



Ай-ТИ Фреш

ТЕХНИЧЕСКИЙ РАЗБОР

AI-агент для экспресс-проверки ТИПОВЫХ ДОГОВОРОВ

Как мы разворачиваем ИИ-ревью аренды, поставки и услуг
до похода к юристу

Июль 2026

itfresh.ru · ИТ-аутсорсинг для юридических лиц

Суть проблемы

Компании до 50 РМ без штатного юриста подписывают договоры аренды, поставки и услуг «на глаз»: вычитывает бухгалтер без времени на 15-20 страниц. В текст проникают пени без потолка, расширенная ответственность, чужая подсудность, автопродлонгация. Разовая вычитка у юриста занимает 1-3 дня и стоит 3 000-15 000 ₽, а решение по сделке нужно принять сегодня.

Почему это важно бизнесу

- Договор подписывается под давлением сроков поставки/аренды без проверки — риски остаются незамеченными до первого конфликта.
- Разовая внешняя юр.проверка стоит от 3 000 ₽ и занимает 1-3 рабочих дня — блокирует сделку, если контрагент требует подписать сегодня.
- Штатный юрист для компании до 50 РМ экономически не оправдан — загрузка 2-5 договоров в месяц не окупает ставку 80-120 тыс ₽.
- Пропущенный пункт о неустойке без потолка или одностороннем расторжении оборачивается убытками,кратно превышающими стоимость проверки.
- Юрист без ИИ-предпроверки тратит время на пересказ типовых пунктов вместо разбора реально спорных условий — переплата за рутину.



Ключевые параметры реализации

100 стр.

Максимальный объём договора за один запрос — лимит нативной PDF-обработки Claude (до 32 МБ).

[Claude PDF support](#)

1М ток.

Контекстное окно Claude Sonnet 5 — весь договор и плейбук риск-паттернов в одном запросе.

[Claude models overview](#)

TTL 1 ч

Время жизни кеша плейбука риск-паттернов — держим весь рабочий день без переплаты за повтор.

[по докам Claude prompt caching](#)

10%

Цена чтения из кеша от базовой стоимости input-токенов при повторных договорах одного типа.

[по докам prompt caching](#)

5 мин

Наш SLA на полный постатейный разбор договора до 20 страниц с выдачей отчёта клиенту.

[наш стандарт](#)

-50%

Скидка на пакетную обработку архива договоров через Message Batches API при ретроаудите.

[Claude Batches API](#)



Приём документа и первичный триаж

Что настраиваем

Входной конвейер: email/Telegram-бот/веб-форма клиента, до 100 стр. на договор

Как мы это делаем

- 1 Клиент присылает PDF/DOCX договора в Telegram-бот или на почту; DOCX конвертируем в PDF, проверяем лимит (32 МБ, 100 страниц на запрос).
- 2 Файл грузим через Files API Claude — получаем file_id один раз и переиспользуем во всех последующих запросах без повторного base64.
- 3 Классификатор на Claude Haiku 4.5 определяет тип договора (аренда/поставка/услуги) и маршрутизирует в профильный subagent Claude Agent SDK.
- 4 Тriage занимает 20-40 с на документ до 20 страниц — клиент получает подтверждение приёма и ETA прямо в чат.

РЕЗУЛЬТАТ

Клиент видит, что документ принят и обрабатывается, в течение минуты после отправки — вместо ожидания ответа юриста «когда будет время».

КЛЮЧЕВОЙ НЮАНС

Держим лёгкий классификатор (Haiku) перед тяжёлым анализом (Sonnet) — это экономит основную часть бюджета на нерелевантных обращениях.



