

КЛЮЧЕВОЙ УПРАВЛЕНЧЕСКИЙ ВЫБОР В ЭПОХУ AI

# Что оставить людям, что коду, а что агенту

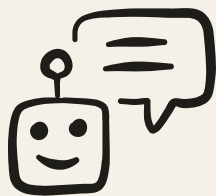
Выигрывает тот, кто правильно проводит линию.



ЛЮДИ



КОД  
(WORKFLOWS)



АГЕНТ

где  
провести  
границу?



# Refat Ametov

CEO & co-founder Devstark · AI Solutions Architect · Researcher & Author

- 01 Ко-фаундер **Devstark** / **SpreadSimple**, строим AI-first продукты
- 02 13+ лет в разработке · 4+ года системно вношу GenAI в прод
- 03 Через меня проходит много проектов: свои + адвайзинг
- 04 Клиенты: Porsche, Forbes-featured стартапы, enterprise EU/US

Поменьше  
теории,  
побольше  
практики.



TELEGRAM  
[@nobilix](#)  
REFAT TALKS: TECH & AI



LINKEDIN  
[in/refat-ametov](#)

ЗАЧЕМ ВООБЩЕ

# «Делать ли» — это уже не вопрос

ВОПРОС ВЕРЫ · ОПЦИОНАЛЬНО

## Верить в (gen)AI — или нет

AGI, «заменит всех», сроки, «пузырь».  
Это про будущее — спорить можно  
бесконечно, и доклад не об этом.

считаем, а не гадаем

ВОПРОС ФАКТА · ИЗМЕРИМО

## Автоматизация и производительность

На достаточно большом количестве  
проектов становится очевидно, что  
автоматизация работает и при правильном  
применении LLM-технологии это сильно  
ускоряют.

ТЕНДЕНЦИЯ ГОДА

# Год harness-инжиниринга

Агенты повзрослели, и рычаг сместился с модели на **harness**, обвязку вокруг модели (тулы, контекст, навыки, память). На ней теперь «программируют», и строят автоматизацию.

## Software 3.0 · Карпаты

Программируешь поведение текстом, поверх harness, а не пишешь код.

## Harness — first-class

skills · контекст · память стали прод-кирпичом (MS Build 2026).

## Модель — коммодити

6 топ-моделей в пределах 0.8 балла; смена harness дает **+22**. Решает обвязка, не модель.

## Managed Agents

Вендоры продают harness как сервис, не только модель.

Созрело первым в кодинге — приемы (skills, файлы, review-петли) едут и в автоматизацию бизнеса. Работает, но это не значит «агент везде»: понятнее всего автоматизировать промптом, поэтому его и тащат повсюду.

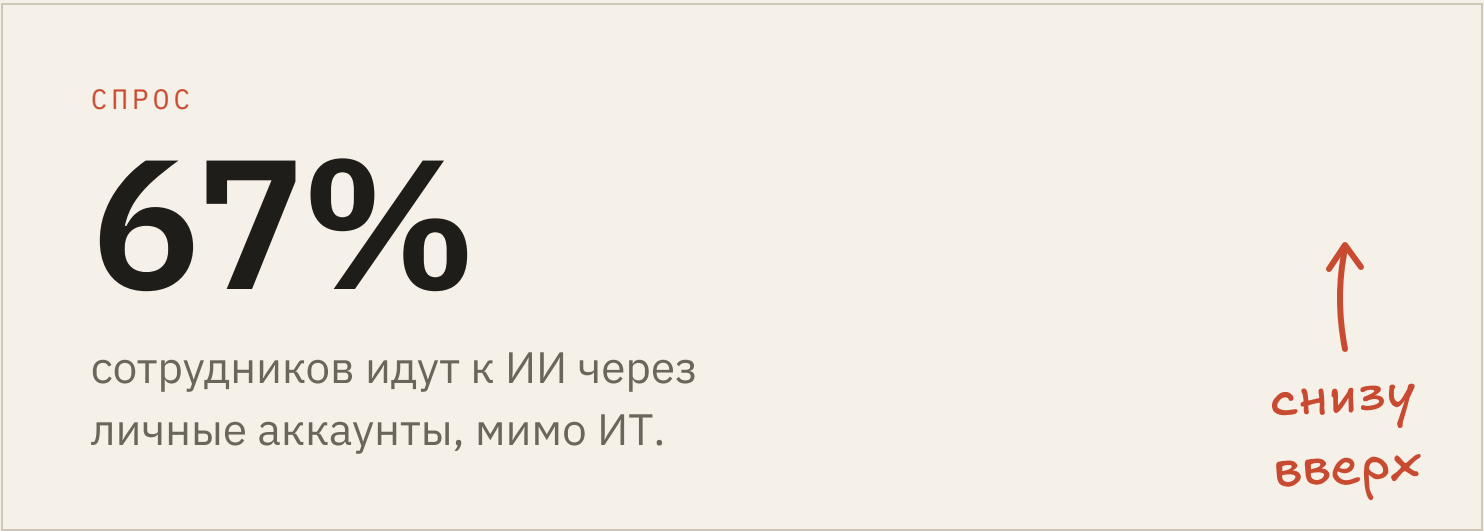
не «везде», а где уместно

# Shadow AI

ИИ уже внедряют снизу, на всех уровнях, кто как умеет.

