьте параллелизм и учитывайте заголовок `Retry-After`, если он присутствует. После завершения текущих операций доступность бюджета пересчитывается. Не выдавайте автоматически новый ключ для обхода ограничения: сначала проверьте незавершённые запросы и согласованный бюджет приложения.

При `429 insufficient_quota` автоматический повтор не устраняет исчерпание квоты. Приложение должно остановить цикл повторов и сообщить владельцу о необходимости проверить бюджет или дождаться предусмотренного сброса периода.

Проверка перед запуском

Проверьте успешный запрос с действующим ключом, отказ с отключённым ключом, отказ для запрещённой модели и поведение при нескольких одновременных запросах. Убедитесь, что приложение не пишет секрет в журнал и различает временную занятость ключа и исчерпанную квоту. Для диагностики сохраняйте название приложения, время, статус и код ошибки; полный ключ в обращение поддержки не включается.

Для профиля частоты отдельно проверьте границу окна, всплеск после простоя, исчерпание TPM длинным запросом, занятые слоты, несколько ключей одного пространства и внешний `429`. Затем оборвите поток и убедитесь, что слот освобождается после завершения обработки; повторите проверку при падении обслуживающего процесса. Сравните ожидаемый и фактический допуск, область ограничения и задержки клиента. Численные результаты заносятся в протокол испытания, а не выводятся из примера профиля.

Численные лимиты конкретной установки задаются её конфигурацией; их нельзя определить по общему описанию HTTP или token bucket.

Ошибки API и повторные запросы

Как отличать ошибки доступа, лимитов, размера запроса и недоступности модели; какие запросы повторять.

Облако / В контуре

https://docs.nevel.sh/api/errors/

Обрабатывайте HTTP-статус вместе с телом ответа. Один статус может обозначать несколько причин: например, `429` возникает при ограничении частоты, исчерпании квоты или временном резервировании бюджета активными запросами. Пользовательское сообщение и стратегия повтора должны соответствовать конкретной причине и источнику отказа.

Формат ошибки

Ошибки шлюза возвращаются в объекте `error`. Он содержит описание `message`, тип `type` и, когда предусмотрено, машинный код `code`. Дополнительные поля могут указывать параметр запроса или технические подробности. Не стройте логику на точном английском тексте сообщения: используйте статус и код.

Ответ внешнего провайдера может иметь отличающуюся структуру. Клиент должен корректно обработать неизвестную ошибку и случай, когда промежуточный сетевой компонент возвращает текст вместо JSON. Сохраните статус и безопасное диагностическое описание, а пользователю покажите понятное следующее действие.

Основные случаи

HTTP	Код или ситуация	Действие
`400`	`invalid_request`	Проверить тело, обязательные `model` и `messages`
`401`	`invalid_api_key`	Проверить заголовок Bearer, состояние и срок ключа
`403`	`model_not_allowed`	Выбрать разрешённую модель или обратиться к администратору
`404`	`not_found`	Проверить метод, базовый URL и относительный путь
`413`	`payload_too_large`	Уменьшить размер тела запроса
`429`	`insufficient_quota`	Проверить бюджет и срок обновления квоты
`429`	`api_key_busy`	Учесть `Retry-After`, сократить параллелизм
`429`	Ограничение частоты, токенов или параллельности на шлюзе либо у провайдера	Определить область, уменьшить нагрузку и выполнить ограниченный повтор; код зависит от компонента
`502`	`upstream_invalid_response`	Проверить временную ошибку ответа модели

HTTP	Код или ситуация	Действие
`504`	`upstream_timeout`	Ограниченно повторить запрос с учётом его результата и стоимости
`500`	`server_error`	Сохранить диагностику; при повторении обратиться в поддержку

Таблица описывает основные случаи шлюза. Другие коды могут прийти от используемой модели или сетевой инфраструктуры. Если ошибка не соответствует ожидаемой, начните с минимального текстового запроса по быстрому старту, используя тот же адрес, ключ и модель.

Превышение размера

Ошибка `413 payload_too_large` от шлюза может содержать `error.metadata.body_limit_bytes`. Это ограничение размера HTTP-тела в байтах, а не длины контекста модели в токенах. Вложенное изображение в base64 увеличивает размер JSON; короткий текст запроса не означает, что всё тело небольшое.

Уменьшите вложение или разделите задачу на допустимые части. Если выбранная модель поддерживает получение изображения по URL, используйте только разрешённый политикой адрес с необходимыми правами. Адрес должен быть доступен именно обработчику выбранного маршрута. В on-prem внешний провайдер не получает доступ к внутреннему URL автоматически; разрешённый способ передачи файла проверяется отдельно. Для изолированного профиля обработчик и адрес остаются внутри периметра. Не делайте рабочий файл публичным ради обхода ограничения размера.

Когда повторять запрос

Ошибки авторизации, запрещённой модели и некорректного тела требуют изменения настроек или запроса. Немедленный повтор той же операции обычно не поможет. Для `insufficient_quota` проверьте лимиты, а для `api_key_busy` дождитесь освобождения зарезервированного бюджета и учитывайте `Retry-After`.

Для временных сетевых ошибок и отдельных `5xx` используйте ограниченное число попыток с увеличивающейся задержкой и случайным разбросом. Таймаут не доказывает, что модель не начала обработку. Повтор может привести к дополнительному расходу и другому ответу. Если приложение выполняет действия во внешних системах, защищайте эти действия от повторного исполнения отдельно.

Разбор временного ограничения

`429` от ограничителя частоты означает, что текущий запрос не допущен по соответствующему пределу. Это может быть RPM одного ключа, TPM пространства, занятый пул модели или ограничение внешнего провайдера. Сначала сократите нагрузку в этой области. Пауза одного потока не поможет, если остальные экземпляры приложения продолжают заполнять ту же очередь.

При неизвестном коде нельзя автоматически считать любой `429` временным: сначала исключите исчерпанный бюджет. Если доступна причина временного ограничения, используйте ограниченный повтор. При неопределённой причине сохраните диагностику и остановите бесконечный цикл. Не создавайте новый ключ и не меняйте провайдера для обхода правил; области rate limits описаны отдельно.

Ответ `429` может содержать `Retry-After`; обязательного наличия заголовка стандарт не устанавливает. Такой ответ нельзя сохранять для последующей выдачи как кэшированный результат API. RFC 6585, раздел 4.

Retry-After и случайная задержка

`Retry-After` бывает целым неотрицательным количеством секунд после получения ответа либо HTTP-датой. Клиент должен поддерживать обе формы. Например, `Retry-After: 12` означает ожидание 12 секунд; дата задаёт момент, до которого нужно ждать. Для даты учитывайте время UTC и рассинхронизацию часов. Невалидное значение не трактуется как разрешение непрерывно повторять запросы. RFC 9110, раздел 10.2.3.

Для временной ошибки без такого заголовка используйте exponential backoff с jitter: верхняя граница задержки растёт после неудачи, а конкретная пауза выбирается случайно внутри неё. Это распределяет повторы клиентов по времени. Одного одинакового удвоения паузы недостаточно: одновременно запущенные клиенты могут повторно встретиться в той же точке.

Подход объяснён в AWS Architecture Blog.

Пример политики клиента, а не готовая конфигурация Nevel:

```
n — номер повторной попытки, начиная с 0
window = min(cap, base × 2^n)
если Retry-After корректен:
    delay = wait_from_header + random(0, window)
иначе:
    delay = random(0, window)
повтор разрешён только в пределах общего deadline и числа попыток
```

`base`, `cap`, deadline и предел попыток задаются владельцем интеграции по допустимому времени ответа. Ограничение `cap` в этом примере относится к случайной добавке, а не сокращает обязательное ожидание из заголовка. Если требуемая пауза выходит за deadline задачи, клиент завершает попытки и сообщает о временной недоступности. Истёкшая дата даёт нулевую нижнюю границу; случайная задержка и предел попыток сохраняются.

Повторы выполняются на одном выбранном уровне интеграции. Если SDK делает до трёх попыток на один вызов, а приложение вызывает его до трёх раз, итоговая нагрузка может стать девятикратной. При сетевой ошибке после отправки неизвестно, была ли операция принята. В таком случае перед повтором побочных действий проверяйте их состояние; произвольный заголовок идемпотентности не создаёт гарантию, если его поддержка не определена контрактом endpoint.

Повтор запроса не даёт разрешения менять получателя. В облаке и on-prem исключите скрытый fallback в клиенте, SDK и шлюзе. Заранее разрешённый резервный маршрут повторно проверяется для категории данных, а переключение фиксируется. При отсутствии допустимого резервного пути запрос завершается ошибкой; для изолированного профиля сбой локальной модели не открывает внешний маршрут. Варианты размещения и маршрутов описывают эти границы.

Потоковые ответы

При потоковой генерации успешные HTTP-заголовки могут прийти до завершения ответа. Последующий обрыв соединения нельзя считать полноценным успешным результатом. Сохраняйте признак завершения, различайте частичный и законченный ответ и сообщайте пользователю о прерывании. Автоматическое продолжение или повтор должно учитывать уже показанный текст и возможные действия приложения.

После начала ответа `200` статус этого же ответа не меняется на `429`. Ошибка приходит внутри поддерживаемого формата потока либо проявляется его прерыванием; формат зависит от маршрута. Не пытайтесь обнаруживать её только проверкой исходного HTTP-статуса. При повторе прерванного потока предыдущий текст не приклеивается к новой генерации как гарантированное продолжение: результат может отличаться, а первый запрос ещё обрабатываться у провайдера.