Постатейный разбор рисков

Что настраиваем

Профильный subagent на договор: аренда / поставка / услуги, плейбук риск-паттернов в system prompt

Как мы это делаем

- 1 Каждому типу договора — свой subagent Claude Agent SDK со своим плейбуком (150-200 типовых риск-формулировок) и ограниченным набором инструментов.
- 2 Плейбук кешируем через prompt caching с cache_control TTL=1 час — повторное чтение кеша стоит 10% от базовой цены input-токенов.
- 3 Claude Sonnet 5 разбирает договор пункт за пунктом: оплата, ответственность, форс-мажор, расторжение, пени, подсудность, автопродлонгация.
- 4 Для многоусловных пунктов (комбинированная неустойка, каскадная ответственность) поднимаем глубину adaptive thinking через effort до high/xhigh — Sonnet 5 управляет...
- 5 Ответ форсируем в Structured Outputs (strict JSON Schema): пункт, категория риска, уровень (красный/жёлтый/зелёный), рекомендация правки.

РЕЗУЛЬТАТ

Полный постатейный разбор договора до 20 страниц укладывается в 2-5 минут вместо 1-3 рабочих дней у внешнего юриста — клиент получает светофор рисков до подписания.

КЛЮЧЕВОЙ НЮАНС

Strict JSON Schema и Citations API нужны вместе: без привязки к странице/пункту оригинала риск-метка выглядит как «мнение модели», а не проверяемый факт.



Привязка к первоисточнику и выдача отчёта

Что настраиваем

Citations API + шаблон PDF-отчёта, доставка клиенту в Telegram/почту

Как мы это делаем

- 1 Каждую риск-метку подтверждаем через Citations API — Claude указывает диапазон страниц PDF, из которого взята цитата пункта договора.
- 2 Собираем итоговый PDF-отчёт: таблица «пункт / риск / что не так / что поправить» плюс сводный светофор по 6 категориям рисков.
- 3 Через MCP-коннекторы (email/telegram, уже развёрнутые в контуре клиента) отчёт уходит в течение минуты после готовности анализа.
- 4 Пачку старых договоров на ретроаудит переключаем на Message Batches API (до 100 000 запросов в пачке, до 24 ч, минус 50% к стоимости).

РЕЗУЛЬТАТ

Клиент получает документ с точными ссылками на страницы и пункты — может сам решить, идти ли с конкретным риском к внешнему юристу, вместо оплаты полной вычитки.

КЛЮЧЕВОЙ НЮАНС

Batch API оправдан только для несрочного ретроаудита пачками — для единичного договора «здесь и сейчас» используем синхронный Messages API, иначе теряем SLA в минутах.



Подводные камни

✗ Без Structured Outputs

Свободный текстовый ответ модели плывёт по формату между вызовами — фиксируем strict JSON Schema, иначе отчёт нельзя стабильно смэппить в PDF-шаблон.

✗ Риск-метка без Citations

Без Citations API модель может сослаться на «типичный риск» без привязки к странице — клиент теряет доверие; постраничные цитаты включаем обязательно.

✗ Один промпт на все типы договоров

Аренда, поставка и услуги имеют разные риск-паттерны (аренда — индексация, поставка — приёмка и сроки); держим отдельный subagent и плейбук на тип.

✗ Sonnet на любую заявку без триажа

Прогон каждого письма (включая нерелевантные) через тяжёлую модель раздувает счёт; ставим лёгкий Naiku-классификатор перед основным анализом.

✗ Кеш короче рабочего дня

Дефолтный 5-минутный TTL кеша плейбука сбрасывается между обращениями за день; переключаем на 1-часовой TTL, чтобы не платить write-цену повторно.

✗ Максимальный effort на каждом пункте

Высокий effort adaptive thinking на каждый пункт раздувает время и стоимость ответа; поднимаем глубину размышления выборочно — только для пунктов с к...

✗ PDF без проверки объёма

Скан на 150+ страниц или файл более 32 МБ падает на лимите нативной PDF-обработки; заранее делим документ на части и агрегируем результат по разделам.

✗ Вывод подан как юр.заключение

Отчёт агента — предварительный разбор для скрининга рисков, не замена юриста по существу спора; эту границу фиксируем прямо в шаблоне отчёта.

