Nevel

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

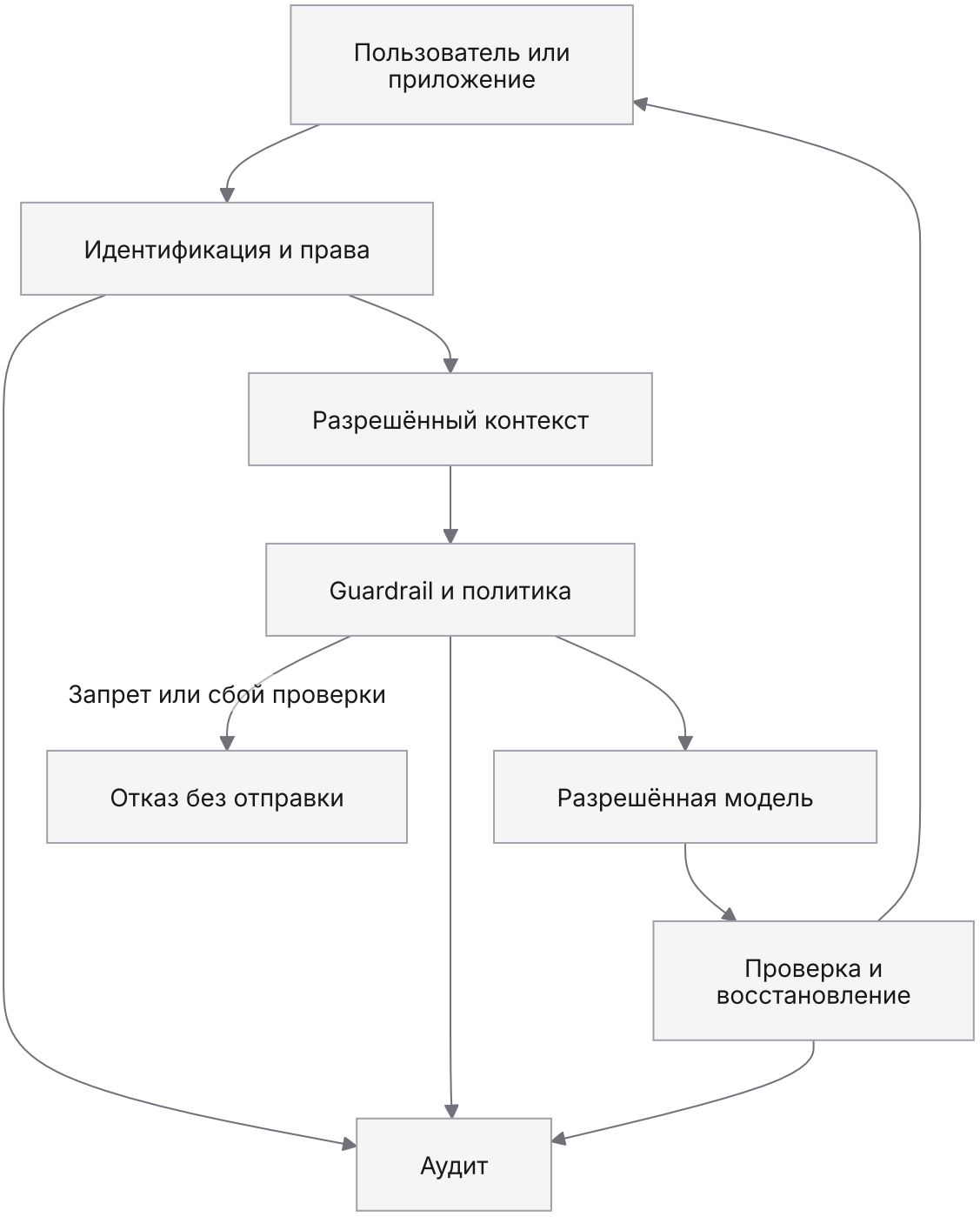


Схема показывает логические проверки, а не обещание одинаковой структуры серверов во всех поставках. Для архитектурной экспертизы существенны точки доступа к открытому тексту, получатели данных и возможность обхода обязательного контроля. Эти границы раскрыты в архитектуре и потоках данных.

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.