- Сигнал:** реальный спрос, люди закрывают свои задачи сами.
- Угроза:** вне контроля, данные наружу, аудита нет, не откатить.

Не проведешь линию ты — проведут стихийно.





ЧТО Я ВИЖУ НА ДЕСЯТКАХ ВНЕДРЕНИЙ

# Ошибаются не в модели, а в выборе инструмента

- 01 **Не видят, что чем делать.** Задачу для детерминированного кода отдают агенту; где нужен человек — суют автоматизацию.
- 02 **Тащат агента всюду.** Модно, но дорого, медленно и непрозрачно.
- 03 **Отбирают у человека лишнее.** Автоматизируют суждение, вкус, живой контакт — то, что должно остаться людям.

Значит, выбор не «внедрять ли AI».

**А что оставить людям, что коду, а что агенту.**

об этом весь доклад

ДОГОВОРИМСЯ О ТЕРМИНАХ

# Код или агент: о чем речь



Граф, запрограммированный **заранее**. Форма задана до запуска, даже если внутри вызывается LLM.

шаг за шагом · жесткая схема  
цена предсказуема

- n8n
- Airflow
- Temporal
- Mastra Workflows
- LangGraph



**Harness** + инструкции текстом + инструменты. Путь выполнения **выбирает сам** — агентный цикл.

цикл решений · автономия  
меняется в разговоре

- Claude Code
- Codex
- OpenCode
- Hermes
- Agent SDK
- Managed Agents

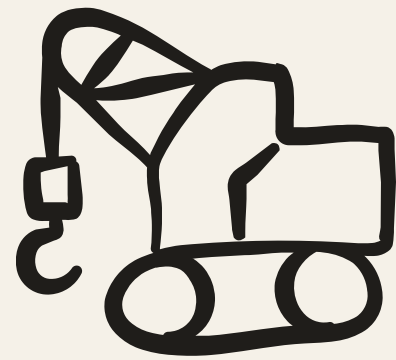
## Не смена поколений.

Два режима — **каждый выигрывает и ломается в своем месте**. Где именно — дальше.

Software 3.0  
harness — «на чем программируют»

ДВЕ РАЗНЫЕ ИГРЫ

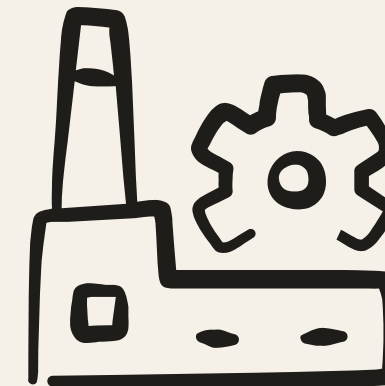
# Augmentation vs. Automation



## Augmentation

Человек в контуре ловит  
ошибки — автономия  
дешева.

VS.



## Automation

Без человека — нужна  
стабильность на скейле.

разные  
критерии  
решения



# Кейс AI-first агентства

- Фаундер, не-инженер
- **Активный пользователь AI**, каждый день в работе.
- Умеет экспериментировать, но очень прагматичен, оставляет только рабочее

## НАГЛЯДНЫЙ ПРИМЕР ТРАНСФОРМАЦИИ

- 01 Активный, но **прагматичный** фаундер, каждый день работает через AI.
- 02 Рано увидел: тот же **harness** из повседневной работы можно повернуть в **автоматизацию**.
- 03 И так же быстро уперся в проблемы, и нашел **баланс**.

разберем кейс

ОБРАТНАЯ СТОРОНА

та же скорость — новый риск

# «I feel like a developer all week. It's just too much.»

— JURRE · МЕСЯЦ СПУСТЯ

Harness и набор навыков — отличный **усилитель эксперта**. Но как **система автоматизации** он ломается, и делает эксперта **незаменимее**, а масштабирование сложнее.

к этому кейсу мы еще вернемся ↩



Все работает, но запутано — «hairy»

НЕ ПОД НОЖ, А ВВЕРХ

# Эксперт незаменим — и растёт

Он держит планку — определяет, что считать **хорошим**.

Это суждение не делегируешь ни коду, ни агенту.