Как правильно

МИНИМУМ

- Синхронный приём договора (email/бот) + классификация типа (Haiku).
- Постатейный разбор Sonnet 5 со Structured Outputs без повышенного effort.
- Citations на каждую риск-метку, отчёт в PDF по фиксированному шаблону.

НОРМАЛЬНО

- + prompt caching плейбука с TTL 1 ч на весь рабочий день.
- + повышенный effort (adaptive thinking) точно на сложные каскадные пункты (неустой...
- + отдельный subagent и плейбук на каждый тип договора (аренда/поставка/услуги).
- + доставка отчёта клиенту через MCP email/Telegram-коннектор без ручной пересылки.

ХОРОШО

- + Message Batches API для ретроаудита архива договоров пачками (-50%, до 24 ч).
- + версионирование плейбука риск-паттернов, регулярная сверка с практикой юриста.
- + интеграция с CRM/Планфикс: риск-отчёт создаёт задачу на согласование правок.
- + SLA-мониторинг времени разбора и точности риск-меток по обратной связи юриста.



Чек-лист самопроверки

- | | |
|--|---|
| ☐ Лимит объёма договора (100 стр./32 МБ) проверяется перед отправкой в Claude PDF support? | ☐ Повышенный effort (adaptive thinking) включается точно, только для пунктов с каскадными условиями? |
| ☐ Договор проходит триаж-классификацию типа (аренда/поставка/услуги) перед основным анализом? | ☐ Есть отдельный subagent/плейбук под каждый тип договора, а не один универсальный промпт? |
| ☐ Ответ модели зафиксирован Structured Outputs (strict JSON Schema), а не свободным текстом? | ☐ Отчёт содержит оговорку «предварительный технический разбор, не юридическое заключение»? |
| ☐ Каждая риск-метка подтверждена Citations API со ссылкой на страницу/пункт оригинала? | ☐ Для пакетного ретроаудита архива подключён Message Batches API со скидкой 50%? |
| ☐ Плейбук риск-паттернов кеширован с TTL, покрывающим рабочий день клиента? | ☐ Доставка отчёта клиенту автоматизирована через MCP-коннектор (email/Telegram/CRM)? |

Если хотя бы на два вопроса ответ «нет» или «не знаю» — тема требует внимания.



Как поможет ITFresh

ITFresh — ИТ-аутсорсинг для юридических лиц до 50 рабочих мест в Москве и области. 15+ лет практики, собственная инфраструктура в дата-центре МТС (8 серверов Dell Xeon Platinum).

- Разворачиваем приёмный контур (Telegram-бот/почта/веб-форма) под приём договоров клиента под ключ.
- Пишем и настраиваем плейбуки риск-паттернов по типам договоров (аренда/поставка/услуги) на Claude Agent SDK.
- Настраиваем Structured Outputs, Citations API и шаблон PDF-отчёта под фирменный стиль клиента.
- Подключаем prompt caching, adaptive thinking (effort) и Message Batches API для оптимальной стоимости под объём кл...
- Сопровождаем и обновляем плейбуки по обратной связи от штатного или внешнего юриста клиента.

15+

лет в ИТ-поддержке

50

рабочих мест — наш профиль

МТС

дата-центр, Москва



КОНТАКТЫ

Обсудить вашу задачу

Сайт **itfresh.ru**

Телефон **+7 903 729-62-41**

Telegram **@ITfresh_Boss**

Бесплатно посмотрим вашу инфраструктуру по этому чек-листу и скажем, где тонко — без обязательств.



itfresh.ru



Техническая база

- 01** PDF support (native document processing) (platform.claude.com — 2026)
- 02** Structured outputs (JSON Schema) (platform.claude.com — 2026)
- 03** Prompt caching (platform.claude.com — 2026)
- 04** Citations API (platform.claude.com — 2026)
- 05** Agent SDK — Subagents (code.claude.com — 2026)
- 06** Message Batches API (platform.claude.com — 2026)
- 07** Models overview & pricing (platform.claude.com — 2026)
- 08** MCP connector (platform.claude.com — 2026)

Основано на официальной документации продуктов и нашей практике внедрения.

