Nevel

ВАРИАНТЫ УСТАНОВКИ

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

Маршрут	Где работает модель и кто её обслуживает	Что пересекает границу платформы	Применение
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
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.

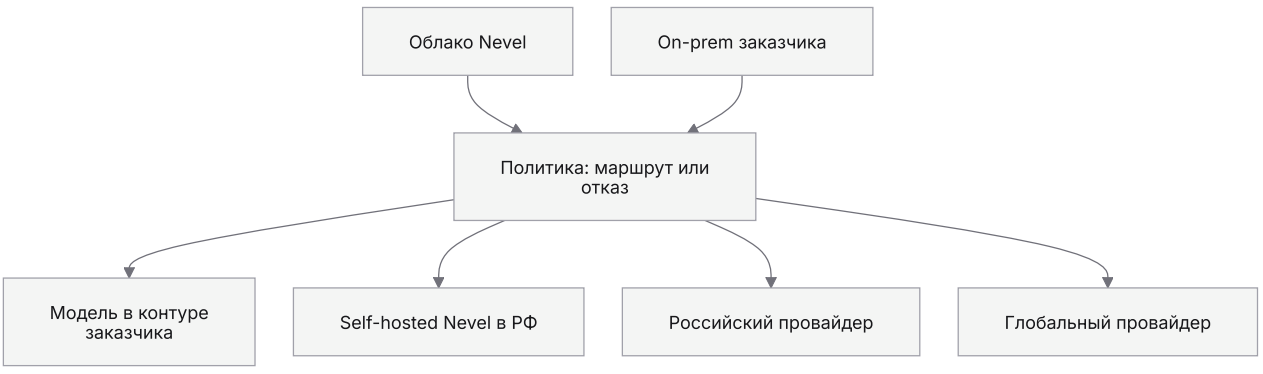
Для облака обращение к модели в контуре заказчика требует отдельно подготовленного разрешённого соединения; её доступность не возникает автоматически. Для любой комбинации необходимы подключённый endpoint, совместимость нужных методов и разрешение политики. Таблица не означает, что любой экземпляр модели 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.