IRONIES OF AUTOMATION · 1983

Чем больше автоматизации, тем **критичнее**, не меньше, роль человека.



**Оракул**

дает экспертизу



**Оценщик**

мерит качество



сюда целимся



**Архитектор**

проектирует процесс

«Когда исполнение перестает быть дефицитом, дефицитом становится суждение.» — BRUNO MEDEIROS

РЕГЛАМЕНТЫ → ВЕРСИОНИРУЕМЫЙ АКТИВ

# Сырье у вас уже есть

01 Армия — уставы и боевые инструкции

02 McDonald's — 600+ страниц регламентов

03 Toyota TWI — «шаг → ключевой момент → почему»



SKILL.md

80 лет  
дисциплины



SKILL.MD

# Обработка заявки

### Шаг 3 — сверка реквизитов

▀ риск: чужой счет → стоп

### Проверка — прогнать на новичке

Экспертиза покидает одну голову →  
**версионизируемый актив**, а не «все в голове у  
Сергея».

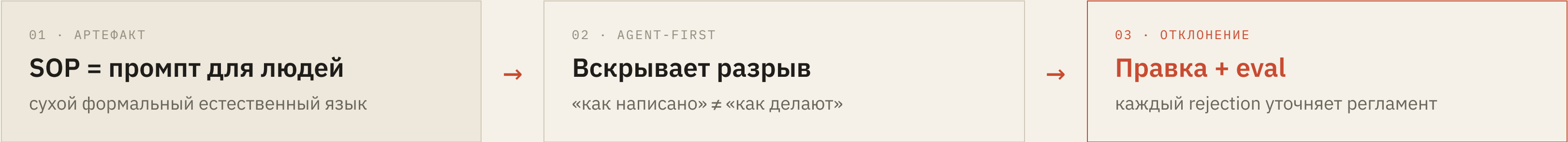
«Best practices encoded in a skill propagate faster than best practices in a blog post.» — TRISTAN HANDY · DBT LABS

КАК ТАЦИТ-ЗНАНИЕ СТАНОВИТСЯ ЯВНЫМ

# SOP — это и был промпт

Тот же регламент, которым «программировали» людей, теперь программирует агента.

← *Living SOP — созревает на отклонениях*



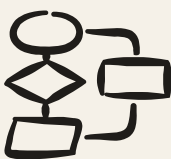
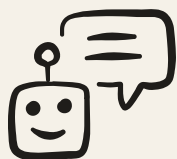
Неизвестный кейс — агент не выдумывает: помечает «**exception, решить не могу**» и отдает человеку.



ТРИ РЕЖИМА ИСПОЛНЕНИЯ

# Люди · Код · Агент

Augmentation мы уже отделили — здесь делим саму работу: **кому отдать исполнение.**

|                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <div></div> <div><b>Люди</b></div> <div>Суждение, вкус, необратимое, живой контакт.</div> <div><div>ЛОМАЕТСЯ</div><div>не масштабируется</div></div> | <div></div> <div><b>Код / процесс</b></div> <div>Событие → цепочка → результат. Граф рисуется заранее.</div> <div><div>ЛОМАЕТСЯ</div><div>негибко</div></div> | <div></div> <div><b>Агент</b></div> <div>Граф не нарисовать: коддинг, дебаг, computer use.</div> <div><div>ЛОМАЕТСЯ</div><div>дорого · «в 3 ночи апрувит не те счета»</div></div> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

ПРАВИЛО ЛИНИИ

Нарисуй блок-схему **до** запуска LLM. Получилось — это **workflow**. Не получилось (поток зависит от того, что найдет модель) — **агент**.

ПРАВИЛО МАРШРУТИЗАЦИИ

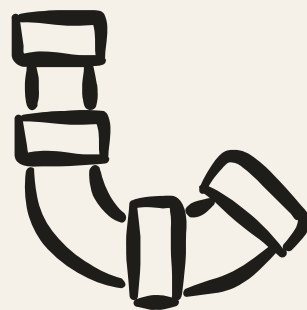
# Кому отдать задачу

|                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <div><div>КАК МЫСЛИТ ИНДУСТРИЯ</div></div> <div><h2>8 уровней автономии</h2><p><del>«выше уровень = лучше»</del></p></div> <div><p>Одномерная лестница: чем выше, тем «продвинутее».</p></div> | <div><div>МОЯ ОСЬ</div></div> <div><h2>По характеру задачи</h2><ul style="list-style-type: none"><li>→ Волатильное — агенту</li><li>→ Вызревшее — в код</li><li>→ Суждение — человеку</li></ul></div> <div><p>не «выше», а «кому»</p></div> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Автономия и score — **две независимые оси**, а не одна лестница. Практический вопрос — не «насколько автономно», а **«кому отдать»**.