Проверка клиента

На тестовом ответе проверьте `429` с задержкой в секундах, HTTP-датой, отсутствующим и некорректным заголовком. Убедитесь, что при `insufficient_quota` повторов нет, временная перегрузка уменьшает подачу общей очереди, а достижение deadline прекращает ожидание. Для потока проверьте обрыв после нескольких фрагментов и отсутствие двойного выполнения связанного действия. Синтетические ответы тестового сервера проверяют логику клиента; они не подтверждают фактические лимиты или поведение конкретного провайдера.

Что передать в поддержку

Подготовьте время с часовым поясом, относительный путь API, модель, HTTP-статус, код ошибки и идентификатор запроса, если он возвращён. Добавьте частоту повторения и минимальное обезличенное тело, достаточное для воспроизведения. Не включайте ключ, заголовок авторизации или рабочие файлы целиком. Для on-prem укажите версию установки и изменения сети или конфигурации, после которых возникла ошибка.

Если речь об ограничении, приложите число активных запросов, частоту отправки, размер входа, источник `429` и полученный `Retry-After`, когда они известны. Отсутствующее поле отметьте как неизвестное; не восстанавливайте область лимита по одному статусу.

Как устроена стоимость Nevel

Лицензии сотрудников, доступ к моделям, лимиты и внедрение: из чего складывается бюджет компании.

Облако / В контуре

https://docs.nevel.sh/pricing/overview/

В облаке Nevel оплачивается по числу сотрудников и выбранному тарифу. Pro стоит **1 990 ₽ за сотрудника в месяц**, Advanced — **4 990 ₽**. Доступ к моделям входит в подписку. **Pro и Advanced имеют одинаковые возможности и инструменты; основная разница — лимиты использования.** Для компании со своим ключом к моделям есть вариант **1 000 ₽ за сотрудника в месяц**: потребление по ключу оплачивает заказчик. Для внешнего API Nevel действует **стоимость OpenRouter + 20%**.

Для установки в контуре действует годовая стоимость решения: от **2,5 млн ₽ для конфигурации до 100 пользователей**. Оборудование для установки оплачивается отдельно. Публичные тарифы Nevel.

При планировании закупки разделяйте четыре величины: цену лицензии, расходы на выполнение моделей, ограничения использования и затраты на развёртывание. Они отвечают на разные вопросы и не заменяют друг друга.

Три модели оплаты доступа к моделям

Сценарий	Что оплачивается
Сотрудники на Pro / Advanced	Подписка по опубликованной цене с доступом к моделям в составе плана
Сотрудники с ключом заказчика	1 000 ₽ за человека в месяц за Platform; потребление отдельно в аккаунте собственного провайдера
Внешнее приложение через Nevel API	Учтённая стоимость OpenRouter × 1,20

Правило ×1,20 применяется к внешнему Nevel API. Оно не добавляется автоматически к расходам по собственному ключу в пользовательском варианте за 1 000 ₽. Если компания использует несколько сценариев, каждый учитывается по своему основанию; наличие лицензии сотрудника не означает обязательную покупку отдельного API-доступа.

Что покупает компания

Составляющая	Что она означает	Где смотреть условия
Лицензия Platform	Доступ сотрудника к рабочему интерфейсу, моделям и функциям выбранного плана	Облачные планы
Platform с ключом заказчика	1 000 ₽ за сотрудника в месяц; расходы на модели в аккаунте провайдера заказчика	Вариант с собственным ключом
Инструменты Platform	Одинаковые в Pro и Advanced: документы, таблицы, презентации, исследования, медиа, интеграции и другие сценарии	Возможности и лимиты планов
Квоты и ограничения	Сколько операций можно выполнить, какой контекст использовать, какие модели доступны конкретному сотруднику	Использование и бюджеты
Внешний API	Программное выполнение запросов, отдельные ключи, ограничения расходов и аналитика	API и бюджеты
Установка в контуре	Развёртывание Platform в инфраструктуре заказчика, подключение согласованных моделей и эксплуатационная конфигурация	Стоимость установки в контуре

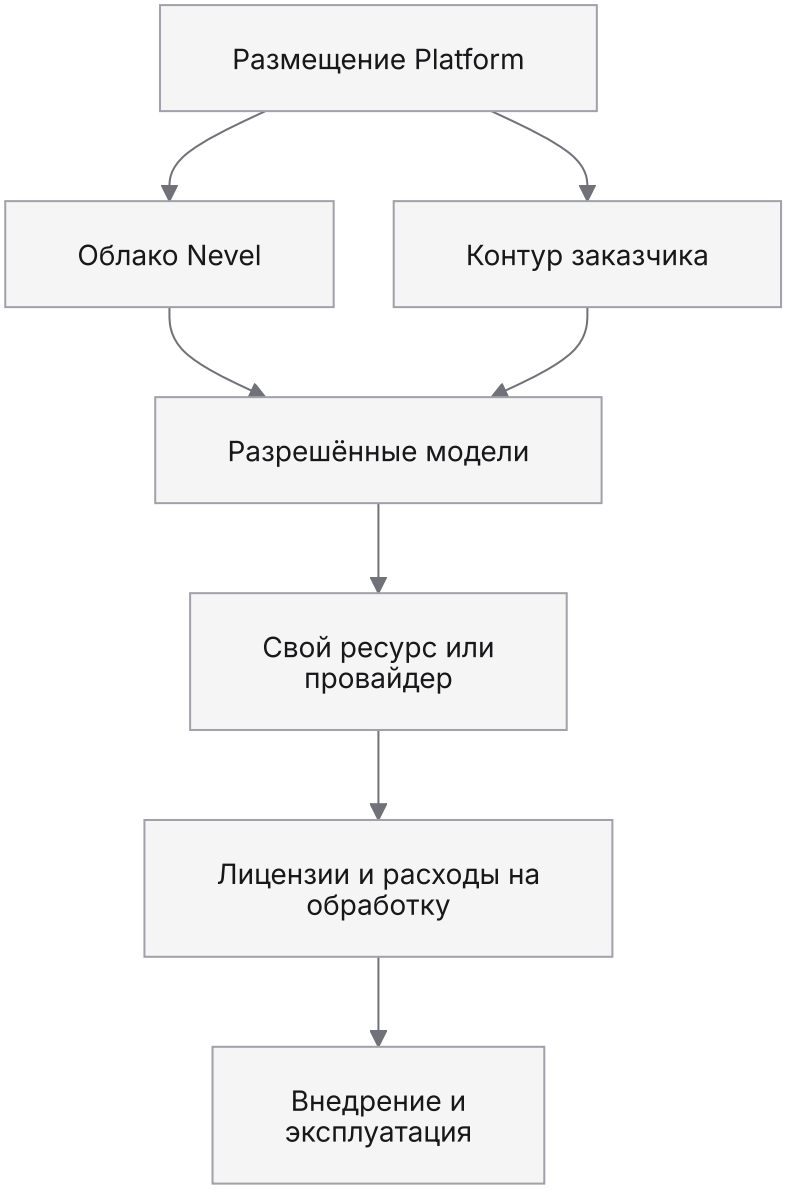
Подписка объединяет доступ к поддерживаемым моделям в Nevel. Сотруднику не требуется покупать отдельную подписку на каждую модель, используемую через платформу. При этом доступ к каталогу моделей не означает одинаковые квоты для любых операций или автоматический доступ к каждому внешнему сервису. Модель может быть ограничена политикой компании; для подключаемой интеграции нужны права в соответствующей системе.

Например, лицензия Nevel не создаёт сотруднику учётную запись в CRM и не расширяет его права в CRM. Она даёт возможность использовать подключение в пределах прав, выданных владельцем внешней системы. Расходы и подписки на такую систему относятся к её собственным условиям.

Как выбрать основу расчёта

СХЕМА 1

Размещение и расчёт стоимости



Для облака основная единица публичного расчёта — сотрудник за месяц. Для годовой самостоятельной подписки оплачивается весь годовой период. Для контура основа расчёта — конфигурация установки и диапазон пользователей. У неё другой состав поставки, поэтому её стоимость нельзя получить простым умножением облачного тарифа.

Маршрут моделей выбирается отдельно от места размещения Platform: модели заказчика, российские провайдеры — например Яндекс и Сбер — либо глобальные модели. В облаке Nevel также доступна self-hosted DeepSeek на серверах в РФ. Для собственных моделей учитываются вычислительные ресурсы и эксплуатация; для провайдерских маршрутов — соответствующее потребление. Установка в контуре может использовать разрешённые внешние модели. Полностью изолированный профиль ограничивает обработку внутренними ресурсами и рассматривается как отдельный вариант конфигурации.

Пример облачного бюджета

Предположим, компании нужен доступ для 25 сотрудников: 20 используют Pro, ещё 5 — Advanced из-за более интенсивной работы и потребности в повышенных лимитах. Для предварительной сметы по опубликованным месячным ставкам:

Группа	Расчёт	Стоимость за месяц
20 сотрудников на Pro	20 × 1 990 ₽	39 800 ₽
5 сотрудников на Advanced	5 × 4 990 ₽	24 950 ₽
Всего	39 800 ₽ + 24 950 ₽	**64 750 ₽**

Это арифметический пример расчёта лицензий, а не предложение о скидке или дополнительном пакете. В него не включены отдельно приобретаемая инфраструктура, работы сверх состава поставки и внешний API по отдельным условиям. Порядок назначения разных планов сотрудникам и учёта лицензий фиксируется для корпоративной поставки.

