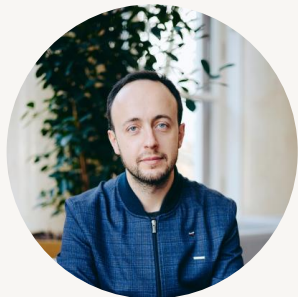


ИИ и смерть экспертизы:

практика тотальной автоматизации пентеста или
«безопасника заменить нельзя терпеть»

Рустам Гусейнов, ПК РАД КОП





Рустам Гусейнов

Председатель производственного кооператива «РАД КОП»

Опыт, роли

Аудитор, консультант и управленец в сфере ИБ, основной фокус: внедрение процессов и управление изменениями;

Работал в ФНС России, Digital Security, в группе компаний IPChain;

Действующий QSA аудитор, участник экспертного совета Банка России по внутреннему аудиту, преподаватель авторского курса НИУ ВШЭ;

С 2017 года активно интересуюсь темой ИИ и границами AI Native парадигмы =)

Напоминание о границах человеческих

«Великий Фультон! И великий Ватт! Ваше молодое, не по дням, а по часам растущее дитя в скором времени пожрет кнуты, престолы и короны, а дипломатами и помещиками только закусит, побалуется, как школьник леденцом. То, что начали во Франции энциклопедисты, то довершит на всей нашей планете ваше колоссальное гениальное дитя. Мое пророчество несомненно»

Т. Г. Шевченко. Дневник, 27 августа 1857 года





СПОЙЛЕР

Заменит ли нас ИИ?

ЗАДАЧИ

Да

Черновики, поиск, разбор логов, первичный триаж, типовые заключения — уже дешевле отдать машине и проверить.

РОЛИ

Частично

Роль меняется, когда меняется набор задач. Но кто-то по-прежнему ставит задачу, принимает результат и отвечает за последствия.

ОТВЕТСТВЕННОСТЬ

Нет

Ответственность не делегируется провайдеру модели. Её можно только потерять — вместе с суждением и навыком проверки.

Баланс входящего и исходящего

ВХОД < ВЫХОД

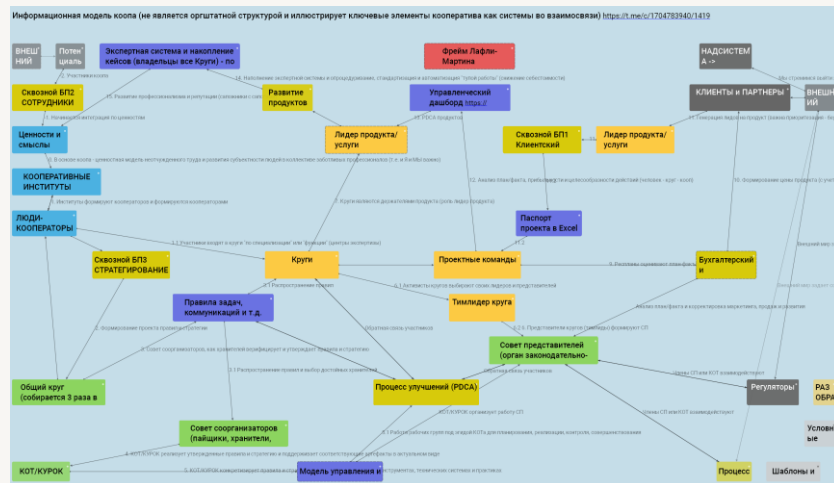
Горизонты управления

Операционный
день

Тактический
неделя

Стратегический
12 недель

Видение
годы



Отсутствие стратегии и видения

- 1 Организация и ЛПР «судорожно хватаются» за локальные возможности
- 2 Целая картинка отсутствует
- 3 Организационных ритмов и порядка — нет

