МОДУЛЬ 5

AGENTS

«Agent» — везде.

1

Дарио Амодей

год — заменят сотрудников

2

Андрей Карпати

десять лет до зрелости

3

Ян ЛеКун

путь тупиковый, нужен другой

Есть ощущение, что отстаёшь.

После лекции — 3 вопроса в голове и рабочая настройка в руках.

Что такое agent

Anthropic, инженерная статья 2024.

Agent = ИИ сам выбирает, что делать дальше.

Agent vs workflow

ВОПРОС	WORKFLOW	AGENT
Кто решает порядок шагов	человек	ИИ
Можно нарисовать схему до запуска	да	нет
Цикл с обратной связью	нет	да

Что НЕ приписывать агенту

Большая языковая модель — не живое существо.

Оболочка из накопленных текстов.
Внешне — живое. Внутри — отражение.

Агент не «решает» — продолжает по данным.

ЧАСТЬ А · 10 МИНУТ

Минимальный агент с нуля

PRODUCT SENSE

Шаг 0: пустая обвязка. Агент видит только модель.

read_file

+1

Агент: «напиш себе новый инструмент».

ВЫЗОВЫ ВИДНЫ

+1

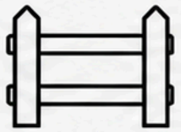
Шаг 3 — добавляем bash. Агент читает ОС, запускает команды.

Это ИИ-модель,
цикл
и достаточно токенов.

Магии нет. Значит можно настраивать.

3 ВОПРОСА ПЕРЕД ДЕЛЕГИРОВАНИЕМ

01



ГРАНИЦА

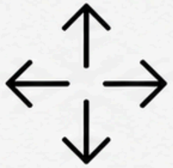
какие шаги сам, на каких остановится и спросит

02

ОБВЯЗКА

из чего собрана способность действовать

03



ЧТО УСИЛИТСЯ

польза, слепые зоны, цена ошибки × N

Все три вместе — рамка работает. Без одного — рассыпается.

3 вопроса перед делегированием: граница / обвязка / × N

Вопросы?

Граница решений

Какие шаги агент делает без вопросов. Где останавливается.

Пример: карточка фичи

САМ	СПРОСИТЬ
Формулировки	Добавление в общий список
Чек по шаблону	Отправка стейкхолдеру

Не поставили — агент идёт до конца. Узнаёшь поздно.

Граница = контракт из M1, перенесённый на каждый шаг. Но не всегда текстовые

ВОПРОС 2 · 4 МИН

Обвязка агента

PRODUCT SENSE

Сама модель умеет одно — продолжать текст.

К модели приделывают части. Вместе — обвязка / harness.

Из чего собран агент: модель + 3 части обвязки

4 части обвязки

ЧАСТЬ	ЧТО ЭТО
Мозг	сама ИИ-модель
Инструменты	что агент умеет вызывать
Песочница	где работает и что там может
Память	что помнит между шагами

Знаете обвязку — видите где сломается.

Что умножается × N

Делегирование умножает — польза, слепые зоны, цена ошибки

Что сдвинулось за полгода

Карпати, октябрь 2025: 4 проблемы агентов.

За полгода часть закрыли.

4 изменения за полгода

ЧТО СДВИНУЛОСЬ	ГДЕ
Computer Use стабилен	Anthropic Q1 2026
Auto Mode без подтверждений	Claude Code, апрель 2026
Subagents в параллель	Claude Code / Codex
A2A между производителями	Microsoft Copilot Studio, апрель 2026

Что не закрыто

01

Длинные задачи (часами): теряет контекст

02

Накопление опыта между запусками: каждый раз с нуля

Слаб хотя бы в одном — это помощник, не сотрудник.