Число выданных лицензий и число активных пользователей за день — разные показатели. Сотрудник может иметь доступ, но не отправлять запросы в конкретный день. Поэтому процент активности помогает оценивать внедрение, но сам по себе не определяет число оплачиваемых лицензий.

Доступ к моделям, контекст и расход

У тарифа есть несколько независимых характеристик:

- **Состав функций:** одинаковый у Pro и Advanced, включая презентации и таблицы.
- **Контекст:** объём информации, который обрабатывается в рамках запроса. Это не число сообщений в месяце.
- **Квоты операций:** ограничения на обращения к моделям или инструментам за период.
- **Бюджет API-ключа:** ограничение расходов внешней интеграции.
- **Политика компании:** разрешённые модели, доступные интеграции и права пользователей.

Одна задача может включать несколько обращений к модели, веб-поиск и обработку файла. Поэтому одинаковое число сообщений не гарантирует одинаковую нагрузку или расход. Разбор этих механизмов и действий при исчерпании лимита — на странице «Использование, кредиты и бюджеты».

Что передать закупкам

Для расчёта и оформления закупки подготовьте один набор исходных данных:

1. Размещение Platform: облако Nevel или контур заказчика; требуется ли изолированный профиль.
2. Число сотрудников на старте и план расширения; каким группам нужны повышенные лимиты Advanced.
3. Период закупки: месяц, год или срок корпоративного договора.
4. Внешние приложения, которым нужен API, и ожидаемая интенсивность запросов.
5. Разрешённые маршруты моделей, ресурсы для собственных моделей и ожидаемый расход у провайдеров.
6. Юридическое лицо, плательщик и ответственный за приёмку документов.

После расчёта в договоре и платёжных документах должны однозначно совпадать продукт, период, число лицензий или конфигурация установки, валюта, итоговая сумма и состав включённых работ. Порядок сверки описан в оплате и документах.

Связанные документы

- Pro и Advanced: цены и лимиты.
- Установка в контуре: конфигурации и бюджет.
- Использование, кредиты и бюджеты.
- Оплата, продление и закрывающие документы.

Источники

Цены на 8 сентября 2026 года: страница тарифов, публичный каталог активных подписок. Обзор продукта — Platform. Размещение и маршруты моделей описаны в соответствующем разделе.

Облачные планы Pro и Advanced

Одинаковые инструменты Pro и Advanced, различия в лимитах, месячные и годовые цены и выбор плана.

Облако

https://docs.nevel.sh/pricing/plans/

В облаке доступны Pro и Advanced с **одинаковыми пользовательскими возможностями и инструментами**. Основная разница между планами — лимиты использования: Advanced подходит для более интенсивной работы. Месячные цены опубликованы на странице тарифов; месячные и годовые варианты доступны в каталоге активных подписок платформы.

Цены и период оплаты

План	Месячная подписка	Годовая подписка	12 отдельных месячных оплат
Pro	1 990 ₽	19 990 ₽ за год	23 880 ₽
Advanced	4 990 ₽	49 990 ₽ за год	59 880 ₽

Публичная месячная ставка указана за сотрудника. Годовая цена — сумма за весь период, а не платёж каждый месяц. По приведённым ставкам годовой Pro дешевле 12 месячных оплат на 3 890 ₽, годовой Advanced — на 9 890 ₽. Это сравнение опубликованных вариантов без промокодов и индивидуальных условий.

Годовые варианты относятся к самостоятельной подписке в интерфейсе платформы. При корпоративной закупке количество лицензий, период, способ расчёта и итоговая сумма отражаются в договоре и счёте. Условия отдельного договора нельзя автоматически вывести из формы самостоятельной оплаты.

Одинаковые возможности, разные лимиты

Возможность	Pro	Advanced
Стандартные и премиальные модели	Включены	Включены
Чаты, ассистенты и пространства	Включены	Включены
Canvas, документы и DOCX	Включены	Включены
Веб-поиск и детальное исследование	Включены	Включены
Генерация и анализ изображений, видео	Включены	Включены
Анализ документов и голосовой ввод	Включены	Включены
MCP-интеграции и работа с подключёнными сервисами	Включены	Включены
Создание и изменение презентаций PPTX	Включены	Включены
Работа с таблицами XLSX	Включена	Включена
Работа с большими документами	Включена в пределах доступных ресурсов	Включена в пределах доступных ресурсов

Возможность	Pro	Advanced
Автоматизации	Включены	Включены
Опубликованный лимит контекста	100 000 токенов	200 000 токенов
Лимиты использования	Лимиты Pro	Повышенные лимиты Advanced

Для презентации, таблицы или документа не требуется выбирать Advanced только ради доступа к соответствующему инструменту. Тариф определяет доступный объём использования. Дополнительно действуют ограничения модели, отдельной операции и административная политика организации.

Если сотрудник не видит инструмент, проверьте выбранное пространство, права и конфигурацию. Исчерпанная квота и отсутствие права — разные причины остановки. Переход на Advanced не выдаёт административных прав и не открывает запрещённый сетевой маршрут.

Как читать ограничения

Контекст — информация для одного обращения к модели: текущий запрос, включённая история, извлечённые данные из файлов и результаты инструментов. Контекст измеряется в токенах; токен не равен одному слову, документу или сообщению. Это не месячный кошелёк и не число ответов.

В опубликованной тарифной матрице для Pro указан контекст **100 000 токенов**, для Advanced — **200 000 токенов**. Это конкретный пример различия лимитов при одинаковом наборе инструментов. Фактический контекст запроса дополнительно ограничивается возможностями выбранной модели и административной политикой: тарифный предел не увеличивает собственное окно модели и не отменяет более строгую настройку организации.

Фактические пределы зависят от модели, тарифа и административной политики. Точные квоты смотрите в действующей конфигурации рабочего пространства и уведомлении о достигнутом лимите. Число доступных запросов, видео или презентаций зависит от конфигурации установки.

Оцените типичные документы, длину диалогов и частоту операций. Короткий текстовый ответ и исследование с несколькими источниками создают разную нагрузку. Большой контекст не гарантирует точность: передавайте релевантные материалы и проверяйте результат.

Корпоративный вариант с собственным ключом

Если компания использует свой ключ к моделям, пользовательская лицензия Platform стоит **1 000 ₽ за сотрудника в месяц**. Компания отдельно оплачивает фактическое потребление в аккаунте провайдера своего ключа. Например, 25 сотрудников — 25 000 ₽ за месяц лицензий, без включения расходов провайдера в эту сумму.

Это отдельная модель корпоративной поставки, а не скидка на самостоятельную подписку Pro или Advanced. Состав включённых инструментов, подключение ключа и ограничения задаются в конфигурации компании; названия Pro и Advanced не следует автоматически переносить на этот вариант. Условия годовой оплаты и включённого потребления определяются корпоративным предложением.

Внешний API Nevel тарифицируется отдельно по правилу **OpenRouter + 20%**. Подробнее — использование, кредиты и бюджеты.

Как выбрать план для рабочих сценариев

Pro подходит как исходный план для полного набора рабочих сценариев: переписки, документов, исследований, изображений, таблиц, презентаций, интеграций и автоматизаций. Проверьте на типичных задачах, хватает ли его лимитов при реальной частоте использования.

Advanced выбирайте для сотрудников, которым регулярно нужны более высокие лимиты того же набора инструментов. Основанием перехода служит интенсивность работы: например, сотрудник выполняет много исследований, обрабатывает большие материалы или часто создаёт файлы. Сравните достигнутые квоты и планируемую нагрузку; само название формата PPTX или XLSX не определяет тариф.

Разделение планов не должно заменять управление доступом. Если сотрудник не видит интеграцию или административное действие, причиной могут быть его роль, права во внешней системе или политика рабочего пространства. Покупка более дорогого плана не делает участника администратором автоматически.

Как проверить текущий план

1. Откройте нужное рабочее пространство. Подписка и настройки относятся к выбранному пространству; при наличии нескольких пространств сначала проверьте его название.
2. В настройках найдите блок подписки. Он показывает план, состояние и дату действия либо следующего списания.
3. Нажмите **«Сменить тариф»**, чтобы увидеть доступные варианты и периоды.
4. Сравните сумму за весь период, валюту, текущий план и действие кнопки перед оплатой. Для годового варианта не принимайте отображаемый месячный эквивалент за полную стоимость года.
5. После оплаты вернитесь к блоку подписки и проверьте выбранный план и дату действия. Сохраните платёжное подтверждение.

Управление оплатой зависит от роли пользователя и конфигурации компании. Если сотрудник не управляет подпиской сам, проверку выполняет ответственный администратор. Для корпоративной поставки источник суммы и срока — документы организации.

Пробный доступ

На публичном сайте предусмотрено **7 дней без карты**. Пробный доступ позволяет проверить модели и рабочие сценарии до оплаты. Он имеет собственные ограничения; бесплатный период не следует интерпретировать как неограниченное потребление любых инструментов.

Для проверки пилота используйте 3–5 типичных задач компании: один документ, исследование вопроса, таблицу или презентацию на выбранном плане, а также нужную интеграцию. Сохраняйте не только удачные результаты, но и причины остановки: квота, отсутствие права, недоступная модель или недостаточный контекст. Это помогает выбрать план по фактическому сценарию.

Пробный доступ через сайт и пилот установки в контуре имеют разный состав. Срок и условия проекта в контуре определяются отдельно; обещание семидневной самостоятельной подписки не задаёт срок такого внедрения.

Частые вопросы

Нужно ли платить отдельно за каждую модель?