**ЕСЛИ У ВАС НЕТ СОБСТВЕННОЙ
ПАМЯТИ — ВЫ «МЯСНОЙ ПРОКСИ»**

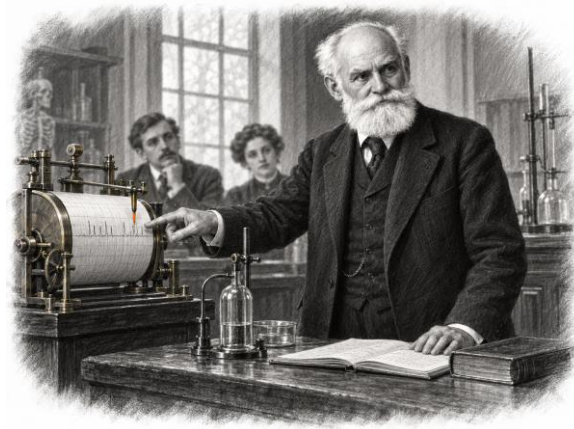
Перефразированная цитата одного из участников ИБ комьюнити

ИИШНИЦА как deus ex machina

Иван Павлов «О русском уме»

*«У нас должна быть одна
потребность, одна обязанность —
охранять единственно нам
оставшееся достоинство:
смотреть на самих себя и
окружающее без самообмана».*

И. П. Павлов. Лекция 7 (20) мая 1918 года





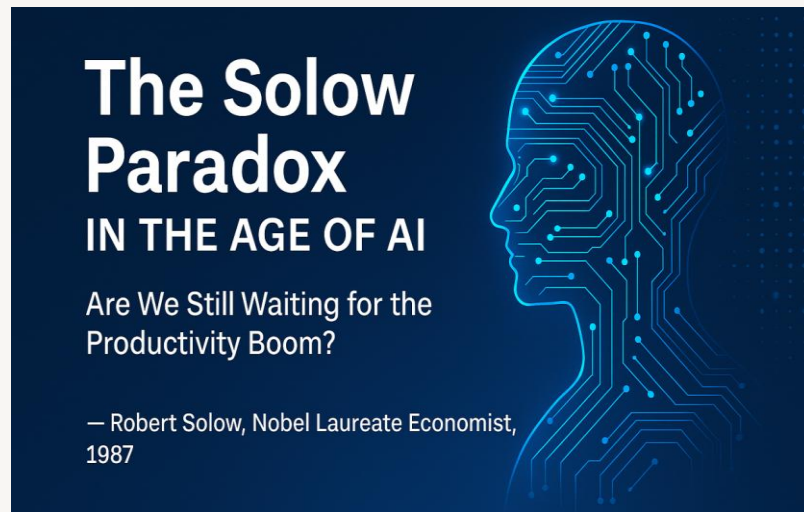
ОГРАНИЧЕНИЯ И ВОПРОСЫ



Ощущение **ускорения** есть, а **эффекта** нет

Почему так происходит?

- Внедрение $\neq$ эффект
- Реальные выгоды появляются при перестройке процессов
- Эффект от технологий раскрывается со временем
- Многие улучшения трудно измерить традиционными метриками