ВОЗРАЖЕНИЕ

# «А почему не все на RPA?»



## Жесткий RPA

Приходит другая форма данных — труба трескается.



## Agent-as-author

Сам перекладывает трубу — под ревью человека.

кто  
переписывает?

МАЯТНИК — К ДЕТЕРМИНИЗМУ

**<15%** включают агентное в automation-сюэты, 2026

**40%+** агентных проектов отменят к 2027

RPA — это код, который переписываешь руками при каждом изменении.  
Ценность не в детерминизме, а в том, **кто его поддерживает.**



ПРАВИЛО ВЫБОРА

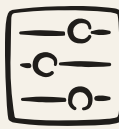
# Делегировать или сотрудничать

ДЕЛЕГИРУЙ КОДУ, ЕСЛИ

- ✓ Repeatable — повторяемая
- ✓ Objective — объективная
- ✓ Checkable — проверяемая
- ✓ Low-risk — низкий риск

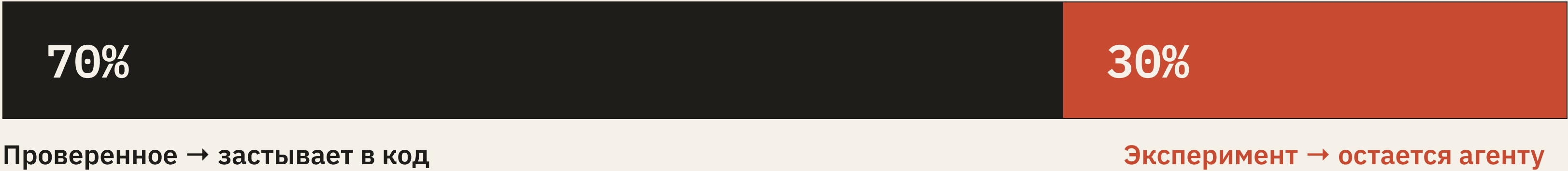
ОСТАВЬ ЧЕЛОВЕКУ, ЕСЛИ

- Ambiguous — неоднозначная
- Judgment-heavy — требует суждения
- Exploratory — исследовательская
- Iterative — итеративная



ЛИНИЯ ПОДВИЖНА

# Линию настраивают, а не фиксируют



Главный навык — не разовый выбор, а цикл: чувствовать, **сколько понимания достаточно**, и знать, когда не кристаллизировать.

КАК ДВИГАТЬ ЛИНИЮ

# Автономию зарабатывают

|                                                                                      |                                                                                    |                                                                             |
|--------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <div>ШАГ 1</div> <div>Полное ревью</div> <div>каждый результат смотрит человек</div> | <div>ШАГ 2</div> <div>Выборочная проверка</div> <div>человек проверяет часть</div> | <div>ШАГ 3</div> <div>Автономно</div> <div>работает само, под аудитом</div> |
| ✓ 6–7 аппрувов без правок → автономно                                                |                                                                                    | ✗ 1 реджект → снова человек                                                 |

65%+

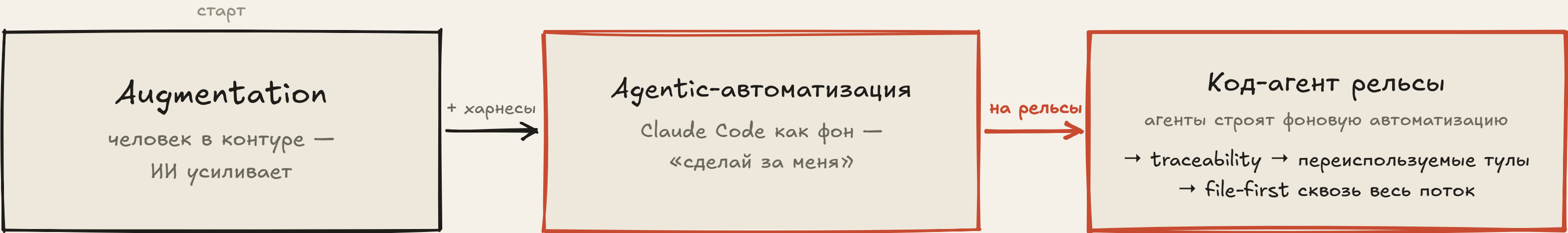
согласований агент обрабатывает **полностью** — поверх LLM  
детерминированные guardrails (лимиты, блок-листы) + слайдер автономии. Ramp

НЕ «АГЕНТ ИЛИ КОД», А ЗРЕЛОСТЬ ПРАКТИКИ

# Кейс прошел три режима

С этим правилом вернемся к кейсу. Сначала ИИ усиливал человека; потом агент начал строить сам; дальше нужен третий слой: **рельсы, чтобы скорость стала управляемой.**