Доступ к стандартным и премиальным моделям входит в опубликованный состав обоих планов. Использование остаётся в пределах возможностей и квот плана, разрешений компании и доступного каталога.

Advanced снимает все ограничения?

Нет. Он повышает лимиты использования при том же наборе возможностей. Ограничения конкретной модели, инструмента, политики компании и внешних API сохраняются. Подробнее — использование и бюджеты.

Можно ли оплатить в рублях?

Да, для российской версии опубликованы рублёвые планы. Доступный способ оплаты показан на шаге оформления. Для корпоративного безналичного расчёта используйте порядок оплаты и документов.

Источники

Цены и периоды на 8 сентября 2026 года: тарифы, активные месячные и годовые планы в публичном API. Для самостоятельной оплаты также действует публичная оферта. Параметры конкретных лимитов доступны в конфигурации рабочего пространства.

Стоимость установки в контуре

Опубликованные годовые конфигурации, состав бюджета и расчёт инфраструктуры заказчика.

В контуре

https://docs.nevel.sh/pricing/onprem/

Для установки Nevel в инфраструктуре заказчика опубликованы три годовые конфигурации. Цена зависит от диапазона пользователей и состава решения. Значение **«от»** — начальная стоимость конфигурации, а не фиксированная цена любого внедрения внутри указанного диапазона. Публичные тарифы Nevel.

Годовые конфигурации

Конфигурация	Диапазон пользователей	Опубликованная стоимость
Стандартное развёртывание	До 100 пользователей	От 2,5 млн ₽ в год
Корпоративное развёртывание	До 1 000 пользователей	От 5 млн ₽ в год
Enterprise с GPU	1 000+ пользователей	От 15 млн ₽ в год

Стоимость оборудования не включена. Требования к инфраструктуре предоставляются для выбранной конфигурации. В частности, название Enterprise с GPU не означает передачу серверов или GPU в собственность заказчика в пределах указанной суммы.

Что входит в предмет поставки

В публичном описании установки указаны развёртывание в инфраструктуре заказчика, контроль над данными, приватные точки подключения моделей, базовая настройка и выделенная команда поддержки. Состав внедрения необходимо читать вместе с архитектурой поставки: где находятся сервисы Platform, где выполняются модели, какие интеграции подключены и кто обслуживает каждый компонент.

Установка в контуре определяет размещение Platform, а не обязательное размещение всех моделей. Заказчик может использовать собственные модели, российских провайдеров, например Яндекс и Сбер, или глобальные модели по административной политике. Для внешних маршрутов в конфигурации явно задаются точки подключения, сетевые потоки и правила передачи данных.

Изолированный профиль — частный случай установки в контуре: модели, инструменты и источники находятся внутри разрешённого периметра, внешние обращения отключены. При сравнении с облаком учитывайте также self-hosted DeepSeek Nevel на серверах в РФ: это размещение модели у Nevel, а не передача локального модельного сервиса заказчику. Варианты установки и маршруты моделей.

На публичной тарифной странице также упоминается базовое обучение при развёртывании. Отдельные образовательные программы не включаются в бюджет внедрения автоматически: учитывайте состав выбранной поставки.

Из чего складывается полный бюджет

Статья	Что учитывать	Кто предоставляет исходные данные
Лицензия и конфигурация Nevel	Число пользователей, срок, компоненты Platform и Guardrail/API	Nevel и закупки заказчика

Статья	Что учитывать	Кто предоставляет исходные данные
Серверная инфраструктура	Вычислительные узлы, память, диски, сеть и размещение	ИТ и инфраструктурная команда
Выполнение моделей	Ресурсы собственных моделей и/или потребление разрешённых российских и глобальных провайдеров, параллельность и контекст	ИТ совместно с владельцами сценариев
Хранение и резервное копирование	Файлы, журналы, резервные копии и требуемый срок хранения	ИТ и ИБ
Интеграции	Подключаемые системы, их доступность, права и условия лицензирования	Владельцы систем заказчика
Эксплуатация	Мониторинг, обновления, резервирование и восстановление	Стороны по матрице ответственности

Эта таблица перечисляет статьи расчёта; она не утверждает, что каждая статья оплачивается отдельным счётом Nevel. Часть ресурсов уже может быть у заказчика. В смете нужно явно указать, какие ресурсы предоставляет заказчик, какие работы входят в поставку и что приобретается дополнительно.

Как определить ресурсы и расходы на модели

Число лицензий описывает доступ людей к платформе. Вычислительная нагрузка определяется одновременными задачами, моделями и инструментами. Если модели выполняются у заказчика, для них нужны собственные мощности; если запрос направлен провайдеру, он использует ресурсы провайдера и учитывается по условиям этого подключения. Сама установка Platform в контуре не означает обязательную закупку GPU для каждой модели.

Сто сотрудников с короткими запросами в разное время и сто сотрудников с одновременной обработкой больших документов создают разную нагрузку. Для смешанного набора маршрутов отдельно оцените внутреннюю мощность и внешний расход.

Для расчёта инфраструктуры соберите:

1. Число активных пользователей в обычный рабочий час и ожидаемый пик.
2. Типовые задачи: чат, обработка документов, изображения, презентации, внешние API-вызовы.
3. Какие модели должны выполняться у заказчика, какие разрешены у российских или глобальных провайдеров и что требуется от результата.
4. Типичный и максимальный объём контекста, размер файлов и длительность операций.
5. Ожидаемую параллельность запросов от сотрудников и автоматизаций.
6. Требования к резервированию и восстановлению.

По этим данным составляется конфигурация. После развёртывания её проверяют на согласованных рабочих сценариях и нагрузке. Название тарифного уровня само по себе не задаёт гарантированное время ответа модели.

Контекст и лимиты в контуре

На публичной странице для On-Premise в строке контекста указано «Без лимитов». Для инженерного расчёта это нельзя читать как бесконечный контекст любой модели: у модели, памяти и вычислительных ресурсов есть технические границы. В проекте указываются фактическая модель, её доступный контекст и принятые ограничения эксплуатации.

Административные лимиты полезны при любом маршруте. Они распределяют собственные ресурсы, ограничивают внешний расход и не позволяют одной автоматизации занять всю доступную мощность. У локальных моделей есть инфраструктурная стоимость; у провайдерских — условия потребления. Не делайте вывод об отсутствии внешнего счёта только из того, что Platform установлена в контуре.

Как сравнить контур с облаком

Сравнивайте одинаковый срок, рабочие сценарии и маршруты моделей. Для облака учитываются лицензии сотрудников и отдельно определённые условия внешнего API. Для контура — годовая лицензия, инфраструктура, эксплуатация и расходы выбранных маршрутов: собственные мощности и/или провайдерское потребление. Если оборудование приобретается на несколько лет, отделяйте разовую закупку от ежегодных платежей.

Pro и Advanced в облаке предлагают одинаковые инструменты и различаются лимитами. Поэтому сравнение с контуром должно включать интенсивность использования, требования к данным и маршруты обработки, а не предположение, что таблицы или презентации доступны только на одном облачном плане.

Не делите начальную годовую стоимость на максимальное число сотрудников без оговорок: такое число не учитывает фактическое заполнение лицензий, состав внедрения и ресурсы для моделей. Практическое сравнение должно показывать полную стоимость выбранного сценария и требования к контролю данных, ради которых выбирается контур.

Что проверить в предложении на внедрение

Проверьте, что в предложении указаны диапазон пользователей, срок лицензии, состав Platform, разрешённые маршруты моделей, ресурсы заказчика, провайдерские расходы, работы по установке и критерии приёмки. Если требуется изолированный профиль, зафиксируйте отсутствие внешних обращений для всего сценария. Отдельно должны быть понятны обновления, резервные копии, поддержка и расширение мощности или числа пользователей.

Технические шаги подготовки площадки описаны в установке в контуре, а состав платёжного пакета — в оплате и документах.

Источники

Цены и состав конфигураций на 8 сентября 2026 года: тарифы Nevel, предложение для компаний, архитектура и безопасность.

Использование, кредиты и бюджеты

Как различаются квоты, контекст и лимиты расходов; как считать API по цене OpenRouter плюс 20%.

Облако / В контуре

https://docs.nevel.sh/pricing/usage/

Цена лицензии, квота операций и денежный бюджет API — разные ограничения. Для внешнего API Nevel действует стоимость **OpenRouter + 20%**. Для сотрудников доступны публичные планы Pro и Advanced либо корпоративный вариант **1 000 ₽ за человека в месяц с ключом заказчика**. Во втором варианте расходы по собственному ключу провайдера оплачивает заказчик.

Четыре вида ограничений

Ограничение	Что измеряется	Что происходит при достижении
Контекст	Объём данных для одного обращения к модели, в токенах	Нужно уменьшить входные данные, изменить способ работы с документом или выбрать подходящую модель/план
Квота операции	Число обращений к модели или инструменту за установленный период	Операция ограничивается до обновления соответствующей квоты
Лимит расходов API-ключа	Денежный объём запросов по конкретному ключу	Новые запросы останавливаются при исчерпании бюджета
Ограничение доступа	Роль пользователя, разрешённая модель, функция или интеграция	Действие недоступно до изменения соответствующего разрешения

Установка большего бюджета API не увеличивает контекст модели и не выдаёт пользователю новые права. Аналогично, годовая оплата подписки продлевает срок доступа, но не превращает суточную квоту в общий годовой запас операций.