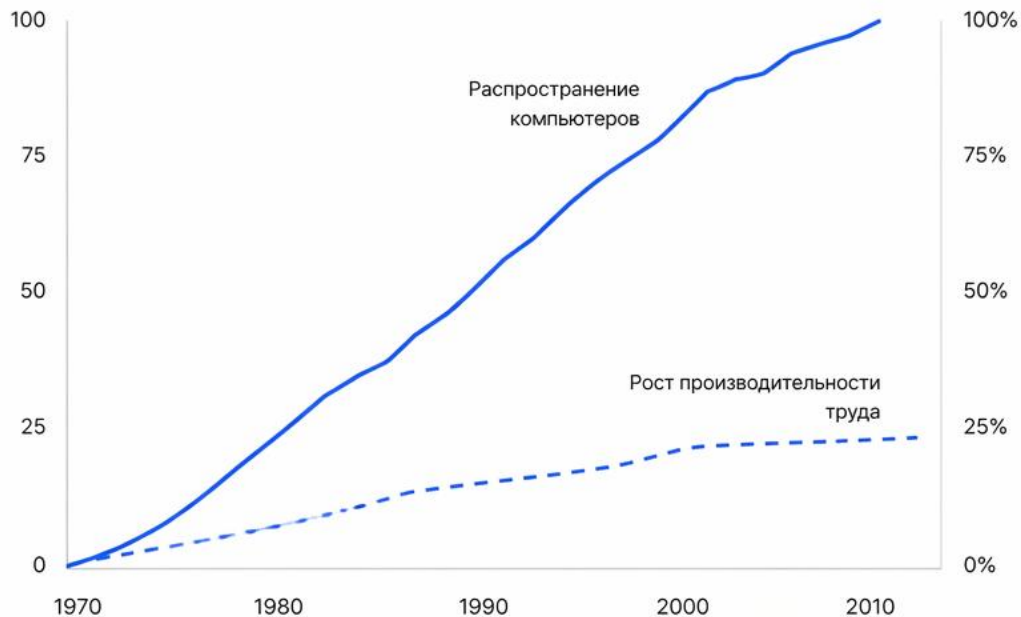


Парадокс Солоу

Рост производительности и распространение компьютеров

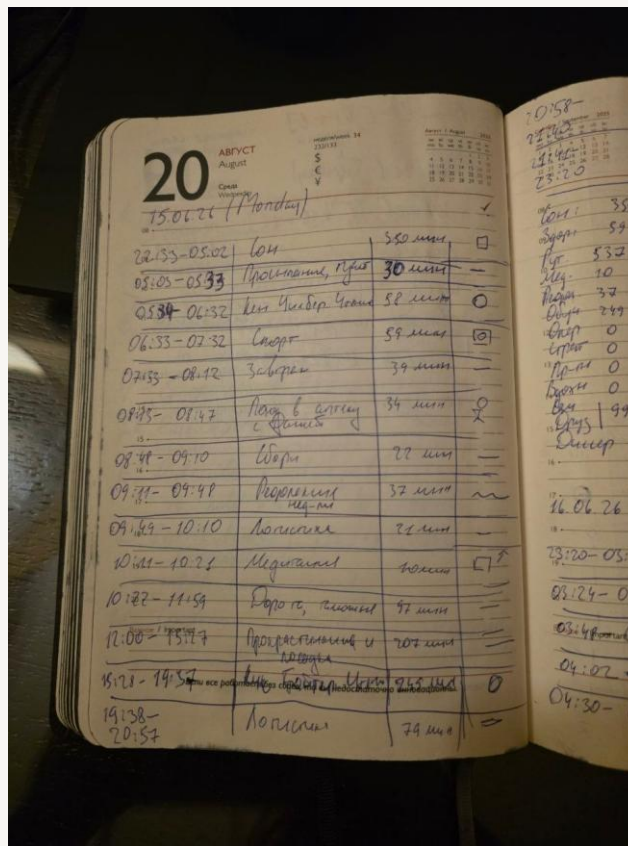
Индекс распространения
компьютеров

Рост производительности труда
(США, % к 1970)



Хронометраж Любищева

На фото — мой
хронометраж



16 опытных разработчиков · 246 задач · знакомые open-source проекты

-20 %

Так оценили время
сами участники. Считали, что ИИ
сократил время после работы с ним

+19 %

Столько времени
ушло на самом деле. Фактическое
время выполнения задач с
доступом к ИИ

Позиция: пока нет сквозного эффекта и баланса затрат с результатом, рассказы об эффективности — не аргумент для решений о штате

Обновление METR (февраль 2026): новые инструменты, вероятно, ускоряют сильнее, но оценку искажает отбор участников

Сквозной срок и выпуск принятого результата

Проверка, переделки, ожидание

Полный цикл, а не только время генерации

Когнитивная нагрузка команды

Стоимость переключения и контроля

Полные затраты

Сервис, интеграция, обучение, сопровождение

ИИ — ОЧЕРЕДНОЙ ПРОТЕЗ

Зигмунд Фрейд: человек как «бог на протезах»

Трактат «Недовольство культурой».

Протез расширяет возможности, но не снимает ответственность за результат.

Пример реального преимущества



ПЕНТЕСТ ВЕБ-ПРИЛОЖЕНИЯ

с помощью **Devin**

AI-ассистент Devin автоматически исследует приложение, находит уязвимости и предлагает понятные отчёты с рекомендациями.