возвращаемся к кейсу ↩



тупик:

нет прозрачности · тяжело поддерживать  
работает только у автора · рутина

и это работает

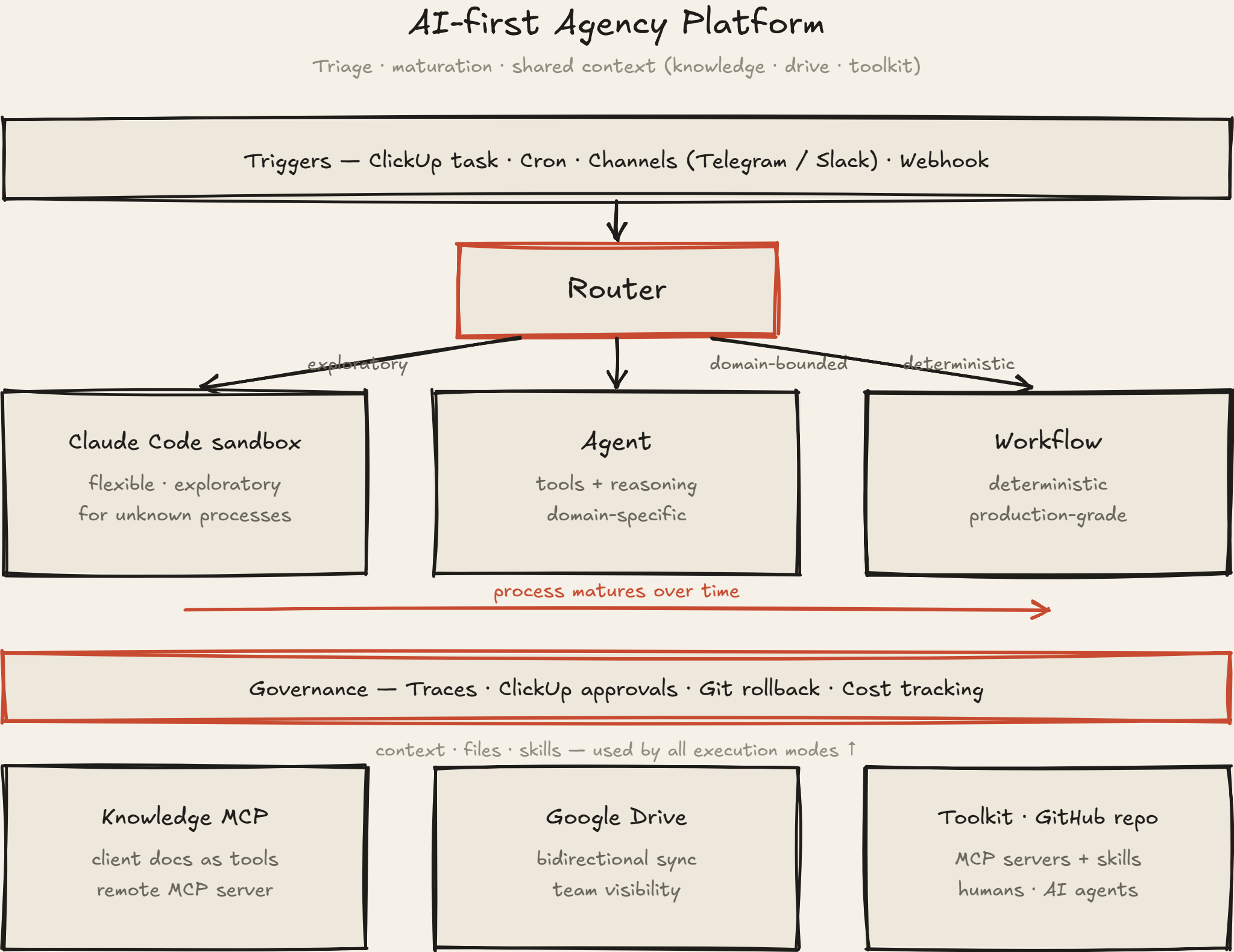
путь кейса · со временем

КАК ПЕРЕСТАТЬ БЫТЬ ТОЧКОЙ ОТКАЗА

# Не чат, а платформа

Единый вход, выбор режима под зрелость процесса, общие рельсы — поддерживаемый механизм, а не магия одного человека.