Что расходует квоты

Запрос пользователя может включать обращение к модели, поиск в интернете, поиск по вложениям, создание изображения, работу с таблицей или другой инструмент. Эти операции могут иметь самостоятельные ограничения. Платформа различает операции и область применения квоты: рабочее пространство, сотрудника внутри пространства либо API-ключ.

Из этого следуют два практических правила:

1. По числу отправленных сообщений нельзя точно определить оставшееся число любых задач. Одно сообщение может запускать несколько действий.
2. Ограничение одного инструмента не означает, что недоступен весь аккаунт. Проверяйте название ограничения в сообщении интерфейса и условия текущего плана.

В интерфейсе различаются окончание подписки и достижение дневного лимита. При дневном ограничении показывается уведомление об обновлении лимита; при окончании подписки — предложение выбрать платный план. Не повторяйте одну и ту же операцию многократно, если причина остановки уже указана: сначала определите тип ограничения.

Pro и Advanced имеют одинаковый набор возможностей и инструментов; различаются лимиты использования. Конкретные квоты зависят от плана, выбранной модели и административной политики; не переводите цену подписки в гарантированное число запросов по произвольному курсу.

Бюджет внешнего API

Стоимость внешнего API рассчитывается по фактической стоимости OpenRouter с наценкой Nevel 20%:

```
Стоимость API Nevel = стоимость OpenRouter × 1,20
```

Если учтённая стоимость OpenRouter для набора запросов составила 100 USD, стоимость API Nevel составит 120 USD. Это пример применения коэффициента, а не цена конкретной модели или обещание определённого числа токенов. Расценки моделей меняются, а разные запросы имеют разный входной и выходной объём.

Для предварительной оценки используйте тариф именно выбранной модели и фактическую структуру запросов. Цена входных токенов и цена выходных токенов могут отличаться; кроме них у провайдера могут учитываться отдельные операции. После выполнения ориентируйтесь на учтённую стоимость запросов в аналитике. Каталог моделей и расценки OpenRouter.

Лимит ключа относится к расходам API Nevel. Не прибавляйте 20% к уже отображаемой в аналитике Nevel сумме второй раз. Рублёвые лицензии сотрудников и долларовый показатель расходов API учитывайте раздельно; порядок рублёвых расчётов по корпоративному договору определяется платёжными условиями, а не произвольным курсом внутри таблицы.

Как задать лимит API-ключа

Эти действия выполняет администратор или владелец, которому доступно управление API-ключами.

1. Откройте настройки нужного рабочего пространства и страницу **«API ключи»**.
2. Для нового подключения нажмите **«Добавить ключ»**. Для существующего ключа выберите **«Редактировать»** в его меню.
3. Задайте понятное название: например, «Сводка обращений» или «Тестирование интеграции». Разные приложения удобнее учитывать отдельными ключами.
4. В поле **«Кредитный лимит, USD»** укажите положительную сумму. Можно использовать до шести знаков после запятой. Пустое поле означает отсутствие ограничения расходов именно по этому ключу.
5. Выберите период сброса. Сумма и период задаются совместно.
6. Сохраните изменения и проверьте строку ключа: лимит, остаток и дату следующего сброса.
7. При создании нового ключа сохраните секрет в защищённом хранилище приложения: полностью он показывается один раз. В отчёте о бюджете используйте название или префикс, а не сам секрет.

Период	Когда начинается новый бюджет
Ежедневно	Каждый день в 00:00 UTC
Еженедельно	В понедельник в 00:00 UTC
Ежемесячно	Первого числа месяца в 00:00 UTC
Никогда	Автоматического сброса нет

Время сброса прямо указано в интерфейсе. Не путайте его с датой оплаты лицензии. Бюджет в 120 USD с ежемесячным сбросом означает ограничение расходов ключа в календарном месяце UTC, а не предоплаченный кошелёк с автоматическим переносом неиспользованного остатка.

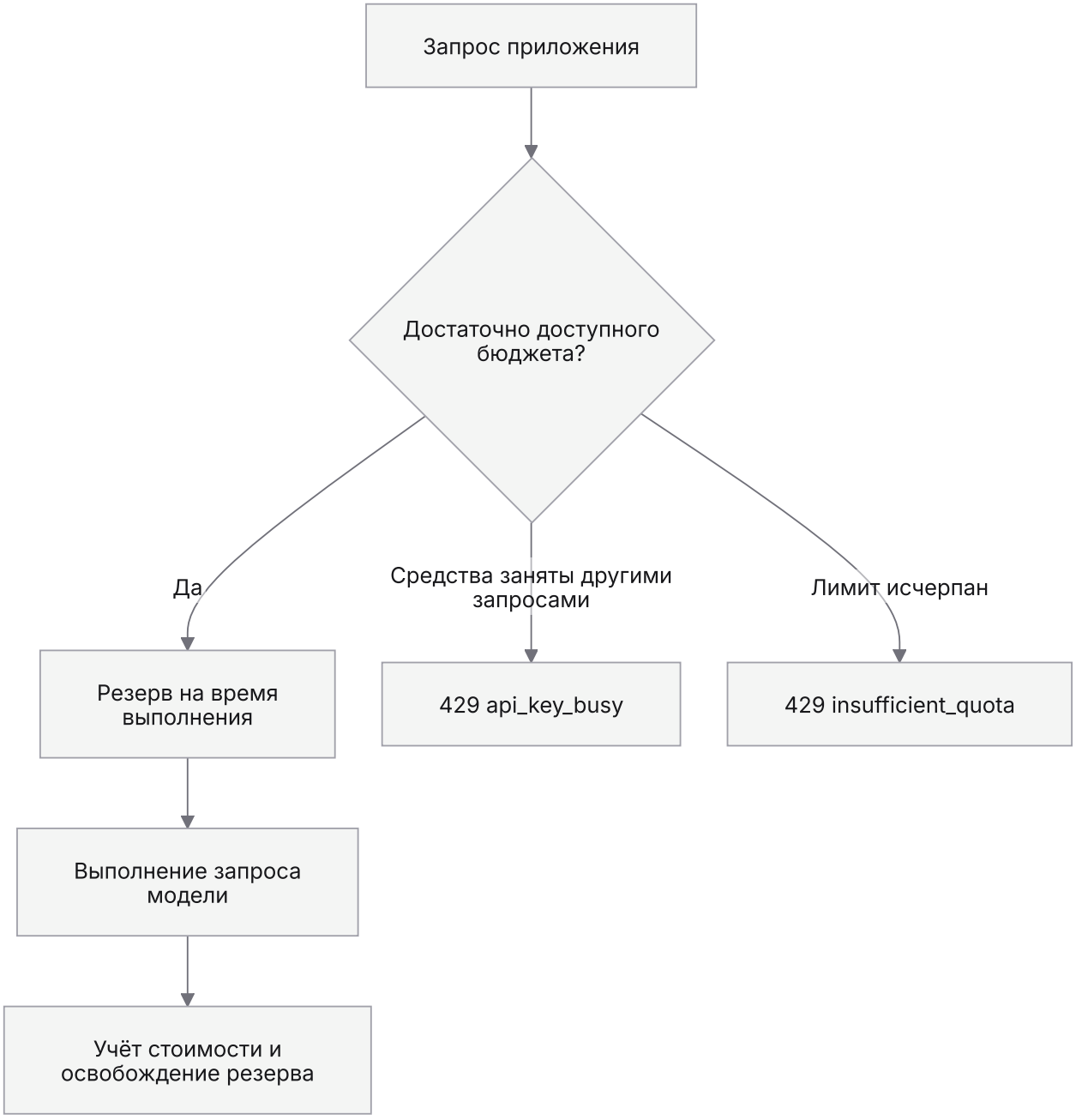
Отсутствие лимита ключа не отменяет ограничения моделей, прав доступа, квот и инфраструктуры. Для промышленной интеграции отдельный лимит полезен даже тогда, когда общий бюджет компании достаточен.

Потрачено, зарезервировано и осталось

Перед отправкой API-запроса платформа резервирует доступную часть бюджета. После получения стоимости от провайдера резерв учитывается по фактическому расходу. Это позволяет ограничивать расход и при параллельной работе нескольких запросов.

СХЕМА 1

Бюджет запроса API



Доступный остаток определяется так:

```
Остаток = лимит − учтённый расход − активные резервы
```

Например, при лимите 120 USD, учтённом расходе 72 USD и активных резервах 18 USD для новых запросов доступно 30 USD. Зарезервированная сумма ещё не означает окончательную стоимость незавершённых запросов.

Если заранее невозможно надёжно ограничить стоимость запроса, на время его выполнения может резервироваться весь доступный остаток. Поэтому новый параллельный запрос иногда получает временную остановку, хотя лимит ещё не потрачен окончательно. Завершение или обрыв клиентского соединения также не гарантирует мгновенного освобождения резерва: системе нужно установить результат и стоимость операции.

Что делать при остановке запросов

Признак	Значение	Действие
`429`, `api_key_busy`	Доступный бюджет зарезервирован незавершёнными запросами	Соблюдайте `Retry-After`, сократите параллельность, дождитесь завершения запросов
`429`, `insufficient_quota`	Лимит исчерпан либо не позволяет покрыть оценку запроса	Проверьте расход и период сброса; уменьшите запрос или измените согласованный бюджет
В строке ключа «Лимит исчерпан»	Бюджет ключа недоступен для новых операций	Сверьте сумму, расход, резервы и следующую дату сброса
Ключ отключён или истёк	Проблема доступа к API	Проверьте состояние и срок ключа; изменение денежного лимита не исправит его статус

Повторять запрос с `insufficient_quota` без изменения условий бессмысленно. Для временного `api_key_busy` используйте ожидание и ограниченные повторы; не запускайте бесконечный цикл. Другие ошибки рассматриваются в справочнике ошибок API.