5 настроек в инструментах

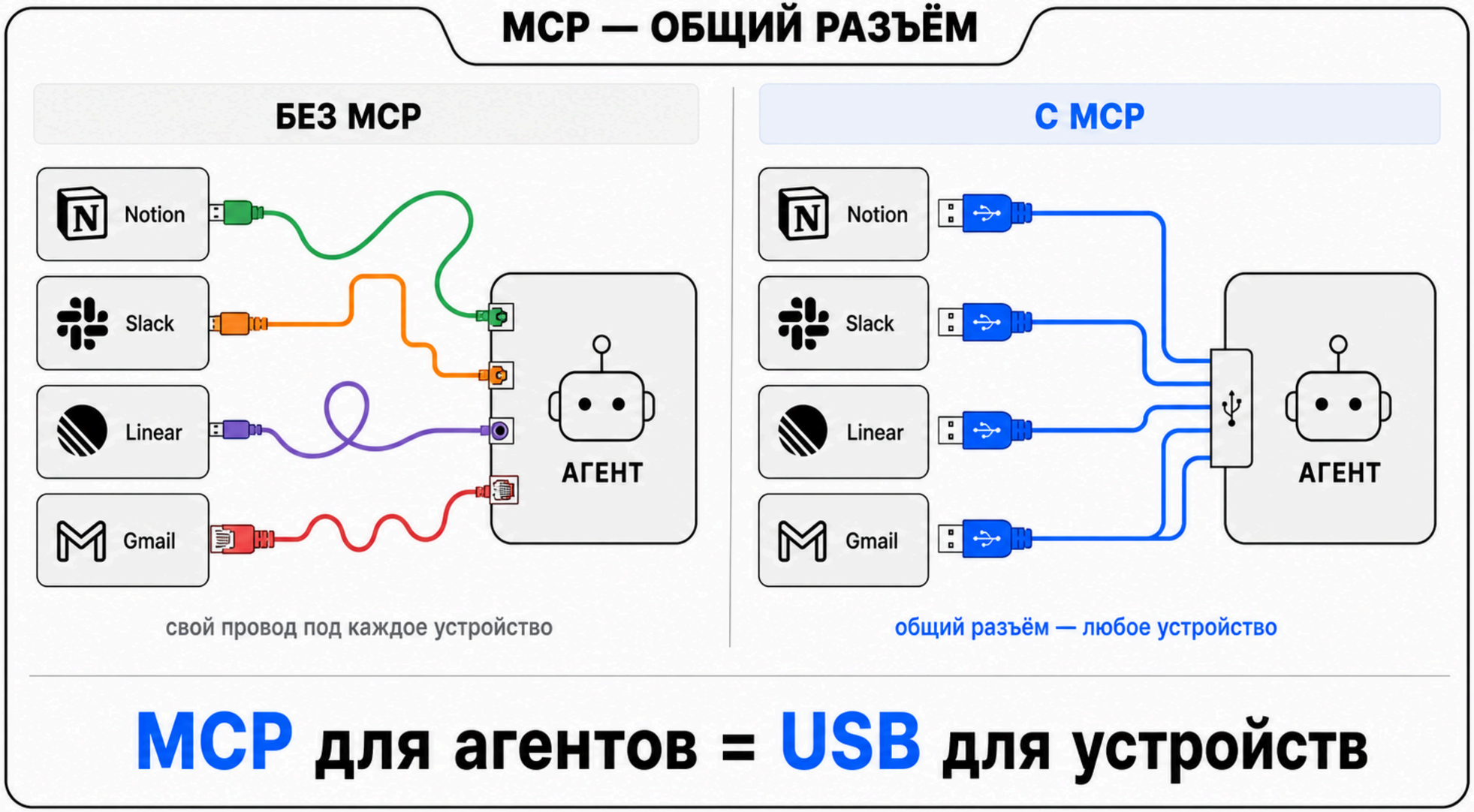
5 настроек

ЧТО	ГДЕ ЖИВЁТ
Подтверждение перед действием	по умолчанию в Claude Code / Codex
Разрешения по типу действий	чтение / запись / сеть / команды
Постоянные инструкции файлом	CLAUDE.md / AGENTS.md
Саб-агенты в параллель	Subagents / фоновые задачи
Подключение внешних инструментов	MCP — open standard

На каждую скоро посмотрим

MCP — общий разъём

Один сервер для Notion → любой агент, понимающий MCP.