**1–2** сервиса в неделю — эксперт упаковал, дальше едет без разработчика





кейс-якорь · CONSTRUCTION DOCS

[ замер ]

тот же процесс

агент · CLAUDE CODE

\$11

за документ

vs.

workflow · MASTRA

<\$1

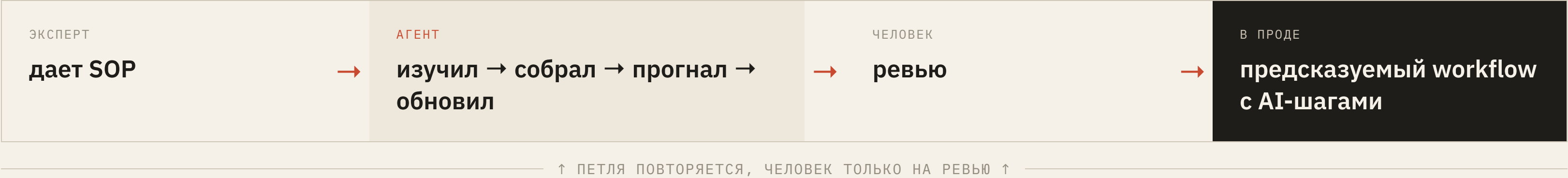
за документ

Одна и та же задача, два режима исполнения: цена агента растёт с объёмом, workflow — почти бесплатно.

Оба режима кратно дешевле ручной обработки — она была на порядок выше левой цифры. Но удешевление и скорость workflow пришли **после того, как агент протоптал дорожку** и стало ясно, как автоматизировать оптимально.

СИГНАТУРНЫЙ ПАТТЕРН

Агент — автор и редактор, не рантайм



«Agents good for prototyping, workflows better for production.»

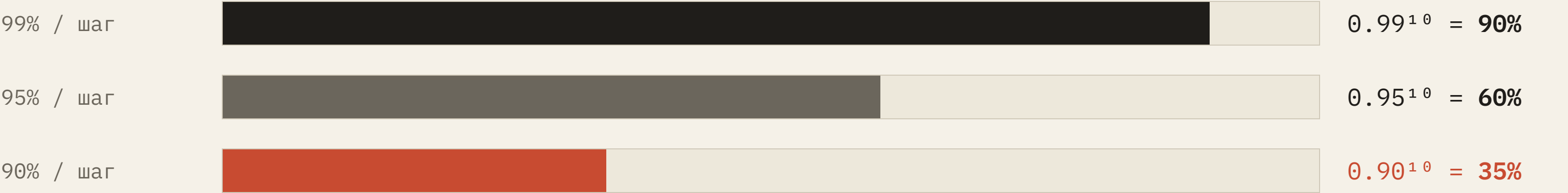
ЧТО ЭТО ДАЕТ

офлайн агент  
компилирует  
повторяющиеся действия  
в детерминированные  
инструменты — на  
масштабе **дешевле и  
надежнее.**

меньше вызовов LLM, выше успех,  
измерено на бенчмарках (arXiv)

ПОЧЕМУ ДЛИННЫЕ ЦЕПОЧКИ РВУТСЯ

# Математика компаундинга



10 шагов по 90% — это **35%** успеха. Поэтому повторяющееся выносим в **детерминированный workflow**, а агента бережем для незаменимого: там, где граф заранее не нарисовать (кодинг, дебаг, computer use).



ПРЕДЕЛ ДЕЛЕГИРОВАНИЯ

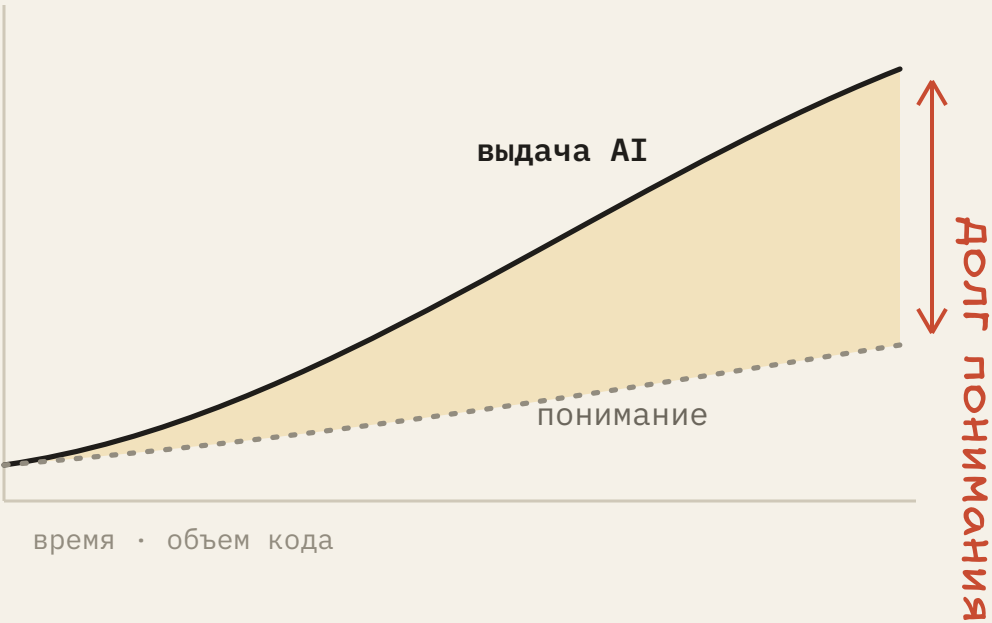
# Делегируй исполнение — не понимание

АИ снимает исполнение. Понимание — **не делегируется.**

Чем мощнее инструмент, тем дороже его отсутствие.