Аналитика для контроля экономики

В аналитике API доступны периоды **неделя, месяц и год**, а также группировки **по модели, ключу и создателю**. Сравниваются расходы, число запросов и токены. Это позволяет ответить на разные вопросы:

- По модели: какая модель формирует основную часть расходов.
- По ключу: какое приложение или автоматизация создаёт нагрузку.
- По создателю: кто отвечает за группу подключений.
- По токенам и запросам: связан ли рост расходов с числом обращений или их размером.

Для ежемесячной сверки выберите одинаковый период во всех сравниваемых отчётах, разделите ключи тестовых и рабочих приложений и сопоставьте рост расходов с изменениями сценариев. Если число запросов не выросло, проверьте увеличение контекста, выходных ответов или переход на другую модель.

Аналитика активности сотрудников показывает использование платформы и групп; это показатель внедрения. Она не измеряет автоматически сэкономленные человеко-часы. Для расчёта эффекта отдельно зафиксируйте длительность задачи до и после внедрения и долю результатов, прошедших проверку сотрудником.

Вариант с ключом заказчика

При корпоративной лицензии **1 000 ₽ за сотрудника в месяц с собственным ключом** компания оплачивает Platform и отдельно несёт расходы по своему ключу провайдера. Например, 25 сотрудников дают 25 000 ₽ месячной стоимости лицензий; стоимость моделей добавляется по фактическим расходам в аккаунте провайдера. Пересчёт валют и расходы провайдера не включаются в эту иллюстрацию.

Это отдельная модель поставки пользовательского доступа. Наценка внешнего Nevel API +20% не добавляется автоматически к расходам этого собственного ключа: правило применяется к запросам внешнего Nevel API как отдельному сценарию. Условия годовой оплаты BYOK определяются корпоративным предложением. При подключении определите владельца ключа, разрешённые модели, бюджет у провайдера и порядок ротации; ключ хранится в защищённой конфигурации, а не в документах, промптах или общем чате.

Размещение Platform и расходы на модели учитываются отдельно. Platform в контуре может использовать собственные модели заказчика, российских провайдеров — например Яндекс и Сбер — и глобальные модели по разрешённым маршрутам. В облаке Nevel доступна также self-hosted DeepSeek на серверах в РФ. Название маршрута само по себе не задаёт дополнительную цену: используйте применимую модель расчёта вашей поставки.

Для собственных моделей расходы могут быть инфраструктурными: GPU, память и эксплуатация. Для провайдерских маршрутов учитывается потребление по условиям подключения. Правило OpenRouter × 1,20 относится к внешнему Nevel API; не переносите его автоматически на собственные вычислительные ресурсы или пользовательский BYOK. **Изолированный профиль** использует только внутренние ресурсы и является отдельной конфигурацией on-prem. Бюджет контура.

Источники и связанные материалы

Условия на 8 сентября 2026 года: публичные тарифы Nevel, возможности API и аналитики, каталог OpenRouter. Подключение описано в API: быстрый старт и ключах и лимитах.

ТАРИФЫ И ОПЛАТА

Оплата, продление и документы

Самостоятельная подписка и корпоративная закупка: суммы, состояния оплаты, отмена продления и закрывающий пакет.

Облако / В контуре

https://docs.nevel.sh/pricing/billing/

Самостоятельная облачная подписка оформляется через интерфейс платформы. Для корпоративной закупки используются договор и согласованный платёжный порядок. Эти варианты нужно различать: дата автоматического списания самостоятельной подписки и отчётный период корпоративного договора могут не совпадать.

Самостоятельная оплата

Согласно публичной оферте Nevel, самостоятельная лицензия предоставляется на условиях 100% предоплаты; другие условия, включая тестовый доступ, могут быть указаны на сайте. Срок лицензии связан с оплаченным периодом. Уже оплаченное вознаграждение не изменяется.

Перед оплатой:

1. Откройте нужное рабочее пространство и текущую подписку.
2. Выберите Pro или Advanced и период — месяц либо год.
3. Проверьте полную сумму списания, валюту и срок. Годовая цена оплачивает весь годовой период.
4. Если применяете промокод, проверьте, что итоговая цена действительно изменилась и соответствует выбранному плану.
5. Завершите оплату доступным для аккаунта способом и сохраните подтверждение.
6. В блоке подписки проверьте название плана, статус и дату следующего списания либо окончания доступа.

В российском каталоге доступны рублёвые планы. Платёжные способы зависят от доступного платёжного оформления; используйте показанный в нём вариант. Не переводите деньги по реквизитам из переписки, которые не связаны с действующим договором или счётом компании.

Как читать состояние подписки

Что видно	Что это означает	Что проверить
План и дата следующего списания	Действует подписка с продлением	Сумму, период и способ оплаты
«Ошибка оплаты» / «Не удалось продлить»	Продление не прошло, может быть назначена следующая попытка	Платёжный инструмент и дату следующей попытки
«Автопродление отключено», «Доступ до…»	Следующее продление отменено, доступ сохраняется до указанной даты	Дату окончания и необходимость нового периода
Окончание подписки	Оплаченный период закончился	Нужен ли новый план или корпоративное продление

Не определяйте успех оплаты только по закрытию платёжного окна. Подтверждением служит согласованное состояние подписки и платёжный документ. Если платёж проведён, а план не обновился, сохраните номер и время операции, сумму, пространство и сообщение об ошибке. Передавайте поддержке эти сведения без номера карты целиком, кода безопасности или платёжного токена.

Отмена автоматического продления

Для подписки с автоматическим списанием откройте управление подпиской и выполните отмену продления. После подтверждения проверьте состояние **«Автопродление отключено»** и дату **«Доступ до...»**. Отключение продления сохраняет доступ до конца действующего периода.

Отмена следующего списания и возврат уже уплаченных средств — разные действия. Не считайте уведомление об отключении продления подтверждением возврата. При переходе между планами проверяйте сумму и начало нового периода на шаге подтверждения; не переносите правила самостоятельного перехода на корпоративный договор автоматически.

Для корпоративной поставки прекращение или продление регулируются договором. Удаление пользователя, отзыв API-ключа и отключение продления личной подписки не являются взаимозаменяемыми действиями.

Корпоративный платёжный пакет

Для закупки организацией подготовьте данные, по которым можно однозначно выпустить и сверить документы:

Данные	Для чего нужны
Полное наименование и реквизиты плательщика	Идентификация стороны и оформление документов
Номер договора и основание заказа	Связь платежа с согласованной поставкой
Продукт и конфигурация	Размещение Platform, пользовательский план или BYOK, внешний API и разрешённые маршруты моделей
Количество лицензий и период	Расчёт стоимости пользовательского доступа
Состав внедрения	Разделение лицензии, работ и ресурсов заказчика
Получатель документов и способ обмена	Доставка счёта и закрывающего пакета

Платёжный пакет может включать договор с приложениями, счёт и УПД либо акт в соответствии с выбранным оформлением. Обмен через ЭДО и его параметры закрепляются для корпоративного договора. В частности, заранее укажите оператора ЭДО и идентификатор организации, если стороны используют такой обмен.

Публичная оферта описывает предоплату самостоятельной подписки. Не выводите из неё срок выставления УПД по индивидуальному договору или обязательную постоплату для всех компаний. Эти условия должны быть прочитаны в подписанном документе конкретной закупки.

Как сверять счёт и закрывающие документы

1. Сопоставьте юридическое лицо, договор и период с внутренним заказом на закупку.
2. Проверьте модель расчёта: цена за сотрудника, собственный ключ заказчика, годовая конфигурация контура или использование внешнего API.
3. Сверьте количество лицензий и правило их учёта. Активность в аналитике и число предоставленных лицензий не обязательно равны.
4. Проверьте цену единицы, валюту, итоговую сумму и налоговое оформление.
5. Для API сопоставьте период и состав учтённых запросов с аналитикой, учитывая, что отображаемые расходы Nevel уже включают применимый коэффициент.
6. Для внедрения проверьте, какие работы и результаты принимаются документом.
7. При расхождении направьте конкретную строку документа, ожидаемое значение и основание расчёта ответственному по договору до завершения внутренней приёмки.

В публичной оферте указано, что вознаграждение не облагается НДС. Для бухгалтерского учёта конкретной корпоративной закупки используются налоговые реквизиты и суммы её договора и платёжных документов; эта страница не заменяет документальное основание операции.

Отдельный учёт API и ключа заказчика

Для внешнего API Nevel действует правило **стоимость OpenRouter + 20%**. Например, 100 USD фактической стоимости OpenRouter соответствуют 120 USD стоимости Nevel. Перед сверкой определите, какой именно период и набор ключей включён в расчёт, и не применяйте наценку второй раз к уже рассчитанной сумме Nevel.

Для пользовательского доступа с ключом заказчика месячная цена лицензии — **1 000 ₽ за человека**. При 25 сотрудниках это 25 000 ₽ за месяц плюс расходы на модели в аккаунте провайдера заказчика. Собственный ключ не означает, что расходы провайдера входят в рублёвую лицензию Platform. Условия годовой оплаты и включённого потребления определяются корпоративным предложением.

BYOK и внешний Nevel API имеют разные основания расчёта. Не прибавляйте наценку API к расходам собственного ключа пользовательской поставки автоматически. Если организация отдельно подключила внешний Nevel API для приложения, его запросы учитываются по правилу ×1,20 отдельно от лицензий сотрудников.