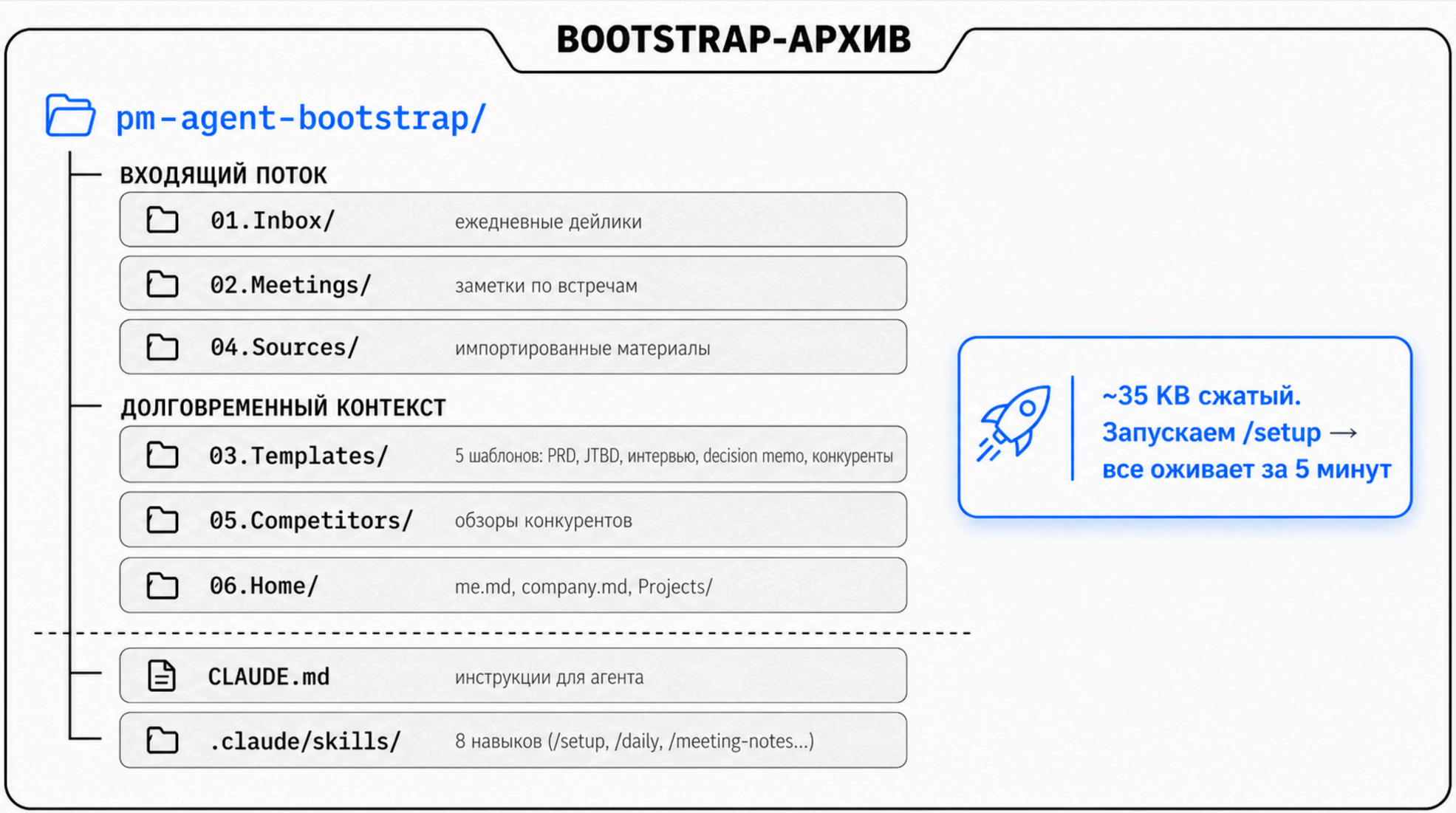
MCP — общий разъём для агента и внешних инструментов

ЧАСТЬ Б · 55 МИНУТ

Препарируем устройство агента

PRODUCT SENSE

Холодный архив



Архив: 6 папок + CLAUDE.md + 8 SKILL.md + 5 шаблонов

6 папок

ПАПКА	ЧТО ВНУТРИ
01.Inbox/	ежедневные дейлики
02.Meetings/	заметки по встречам
03.Templates/	5 шаблонов
04.Sources/	внешние материалы
05.Competitors/	обзоры конкурентов
06.Home/	корневой контекст: me.md, company.md

Живой /setup

15 вопросов по одному

РАЗДЕЛ	ВОПРОСОВ
Про автора	5
Про продукт	6
Про текущий проект	4

Чат-чек 2: что увидели? Что произошло?