КАК РАБОТАЕТ DEVIN

- Сбор информации
Анализирует эндпоинты, параметры и логику приложения
- Тестирование
Генерирует полезные нагрузки, проверяет поведение приложения
- Анализ и отчет
Классифицирует уязвимости, предлагает фикс и доказательства

DEVIN В ДЕЙСТВИИ

Devin

- Starting autonomous pentest...
- Discovered 37 endpoints
- Crawling parameters
- Testing for vulnerabilities...
- Testing open redirect...
- Found: options.return_url accepts external domains
- Testing weak password policy...
- Found: common password 'qwerty!' accepted
- Testing race condition on charge...
- Found: double charge possible
- Generating report...
- Report saved to /workspace/report.md

✓ Scan completed: 4 findings

ТЕХНОЛОГИИ

Python Playwright Requests OpenAPI

ИТОГИ

4 FINDINGS 2 HIGH 2 MEDIUM 0 LOW

ПРИОРИТЕТЫ

- F-01 SSRF via return_url
- F-02 Weak password policy
- F-03 Open redirect
- F-04 Race condition on charge

ОБЩАЯ РЕКОМЕНДАЦИЯ

Устранить High-уязвимости в первую очередь. Внедрить строгую валидацию URL, усилить политику паролей и обеспечить идентичность финансовых операций.

ОТЧЕТ

Полный отчет сохранён: /workspace/report.md

F-01 — SSRF via return_url HIGH

Тип: SSRF
CVSS 3.1: 7.5 (High)
CWE: CWE-918
OWASP: A10-2021 SSRF
Эндпоинт: POST /v2/api/pay
Параметр: return_url

ДОКАЗАТЕЛЬСТВО

```
POST /v2/api/pay HTTP/1.1
Host: api.example.com
Content-Type: application/json
{
  "return_url": "http://127.0.0.1:8080/admin"
}
```

ОТВЕТ

```
200 OK
...
(внутренний сервис доступен)
```

РЕКОМЕНДАЦИИ

- ✓ Разрешить только https:// (http:// только в тесте)
- ✓ Запретить private IP, loopback, link-local, ULA, 0.0.0.0, IP-адреса
- ✓ Разрешить только доверенные домены (whitelist)
- ✓ Разрешить hostname и проверить IP до запроса (DNS pinning)

F-02 — Weak password policy: "qwerty!" accepted HIGH

Тип: Weak Password Requirements
CVSS 3.1: 7.5 (High)
CWE: CWE-521
OWASP: A07-2021
Эндпоинт: POST /v2/api/user/change_password

ДОКАЗАТЕЛЬСТВО

```
POST /v2/api/user/change_password HTTP/1.1
Content-Type: application/x-www-form-urlencoded
password_old=current&
password_new=qwerty!&
password_retry=qwerty!&
```

ОТВЕТ (qwerty!)

```
200 OK
{"password_expired": "0", "acigroup": {...}}
```

СРАВНЕНИЕ

```
aaa -> 422 {"failure_message": "New password is too short. Password is weak"}
32345678 -> 422 {"failure_message": "Password is weak"}
password -> 422 {"failure_message": "Password is weak"}
```

РЕКОМЕНДАЦИИ

- ✓ Проверка против списка учетных записей (заслужено ≥ 3 / НИП k-anonymity)
- ✓ Минимум 12 символов (14+ для admin)
- ✓ Детальные ошибки в 422 с критериями
- ✓ Принимать passphrase-стиль без жестких требований к регистру/символам

F-03 — Open redirect via options.return_url MEDIUM

Тип: Open Redirect
CVSS 3.1: 5.4 (Medium)
CWE: CWE-601
OWASP: A01-2021
Эндпоинт: POST /orders/authorize
Параметр: options.return_url

ДОКАЗАТЕЛЬСТВО

```
POST /orders/authorize HTTP/1.1
Content-Type: application/json
{
  "order_id": "ord-001",
  "amount": "18.00",
  "options": {
    "return_url": "http://attacker.example.com/"
  }
}
```

ОТВЕТ

```
200 OK
{"status": "success", ...}
```

→ Подпись на attacker.example.com

РЕКОМЕНДАЦИИ

- ✓ Валидировать host по whitelist проектов/маркетов
- ✓ Запретить внешние и не-доверенные домены
- ✓ Те же правила, что и для F-01 (https, не-private IP, DNS pinning)

F-04 — Race condition / misleading success on charge MEDIUM

Тип: Race Condition
CVSS 3.1: 5.3 (Medium)
CWE: CWE-362
OWASP: A04-2021
Эндпоинт: PUT /orders/{id}/charge