Подробные определения расходов, резервов и периодов лимита приведены в использовании и бюджетах.

Что подготовить к продлению корпоративного договора

Зафиксируйте число сотрудников и планы на следующий период, нужные лимиты, новые интеграции, использование API и изменения инфраструктуры. Pro и Advanced предоставляют одинаковые инструменты; пересмотр плана связан прежде всего с объёмом использования.

Для установки в контуре отдельно перечислите маршруты моделей: собственные ресурсы заказчика, российские провайдеры или глобальные модели по политике компании. Изменение маршрута может менять структуру затрат и требования к эксплуатации даже без переноса Platform. Для изолированного профиля сохраняется условие внутреннего выполнения всего сценария. Это позволяет связать продление с фактическим составом решения.

Текущие цены доступны в планах Pro и Advanced и конфигурациях установки в контуре. Юридические материалы собраны в разделе «Документы для согласования».

Источники

Условия оплаты на 8 сентября 2026 года: публичная оферта, тарифы, активные планы самостоятельной подписки.

ЮРИДИЧЕСКИЕ ВОПРОСЫ

152-ФЗ: согласование обработки данных

Пошаговая подготовка к использованию Platform и Guardrail с персональными данными и официальные источники РФ.

Облако / В контуре

https://docs.nevel.sh/legal/152-fz/

Nevel помогает организовать контролируемую обработку запросов, но решение о допустимости конкретного сценария принимается по его данным, участникам и настройкам. Для согласования юридическая служба, ИБ и ИТ определяют разрешённые операции, состав документов и приёмочные проверки.

Сначала опишите процесс

Не согласовывайте «использование искусственного интеллекта вообще». Выберите операцию: например, подготовку ответа по обращению клиента, анализ договора или поиск по внутренним документам. Для неё заполните карточку.

Поле	Вопрос для согласования
Цель	Для какого результата обрабатываются сведения
Люди	Чьи данные встречаются: сотрудники, клиенты, представители контрагентов
Состав	Какие конкретные поля и документы передаются
Основание	На каком основании выполняется каждая операция
Участники	Кто определяет цель, кто действует по поручению, кто получает данные
География	Где выполняются обработка, хранение и резервирование
Доступ	Какие пользователи, приложения и исполнители получают доступ
Сроки	Когда прекращается обработка и как очищаются все экземпляры
Проверка	Какие доказательства нужны перед запуском

Сохраните карточку рядом со схемой потоков данных. При добавлении модели, подключения или нового типа документов обновляйте обе записи.

Роли и основание обработки

Статья 3 определяет оператора через цели, состав данных и действия. Статья 6 устанавливает основания обработки и условия поручения другому лицу; согласие — одно из оснований. Статья 9 требует отдельного оформления согласия от иных подтверждаемых субъектом документов. Официальный текст 152-ФЗ.

Для внедрения составьте матрицу «участник — данные — операция — основание». Роли определяйте отдельно для содержимого рабочих документов, регистрации пользователей, расчётов и поддержки. Не считайте, что один договор автоматически описывает все эти процессы одинаково.

Если Nevel обрабатывает сведения по поручению заказчика, согласуйте поручение и доказательства исполнения до передачи данных. Для on-prem отдельно рассмотрите обслуживание и разрешённые модельные маршруты: доступ инженера, передача диагностического архива или вызов внешнего провайдера может добавить участника обработки. Размещение платформы у заказчика само по себе не определяет состав всех получателей.

Локализация и внешние модели

Часть 5 статьи 18 ограничивает перечисленные в ней операции при сборе данных граждан РФ с использованием зарубежных баз, с установленными исключениями. Локализация и трансграничная передача оцениваются отдельно; статья 12 предусматривает предварительное уведомление о намерении такой передачи. 152-ФЗ, изменения № 23-ФЗ от 28.02.2025.

Практический порядок для внешнего модельного маршрута в облаке и on-prem:

1. Получите перечень сервисов, стран и обрабатываемых полей по каждому маршруту.
2. Разберите операции до модели: загрузку, распознавание, индексацию, диагностику и копирование.
3. Установите, остаются ли передаваемые сведения персональными после обработки Guardrail.
4. Проверьте актуальную редакцию правил передачи, статус страны, требования к уведомлению и возможные ограничения.
5. Разрешите маршрут только после фиксации решения ответственного лица.

Место установки платформы и маршрут модели фиксируются отдельно. В on-prem платформа находится в инфраструктуре заказчика, а политика может разрешать локальную модель, российский API, self-hosted модель Nevel на серверах РФ или глобального провайдера. Например, DeepSeek на серверах Nevel в РФ и внешний API модели с таким же названием имеют разные цепочки получателей. Матрица размещения и маршрутов помогает описать конкретную конфигурацию.

Передача из инфраструктуры заказчика российскому получателю является выходом за её границу, но сам этот факт не устанавливает трансграничную передачу. География оценивается по фактическим обработчикам, площадкам и операциям всей цепочки, включая поддержку и резервирование. Страна регистрации бренда или адрес интерфейса не заменяет эту схему.

Для отдельного изолированного профиля подтвердите нахождение всех обработчиков в контуре и исключите внешнее переключение при сбое. Замена имён маркерами сама по себе не является основанием считать любой внешний вызов допустимым. Подробности — Guardrail.

Изменения правил передачи с 26 июля 2026 года

Федеральный закон № 265-ФЗ изменил статью 12: при включении страны в перечень учитываются её регулирование и применяемые меры защиты, соответствующие положениям Конвенции. Из пункта 2 части 5 и частей 10–11 исключены положения, отдельно выделявшие участие государства в Конвенции. Поправка к статье 12 действует с 26 июля 2026 года — даты официального опубликования; отложенный срок для статьи 2 закона № 265-ФЗ к этой поправке не относится. Официальная публикация № 265-ФЗ, PDF: статья 1 на страницах 1–2, статья 7 на страницах 5–6.

При согласовании не используйте прежнее участие страны в Конвенции как самостоятельное достаточное условие для выбора порядка передачи. Проверьте действующий перечень Роскомнадзора, сведения о получателе и применимый порядок по актуальной статье 12. Проверка самой поправки не подтверждает актуальность конкретного списка стран или допустимость отдельного вызова модели.

Маскирование не заменяет оценку передачи

В Guardrail идентификаторы заменяются масками, а соответствия позволяют восстановить разрешённый ответ. Это конкретный технический механизм уменьшения раскрытия. Для юридической оценки отдельно рассматриваются возможность определения человека по оставшемуся контексту, доступ к соответствиям, роли участников и требования к выбранному методу. Нельзя автоматически заключить ни что любая обратимая замена запрещена, ни что замена ФИО освобождает от всех требований к персональным данным.

Актуальные требования к методам обезличивания, включая введение идентификаторов, разобраны со ссылкой на официальный приказ Роскомнадзора № 140 в мерах защиты РФ. При внешнем маршруте проверяется вся цепочка: для API через OpenRouter этот участник учитывается наряду с конечным провайдером. Использование своего ключа модели не меняет географию обработки и правовой статус содержания.

Техническая политика должна выражать конкретную границу запрета. Не используйте слово «внешний» одновременно для чужой инфраструктуры и другой страны.

Правило для категории	Допустимое техническое решение
Только в инфраструктуре заказчика	Локальная обработка всеми затронутыми компонентами; при её недоступности — отказ. Российское облако тоже находится за этой границей

Правило для категории	Допустимое техническое решение
Только согласованные получатели и обработка в РФ	Проверенный маршрут к локальной модели, self-hosted Nevel в РФ или российскому провайдеру, если вся требуемая цепочка соответствует решению
Разрешён согласованный глобальный маршрут	Только указанная цепочка получателей после оценки применимых требований и фиксации решения; название модели не даёт общего разрешения

Правило категории действует независимо от обнаружения конкретных сущностей детектором. Резервный маршрут заранее включается в разрешённый набор, повторно проверяется для данных и фиксируется в аудите. Скрытое переключение на другого получателя запрещено. Приёмка включает отказ основного маршрута и проверку того, что данные не уходят за разрешённую границу. Технические сценарии описаны в Guardrail и потоках данных.

Меры защиты и организационные документы

Для подготовки проекта используйте рабочую таблицу требований. Она связывает задачу с конкретным результатом, который можно проверить.

Задача	Результат для комплекта внедрения
Классифицировать информацию	Перечень данных, субъектов и допустимых сценариев
Определить требования к системе	Заключение о применимых нормах и модель угроз
Ограничить доступ	Матрица ролей, процедура выдачи и отзыва, тесты изоляции
Управлять получателями	Реестр моделей, обработчиков и разрешённых маршрутов
Настроить эксплуатацию	Политики хранения, резервирования, удаления и реагирования
Работать с обращениями	Ответственный, регистрация запроса, поиск связанных данных
Подтвердить запуск	Протокол испытаний с конфигурацией и доказательствами

Для информационных систем персональных данных при определении требований рассматривается постановление Правительства РФ № 1119. Для систем государственных органов, государственных унитарных предприятий и учреждений отдельно проверьте применимость приказа ФСТЭК России от 11.04.2025 № 117. Приказ № 117 действует с 1 марта 2026 года вместо приказа № 17; оговорка о ранее выданных аттестатах и карта технических мер приведены в разборе требований РФ. Уровень защищённости и необходимый набор мер нельзя назначать по одному названию продукта.

