Nevel

ТАРИФЫ И ОПЛАТА

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.

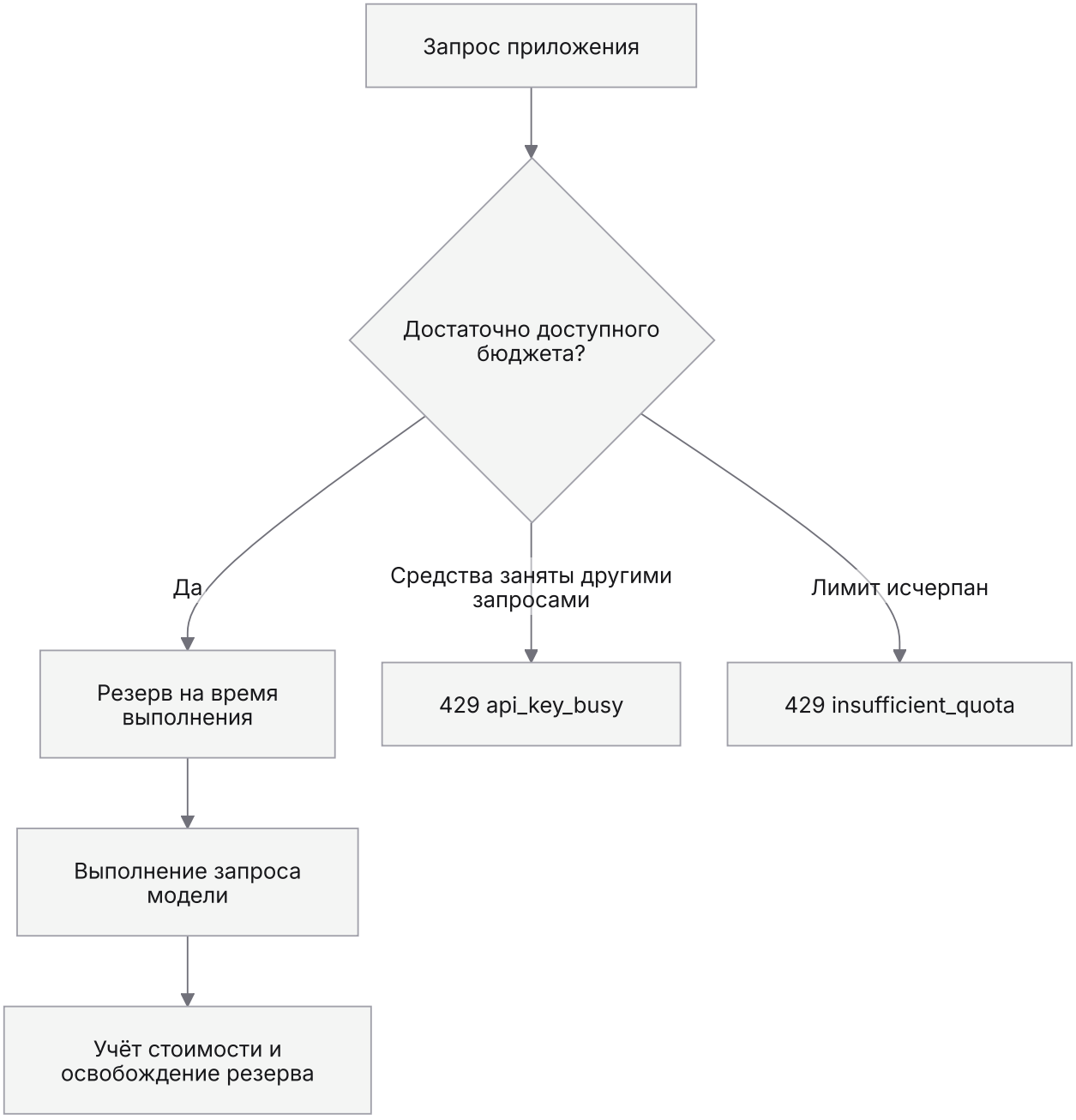
Отсутствие лимита ключа не отменяет ограничения моделей, прав доступа, квот и инфраструктуры. Для промышленной интеграции отдельный лимит полезен даже тогда, когда общий бюджет компании достаточен.

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.