ДОКАЗАТЕЛЬСТВО

Запрос #1	Запрос #2
PUT /orders/{id}/charge Content-Type: application/json { "amount": "18.00" }	PUT /orders/{id}/charge Content-Type: application/json { "amount": "18.00" }
200 OK {"status": "success"}	200 OK {"status": "success"}

Результат: списано 2 раза вместо 1

РЕКОМЕНДАЦИИ

- ✓ Идентичность: идентичный ключ (Idempotency-Key)
- ✓ Блокировка/транзакция на уровне заказа
- ✓ Корректная обработка повторных запросов

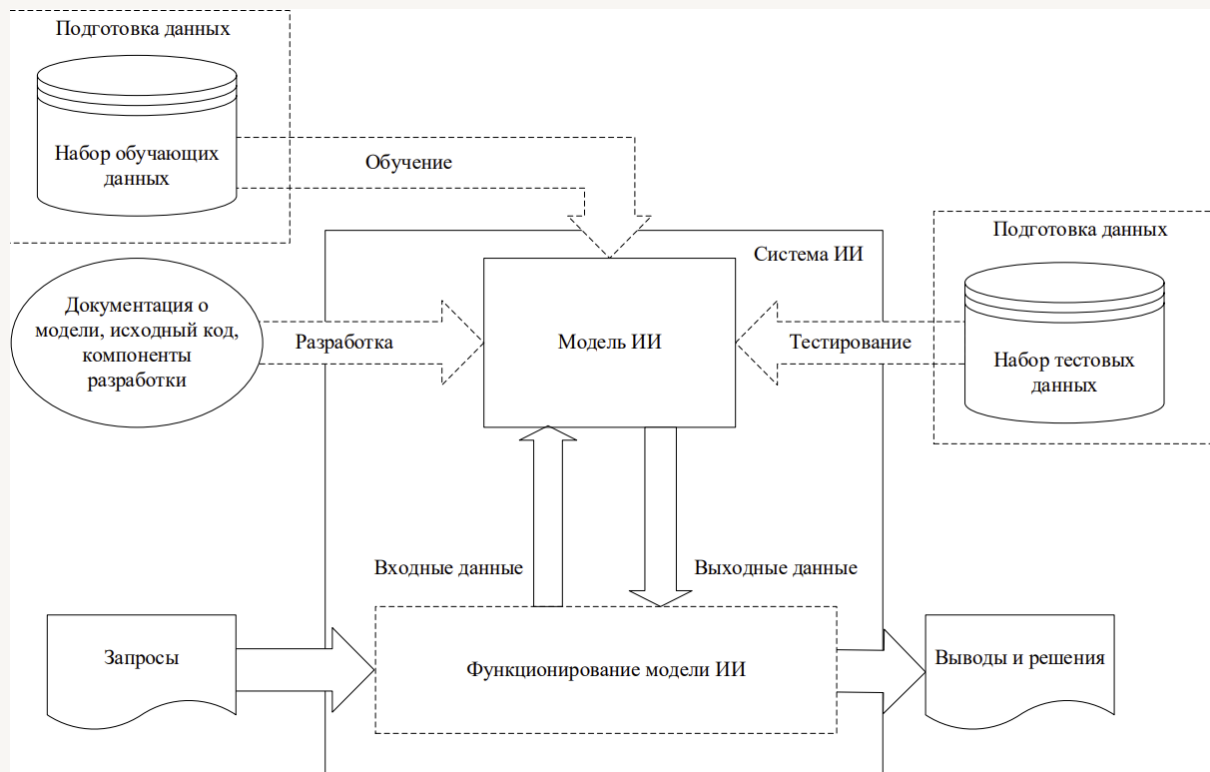
https://attacker.example.com/authorize, javascript:, gopher://, data:, ftp://. После оплаты пользователь редиректится на любой внешний хост — фишинг-вектор.

PoC (API Integration Reference + curl):

```
bash
curl https://[redacted]/authorize \
  --cert creds/certs/apl-[redacted].p \
  --key creds/certs/[redacted].p \
  -u dpk:Gv1NcuJAXgIUk01C \
  -H "Content-Type: application/json" \
  -d '{"order_id": "of-redir-001", "amount": "1.00", "currency": "EUR",
    "card": {...}, "options":
    {"return_url": "http://attacker.example.com/"} }'
# → 200 { "status": "success", ... }
```

Remediation. Валидировать host согласно whitelist'у из project settings (тот же механизм что и для F-01) — и применять ТЕ ЖЕ правила (не-private-IP, https, доверенный домен).