Инциденты и обращения субъектов

При событии, подпадающем под часть 3.1 статьи 21, уведомление регулятора об инциденте направляется в течение 24 часов, о результатах внутреннего расследования — 72 часов; отсчёт связан с выявлением указанного в норме инцидента. Статья 21 152-ФЗ.

В договоре и инструкции поддержки согласуйте эскалацию так, чтобы у заказчика оставалось время на оценку и уведомление. Включите контакты, замещение, обязательные сведения, способ передачи доказательств и действия вне рабочего времени. Срок ответа технической поддержки нельзя автоматически считать допустимым сроком уведомления регулятора.

Для обращений субъектов подготовьте поиск по всем связанным данным, проверку полномочий заявителя, исполнителя и фиксацию результата. Процедуру удаления соедините с резервными копиями и внешними получателями. Требования к подтверждению уничтожения закреплены в приказе Роскомнадзора № 179; технический порядок описан в хранении и удалении.

Итог согласования

Результат — разрешение на конкретный сценарий с указанными данными, маршрутами, настройками и ограничениями. К нему прикладывают договорный пакет, схему, матрицу доступа и протокол приёмки. Новые категории данных, зарубежный получатель или иной способ размещения требуют повторной оценки затронутой части.

Нормативная база и область применения

Дата нормативной справки: 8 сентября 2026 года. Базовая ссылка ведёт на официальный сводный текст в состоянии на 1 сентября 2025 года; для статьи 12 дополнительно учитывается закон от 26.07.2026 № 265-ФЗ. Эта нормативная база не является полной действующей редакцией всех применимых актов. Перед согласованием внешней передачи нужны актуальные условия статьи 12, действующий перечень стран Роскомнадзора, последующие изменения закона и применимые подзаконные и отраслевые требования.

Официальные источники: официальный текст 152-ФЗ, № 23-ФЗ, № 156-ФЗ, № 265-ФЗ, приказ ФСТЭК № 117, приказ Роскомнадзора № 179. Изменения, применимость к отрасли и итоговую редакцию комплекта проверяет согласующий юрист заказчика.

ЮРИДИЧЕСКИЕ ВОПРОСЫ

Документы для закупки и согласования

Состав договорного пакета, поручения на обработку данных и материалов для юридической службы, ИБ и ИТ.

Облако / В контуре

https://docs.nevel.sh/legal/documents/

Для согласования Nevel соберите единый комплект, в котором договорные условия совпадают с технической схемой и порядком эксплуатации. Юристы, закупки, ИБ и ИТ должны работать с одной версией решения и согласованными документами.

Обязательства сторон, стоимость, состав лицензий, сроки, ответственность и условия поддержки устанавливаются договором и документами конкретной поставки.

Базовый комплект

Документ	Какие вопросы закрывает	Кто проверяет
Договор и спецификация	Предмет, продукты, объём, срок, оплата, права использования, порядок изменений	Юристы и закупки
Описание поставки	Место установки платформы, отдельно маршруты моделей, компоненты, подключения и площадки	ИТ и ИБ
Поручение на обработку	Роли сторон, данные, операции, цели и условия обработки, если применимо	Юристы и ответственный за ПДн
Условия конфиденциальности	Допустимое использование информации, доступ исполнителей, передача материалов	Юристы и ИБ
Регламент поддержки	Каналы, время работы, категории обращений, эскалация и доступ	ИТ и владелец сервиса
Политика хранения	Категории, сроки, копии, удаление и подтверждение	ИБ, ИТ и юристы
Программа и акт приёмки	Проверяемые сценарии, результаты, ограничения, критерии устранения замечаний	Заказчик и исполнитель
План завершения	Выгрузка, переход, отзыв доступа, удаление и оставшиеся обязательства	ИТ, закупки и юристы

Названия файлов могут отличаться. Существенно, чтобы каждому вопросу соответствовал согласованный пункт и ответственный. Не создавайте несколько документов с противоречивыми сроками или разными названиями одного компонента.

Предмет и границы поставки

Укажите, какие части используются: Platform, Guardrail, внешний API и работы по внедрению. Отделите право использовать программное обеспечение от работ по установке, настройке, интеграции и сопровождению. Для каждого результата определите способ передачи и критерий принятия.

Для облака внесите адреса доступа, границы эксплуатационной ответственности, разрешённые площадки и порядок изменения провайдеров. Для on-prem — состав устанавливаемых у заказчика компонентов, инфраструктурные предпосылки, версии, модель обновления и доступ сопровождения. On-prem не означает автоматически запрет внешних моделей: модельные маршруты описываются отдельным приложением для обоих вариантов размещения платформы.

В приложении перечислите локальные модели заказчика, российские API, self-hosted модели Nevel на серверах РФ и глобальных провайдеров, разрешённых конкретной поставке. Например, self-hosted DeepSeek у Nevel и внешний API DeepSeek нельзя объединять одной строкой без указания получателей и площадок.

Поле паспорта маршрута	Что зафиксировать
Модель и размещение	Конкретный экземпляр, оператор инфраструктуры, площадки обработки и хранения
Цепочка получателей	Nevel, агрегатор при его использовании, конечный провайдер и вспомогательные обработчики
Данные и операции	Разрешённые категории, маскирование, OCR, embeddings, журналы и сроки
Границы	Выход из инфраструктуры заказчика и отдельно география обработки
Основной и резервный путь	Заранее разрешённые маршруты, условия переключения, запреты и запись в аудите
Проверка	Версия настройки, результат разрешённого вызова и отказа запрещённому получателю

Изолированный профиль указывается явно, если все рабочие обработчики должны находиться у заказчика. Для него отдельно фиксируются отсутствие внешнего модельного fallback и порядок обновлений, поддержки и передачи диагностики. Матрица вариантов раскрывает различия.

Проверьте права на используемые модели и внешние сервисы отдельно. Включение API в техническую схему не означает, что условия всех внешних правообладателей автоматически становятся одинаковыми с условиями Nevel.

Поручение на обработку персональных данных

Если применима обработка по поручению, его содержание сверяется с частью 3 статьи 6: закон определяет перечень обязательных условий поручения, включая данные, операции, цели, конфиденциальность, безопасность, подтверждающие сведения и уведомление оператора об инцидентах. Официальный текст 152-ФЗ.

Для подготовки поручения используйте приложение с конкретными полями:

- категории субъектов и данных с примерами допустимых документов;
- операции каждого участника и разрешённые цели;
- площадки, обработчики, модели и география;
- доступ работников исполнителя и порядок привлечения других лиц;
- подтверждение мер, взаимодействие при проверках и обращениях;
- сроки хранения, возврат, прекращение обработки и удаление;
- контакты и процедура передачи информации об инциденте.

Этот перечень помогает собрать исходные сведения для юриста и не является готовой универсальной формой поручения. Для оценки процесса используйте 152-ФЗ: согласование обработки.

SLA и поддержка

Для измеримого обязательства нужны объект измерения, период, методика, источник данных, исключения и последствия нарушения. «Высокая доступность» без этих параметров не позволяет проверить исполнение.

Согласуйте отдельно доступность интерфейса, обработку API, работоспособность внешней модели и устранение ошибок. Внешний провайдер может быть недоступен при работающем Nevel; резервный маршрут при этом должен оставаться допустимым с точки зрения данных и географии.

Для поддержки различайте регистрацию обращения, первый ответ, начало работ, временное восстановление и окончательное устранение. Контакты для инцидента ИБ и коммерческих вопросов также разделяют, чтобы срочная информация не зависела от обычного процесса закупок.

Доказательства для ИБ и ИТ

К технической части приложите обзор безопасности, потоки данных, Guardrail, доступ и аудит, хранение. В конкретной поставке дополните их схемой с площадками и протоколом испытаний.

Если требуется сертификация, аттестация или включение в реестр, проверьте реально предоставленный документ: объект, номер, срок, версию, ограничения и отношение к закупаемой конфигурации. Общая ссылка на законодательство не подтверждает наличие документа. Документы партнёра или площадки не следует автоматически приписывать самому продукту.

Приёмка и изменение условий

Приёмку свяжите с конкретными действиями: вход пользователя, работа разрешённой модели, отказ запрещённого маршрута, отзыв ключа, журналирование, восстановление и удаление тестовых данных. Для on-prem проверьте разрешённые внешние зависимости и блокировку остальных; для изолированного профиля — отсутствие внешних рабочих вызовов. В обоих случаях проверьте сбой основного маршрута и отсутствие скрытого переключения. Укажите версию конфигурации и перечень замечаний с решениями по каждому.

В порядке изменений определите, кто согласует новую модель, подключение, площадку или состав данных. Изменение коммерческого тарифа и изменение разрешённого маршрута обработки могут требовать разных участников согласования.

Завершение договора

До прекращения доступа согласуйте формат выгрузки, её полноту, период передачи и проверку читаемости. Затем зафиксируйте отзыв ключей и доступов, остановку подключений, очистку рабочих экземпляров, порядок обращения с резервными копиями и необходимые подтверждения.

Архив, переданный заказчику, должен иметь понятный состав и ответственного за получение. Если часть сведений сохраняется по установленному основанию, итоговый документ перечисляет категорию, ограничение использования и срок. Формулировка «всё удалено» применяется только к реально проверенной области.

Нормативная основа

Нормативная основа — 152-ФЗ. Состояние нормативной справки на 8 сентября 2026 года и версии актов указаны в разделе о 152-ФЗ. Окончательные формы, состав пакета и обязательства согласуют стороны с учётом применимых требований.