Мозг — выбор модели

В SKILL.md — поле model:

Точечный выбор

DECISION-MEMO	DAILY
model: opus	`model: haiku`
глубокое мышление	`дешёвая рутина`

Глобально — `claude --model` opus|sonnet|haiku.

Алиасы (2026-05)

01

opus / sonnet / haiku

02

opusplan — Opus планирует, Sonnet исполняет

03

opus[1m] / sonnet[1m] — расширенное окно

Дёшево/быстро или умно/дорого — выбираете сами.

Инструменты + MCP

B SKILL.md — allowed-tools.

Скобки = whitelist параметров. Bash(git *) разрешает git status, не rm -rf.

MCP-сервер для Яндекс.Почты

ФУНКЦИЯ	ЧТО ДЕЛАЕТ
list_folders / search_emails	найти папки и письма
read_email / download_attachment	прочитать
send_email	отправить
move_email / delete_email	переместить / удалить

Встроенные предохранители

01

Allowlist получателей

02

Rate-limit на отправку

03

Детекция prompt-injection во входящих

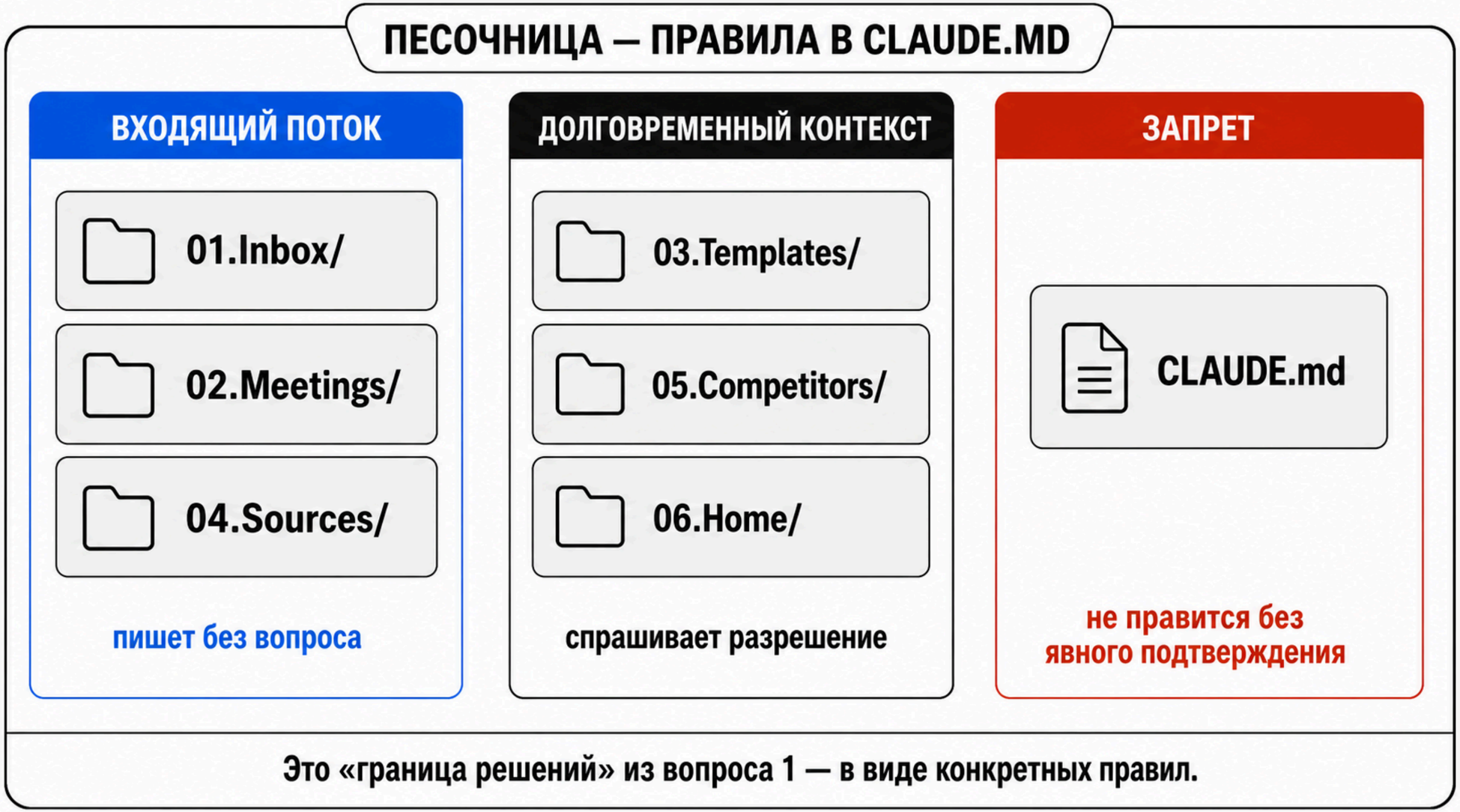
04

Валидация CRLF

Не «может всё» — ровно вот это.

Дальше — граница

Песочница



Песочница: правила в CLAUDE.md

Деление по риску записи

КЛАСС	ПАПКИ	ПОВЕДЕНИЕ