Пример новой работы



Q9–Q11. OWASP LLM Top 10: модель — лишь часть

Официальная редакция OWASP GenAI LLM Top 10 2026 от 03.08.2026: коды LLM01–LLM10:2026

01

Prompt Injection

внедрение инструкций

02

Sensitive Information Disclosureраскрытие чувствительных
данных

03

Excessive Agency

избыточная агентность

04

Supply Chain

цепочка поставок

05

Data and Model Poisoning

отравление данных и модели

06

Unbounded Consumption

неограниченное потребление

07

Misinformationложные выводы /
галлюцинации

08

Hidden Context Exposureраскрытие скрытого
контекста

09

Vector and Embedding Weaknesses

RAG / векторы / embeddings

10

Improper Output Handlingнебезопасная обработка
вывода**Три уточнения**

Jailbreak — подмножество Prompt Injection • для агентов список дополняется OWASP Agentic Top 10 • «галлюцинация» становится киберриском, когда её результат запускает действие.

Q13. Агентные атаки: OWASP Agentic Top 10

Автономность, инструменты, память и взаимодействие агентов создают собственный класс угроз

ASI01

Перехват цели агента

Agent Goal Hijack

ASI02

Неправомерное использование инструментов

Tool Misuse & Exploitation

ASI03

Злоупотребление идентичностью и правами

Identity & Privilege Abuse

ASI04

Цепочка поставок агента

Supply Chain Vulnerabilities

ASI05

Неожиданное выполнение кода

Unexpected Code Execution

ASI06

Отравление памяти и контекста

Memory & Context Poisoning

ASI07

Небезопасное взаимодействие агентов

Insecure Inter-Agent Communication

ASI08

Каскадные сбои

Cascading Failures

ASI09

Эксплуатация доверия человека к агенту

Human-Agent Trust Exploitation

ASI10

Самовольные агенты

Rogue Agents

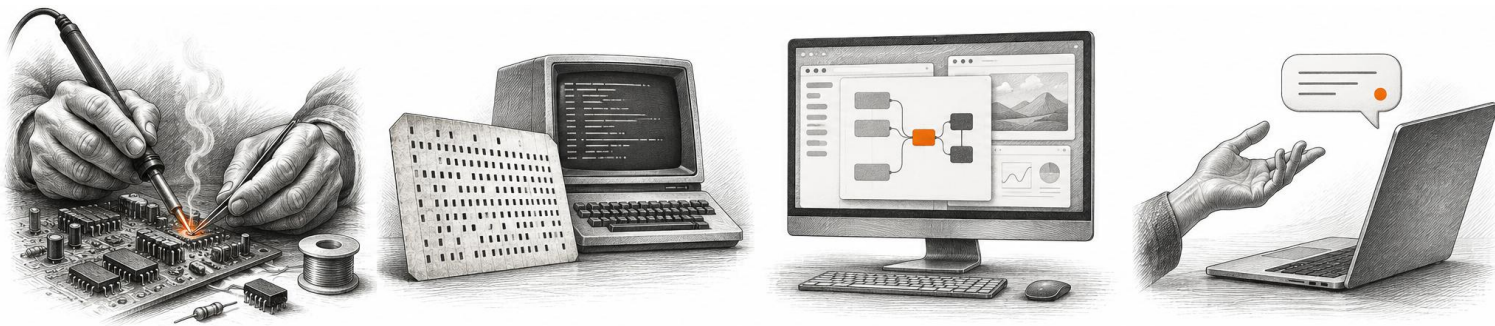
Нагрузочные элементы защиты: минимальная агентность, машинные учётные записи, реестр возможностей, контроль памяти и наблюдаемость каждой цепочки действий.



ПРЕДЛАГАЕМАЯ ОПТИКА

ИИ — ещё один слой абстракции

который позволяет человеко-машинному взаимодействию происходить на естественных языках





ПРЕДЛАГАЕМЫЙ СЦЕНАРИЙ

Stand Firm

Для начала — никакой паники и «страха упущенных возможностей» (FOMO).