THEORY BUILDING · 1985

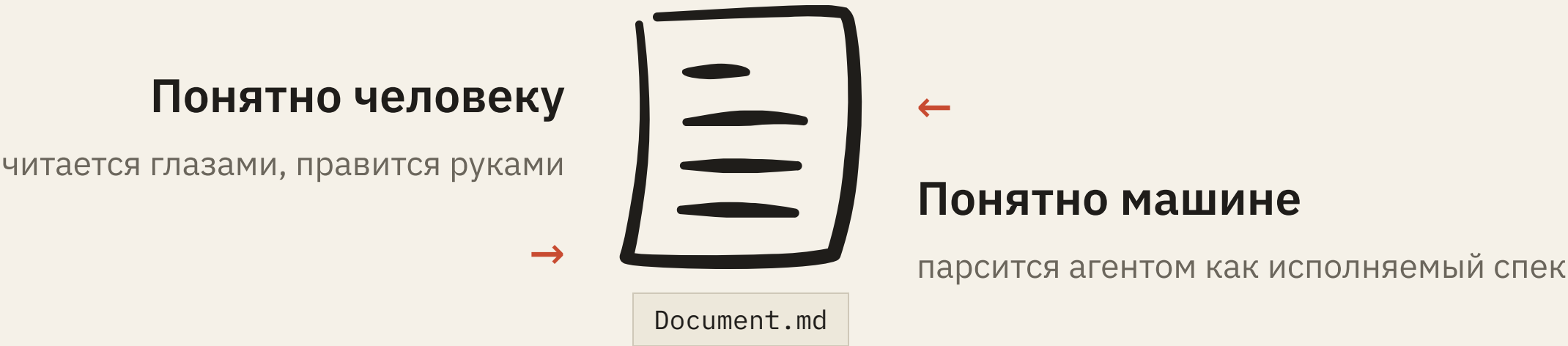
Понимание не живет в артефактах — оно в голове команды. — P. Naur



АИ пишет быстрее, чем растёт понимание. Разрыв копится как **долг** — отдаешь при отладке и владении.

АХ · АГЕНТНЫЙ ОПЫТ

# Ода файлам и переносимости знаний



АХ — свойство среды: насколько она **читаема для агента**. Ставка на файлы.

ПРАВИЛО

Берите форматы, которые **не нужно распаковывать** — удобные и человеку, и агенту: **Markdown, а не .docx; файлы и синхронизация, а не тяжелая БД.**

Думайте про agent experience — это экономит токены, latency и нервы.

HUMAN-IN-THE-LOOP

# Ревью — это поверхность дизайна



**Смотрю**

агент показал результат



**Запускаю**

одним нажатием в потоке



**Принято**

прямо в ClickUp / чате

Качество HITL-UX решает: устойчивое владение — или ревью-театр.

**16% → 54%**

содержательного ревью, когда поверхность ревью **спроектирована** (при <1% ложных). Ревью стало новым узким местом — проектируйте его, а не терпите. Anthropic

НОВЫЙ НАЛОГ

# Счет за токены — не главная цена



счет за токены

≠



в 3 ночи читаешь  
логи как детектив  
вот это — настоящая цена

«The token bill isn't the cost. The 3 a.m. trace you read like a detective — that's the cost.»

|                                                               |                                                                       |                                                                                     |
|---------------------------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| <div>10–100×</div> <div>токенов жгут агенты против чата</div> | <div>30×</div> <div>\$0.04→\$1.20 за тот же запрос, 3 года · EY</div> | <div>Рельс = kill-switch</div> <div>лимит бюджета · лимит вызовов · авто-стоп</div> |
|---------------------------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------|

КОНТР-ХАЙП ДЛЯ БОРДА

# Увольнения ≠ ROI

«Сокращения освобождают бюджет, но не создают возврат.»

GARTNER · ОПРОС 350 РУКОВОДИТЕЛЕЙ

И парадокс: «автономный бизнес создаёт людям больше работы, а не меньше».