Входящий поток	01 / 02 / 04	пишет без вопроса
Долговременный	03 / 05 / 06	спрашивает
CLAUDE.md	корень	запрещено

У Codex — 3 режима

01

read-only — только чтение, по умолчанию

02

workspace-write — запись в текущую папку, без сети

03

danger-full-access — без ограничений, только в контейнере

«Дать больше свободы» — это не решение разработчика, а конкретные настройки.

Память

Две части памяти

CLAUDE.MD	ПАПКА НА ДИСКЕ
устойчивый контракт	вся история работы
грузится в начале сессии	ничего не теряется между запусками

строк / 25KB лимит

200

CLAUDE.md иерархия — 4 уровня

01

Managed

02

Project
./CLAUDE.md

03

User
~/.claude/CLAUDE.md

04

Local
./CLAUDE.local.md

Та же папка из М3 — теперь агент читает её сам.

Навыки

Не пятый компонент — отдельный слой поверх четырёх.

YAML-шапка SKILL.md

ПОЛЕ	ЧТО ЗАДАЁТ
description	триггер-фразы
allowed-tools	разрешённые инструменты
arguments	$name — подставляется при вызове
model / effort	модель + сила «думания»

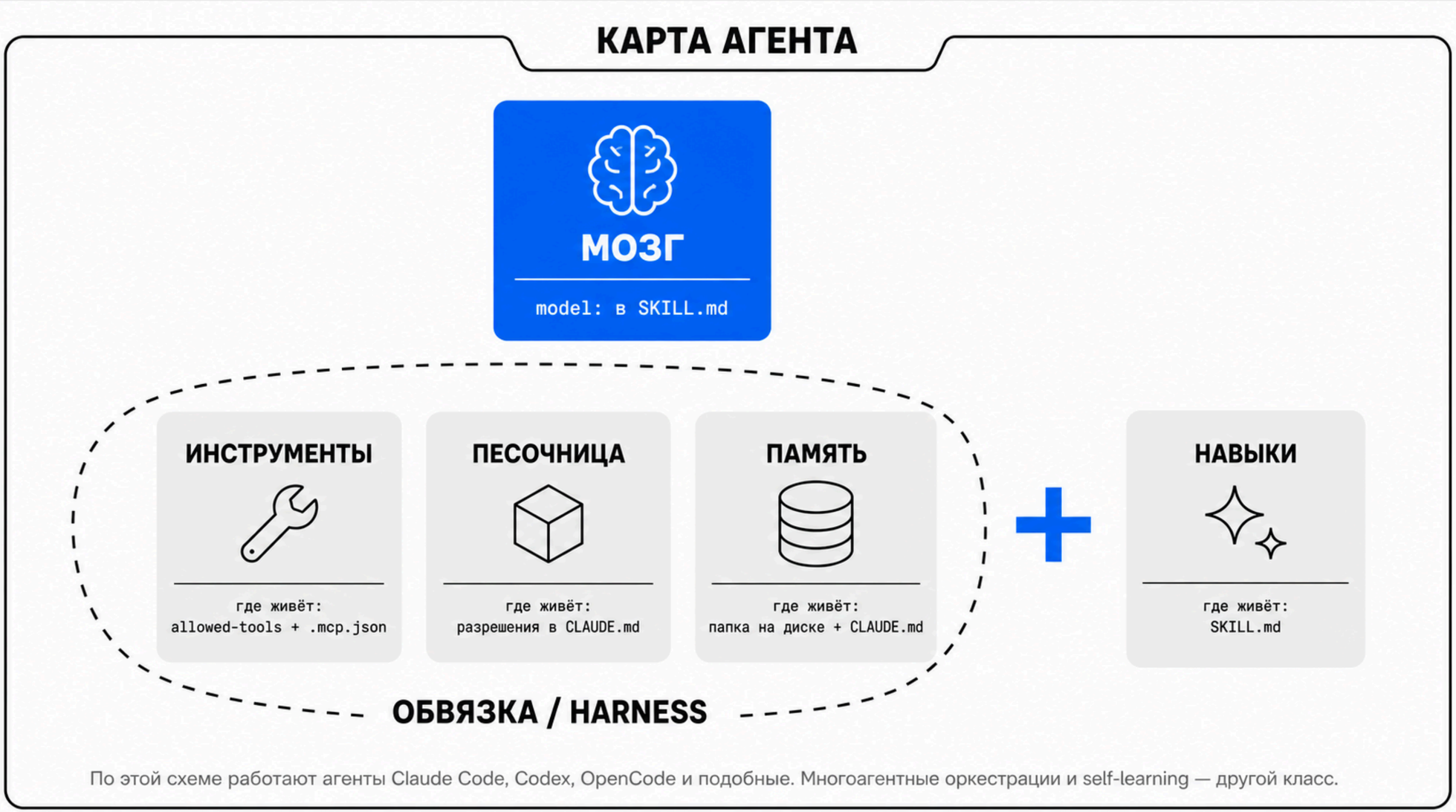
На старте грузятся только описания. Тело — когда вызван.