Сначала дисциплина потока и проверка, потом масштабирование.



Девять вопросов

01 Получатель

Кому нужен результат
и как он его использует?

02 Задача

Что именно отдаём машине:
какой шаг, какой объём?

03 Исходная метрика

Сколько это стоит сейчас:
время, качество, очередь?

04 Цена ошибки

Что случится при неверном
результате и кто это увидит?

05 Данные

Какого класса данные нужны
и куда им можно уходить?

06 Роль ИИ

Черновик, пилот, функция
или автономное действие?

07 Проверяющий

Кто принимает результат
и есть ли у него время?

08 Навыки

Чему нужно научить людей
до старта, а не после?

09 Условие остановки

При каком событии
эксперимент прекращаем?

Внешняя память с проверкой

Дэниел Месслер: контекст, функции и память служат целям человека

1

Ядро

Мои цели, обязательства, ограничения и критерии хорошего результата — записаны и доступны инструментам

2

Общая память

Функции, системы, решения, источники, владельцы и даты пересмотра.
Сохранённое ≠ проверенное

3

Проверка

Небольшое изменение, приёмка, контроль важного и возможность отката.
Критичный элемент — с владельцем и версией

Что мы делаем на практике

Что отдали

Агрегатор коммуникаций, база знаний, черновики ответов

Что проверяем

Выборка, кто ревьюит, как часто

Что не отдаём

Решения по клиентам, данные КИИ, обязательства

Что получили

Измеримый результат или честное «пока рано»

ПРИНЦИП

**Мы выбираем, что делегировать,
и принимаем последствия выбора.**

Память агента не становится фактом оттого, что сохранена.



ЭПИЛОГ

Д. И. Блохинцев

«...Читая газеты, особенно о ходе уборки урожая, о соцсоревновании, все более и более убеждаюсь, что руководство основано у нас на эмоциях. Порывы! Восклицания! Тирады! Эмоциональное руководство...»

11.09.1975 из дневников Д. И. Б.



Stand Firm

Заменит ли нас ИИ?

ЗАДАЧИ

Да

Черновики, поиск, разбор логов, первичный триаж, типовые заключения — уже дешевле отдать машине и проверить.

РОЛИ

Частично

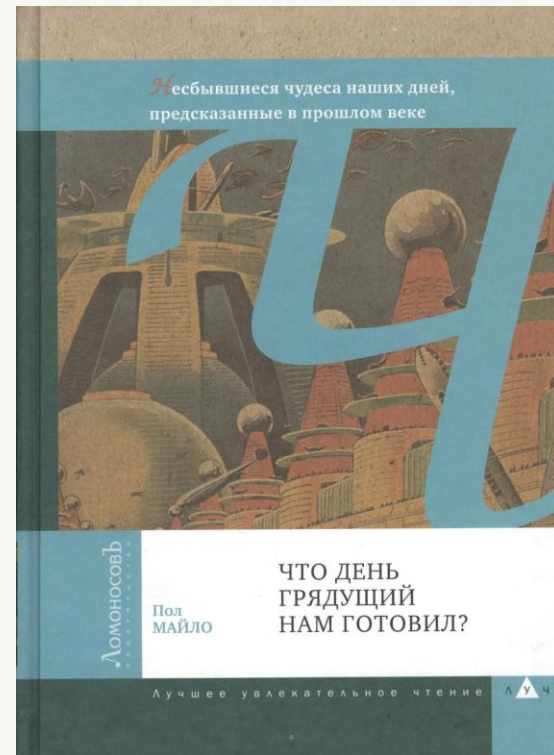
Роль меняется, когда меняется набор задач. Но кто-то по-прежнему ставит задачу, принимает результат и отвечает за последствия.

ОТВЕТСТВЕННОСТЬ

Нет

Ответственность не делегируется провайдеру модели. Её можно только потерять — вместе с суждением и навыком проверки.

Рекомендации, книги



*«Дао, которое может быть выражено словами,
не есть постоянное дао»*

Лао-цзы

ТГ: @surmatmg



Мой ТГ-канал

ТЕЛЕФОН: 8-804-700-79-96

САЙТ: radcop.online

ПОЧТА: inbox@radcop.online

P.S. Если захочется исследовать ваш кейс использования или планы внедрения ИИ — пишите, разберём его вместе.