→ ROI не из увольнений, а из редизайна процесса

ДЕЛАЛИ

~80%

компаний сокращали штат под AI

ПОЛУЧИЛИ ПО ROI

≈ 0

сокращали одинаково и лидеры, и аутсайдеры по возврату



ИНФРАСТРУКТУРА МАСШТАБА

# Шеринг примитивов через управляемую шину



Шеринг примитивов невозможен без governance-слоя, и это прямой ответ Shadow AI.

META · RULE OF TWO

не больше 2 из 3 в одной сессии: недоверенный ввод · чувствительные данные · действие вовне

GARTNER

спрашивай «какой это ТИП», не «одобрен ли». К 2027 40% отключат/понижат агентов из-за governance-дыр

СОВЕТ ДИРЕКТОРУ



# Назначь владельца

|                                                                                                                                                                                                                               |                                                                                              |                                                                                                |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| <div>01</div> <div>Профильный эксперт</div> <div>у каждой автоматизации есть конкретный владелец, не отдел</div>                                                                                                              | <div>02</div> <div>Дай ему владение</div> <div>право менять процесс и отвечать за него</div> | <div>03</div> <div>Никакого «поставил и забыл»</div> <div>это не устойчивый режим работы</div> |
| <div><div>«Почему человек не смог это остановить?»</div><div>Настоящее владение — это <b>защищенное право решать</b>: время · информация · право остановить · психбезопасность. Иначе «владелец» — лишь табличка.</div></div> |                                                                                              |                                                                                                |

КЕЙС · АГРЕГАТОР СОБЫТИЙ

[ замер на проде ]

# LLM работает, но копится долг качества

Американский стартап, агрегатор событий: ~6K событий в день. Сбор держится на универсальном LLM — **68%** источников, широкое покрытие быстро. Но за ширину платим качеством.

| НА ОДНОМ ПРОДЕ    | LLM-СБОР | ДЕТЕРМ. ПАРСЕР |        |
|-------------------|----------|----------------|--------|
| Публикуется       | 67%      | 81%            | +14 пп |
| Событий за запуск | 5        | 37             | ×7     |
| Дубли             | 22%      | 7%             | ×3     |
| Брак на ревью     | 13%      | 4%             | ×3     |

80%

ручной проверки приходит с LLM-источников: долг качества отдаем руками.

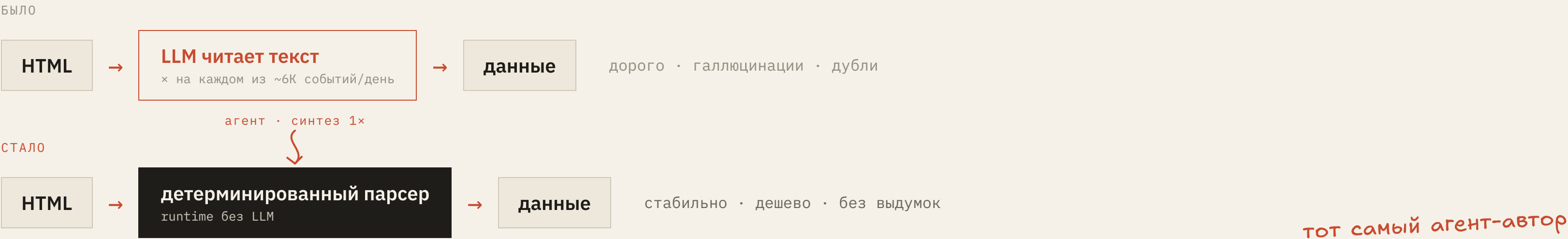
ширина ценой качества



КЕЙС · РЕШЕНИЕ

[ внедряем ]

# Агент пишет парсер, а не парсит



LLM уходит из рантайма в авторы. Дальше работает чистый код: дешевле, стабильнее, **+14 пп к публикации.**

Каждый отклоненный кейс становится правилом, и парсер созревает на отклонениях.

ЧТО СДЕЛАТЬ ЗАВТРА

# Возьмите один регламент и проведите по нему линию

|                                                                                                                                                         |                                |                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|----------------------------------|
| <div>ЧТО</div> <div>людям</div>                                                                                                                         | <div>ЧТО</div> <div>коду</div> | <div>ЧТО</div> <div>агенту</div> |
| <div>для SMB Не нанимай CAIO — назначь <b>owner + 1–2 champion</b> с защищенным временем 30–60 мин/нед. Начни с <b>3–5 кейсов</b> в ядре бизнеса.</div> |                                |                                  |

Перестаньте быть узким горлышком — станьте владельцем линии.

ВОПРОСЫ · ОТВЕТЫ · КОНТАКТЫ

# Спасибо за внимание

С удовольствием отвечу на вопросы.  
Контакты для связи на QR.



TELEGRAM  
[@nobilix](https://t.me/nobilix)



LINKEDIN  
[in/refat-ametov](https://www.linkedin.com/in/refat-ametov)

давайте на связи