Сводная карта

Разобрали 4 компонента + слой навыков.

Карта в одной строке

КОМПОНЕНТ	ГДЕ ЖИВЁТ
Мозг	поле model: в SKILL.md
Инструменты	allowed-tools + MCP
Песочница	разрешения в CLAUDE.md
Память	папка + CLAUDE.md как контракт
Навыки	SKILL.md — собирает 4 компонента



Карта агента: мозг + инструменты + песочница + память + навыки

Чат-чек 3: какой компонент критичнее для вашей задачи?

Прогон /competitor-scan

/competitor-scan ...

Что наблюдаем

ШАГ	КОМПОНЕНТ
Читает Home/me.md, company.md	память
Спрашивает разрешение на сайт	граница / песочница
Собирает таблицу через модель	мозг + инструменты
Пишет черновик в Competitors/	память + правила записи

Прогон /competitor-scan: что наблюдаем по компонентам

Прогон /meeting-notes

/meeting-notes на готовом транскрипте. Карта универсальна.

А где же автономия

"The technique is deterministically bad in an undeterministic world."

Ralph-режим — Codex / OpenCode.
Одна команда → агент сам отрабатывает план.

Auto Mode — Claude Code, апрель 2026. Снимает подтверждения на безопасных операциях.

3 режима автономии

РЕЖИМ	ЧТО ДЕЛАЕТ
Ralph	одна команда → план сам
Auto Mode	без подтверждений на безопасных
dangerously-skip-permissions	полное снятие — без защитной обвязки не трогать

Это уже про управление автономией.
Через 5 дней в M6.

Что делать дальше

Bootstrap-архив — ZIP, ~35 КВ.

Ссылка в чате после лекции.

15-30 минут до рабочей настройки

Дома

01

Открыть свою папку-экзокортекс из М3

02

Распаковать архив

03

Запустить Claude Code, сказать «настрой меня» (/setup)

04

Прогнать одну реальную задачу: /competitor-scan или /meeting-notes

Через 10-15 минут — рабочая настройка под ваш продукт.

Куда ложатся модули М1–М4

M5 опирается на M1-M4: контракт → граница, классы → разрешения, экзокортекс → vault

PRODUCT SENSE

M1-M4 → M5

МОДУЛЬ	ЧТО СТАНОВИТСЯ В M5
M1 контракт	граница решений
M2 классы	правила разрешений
M3 экзокортекс	та же папка — теперь читает агент
M4 определение agent	точка отправления

Чат-чек 4: какую задачу прогоните первой?

Q&A

ДО ВСТРЕЧИ

M5 → M6