

Разработка с помощью агентов

Промышленный вайбкодинг



Дмитрий Колодзев, Промсофт



Я и мои слайды

- kolodezev.ru
- [free-llm-проxy](#) + пример настройки qwen
- [llm-kolodezev.pdf](#)
- [free-llm-проxy.md.txt](#)
- Датазавтраки
- Датафесты
- Капитанский Мостик
- ReliableML
- Канал в телеге



Кратко

- Спецификация → План → Код → Валидация → Повторить
- Инструменты: claude code, qwen, codex, opencode
- Модели: встроенные в инструменты, z.ai, openrouter
- Платно, бесплатно, свой сервер
- Решение проблемы КВН
- MCP сервера
- LSP сервер
- Береги контекст смолоду
- Изоляция, отдельные пользователи, Господи — жги!
- Навайбкодил тут

Если кто не в курсе

```
Qwen - free-llm-proxy
API Key | free-model (/model to change)
/projects/2026/free-llm-proxy

Tips: Используйте /bug, чтобы сообщить о проблемах разработчикам.

> привет

✦ Привет! Как я могу помочь тебе сегодня?

> В файле @spec/free-llm-proxy.md описана реализация простого прокси к LLM -
посмотри, проверь корректность реализации

✓ Read File Read file free-llm-proxy.md

✦ Анализирую спецификацию и проверяю реализацию в репозитории.

.: Провожу настройку методом тыка... (2s · ↓ 2 tokens · esc to cancel)

Введите сообщение или @путь/к/файлу

? — горячие клавиши 13.8% использовано
```

Эволюция

- Предсказать следующий символ → LLM
- Следующий символ в цикле → Подсказка
- Подсказка в цикле → Чат
- Чат в цикле → Агент
- Агент в цикле → Оркестратор? OpenClaw? SkyNet?

- Чат + RAG → Cursor
- Cursor — RAG — Интерфейс + grep → Claude
- Claude + loop → OpenClaw

Три составные части

- LLM (чем умнее, тем дешевле выходит)
 - Инструмент подготовки контекста (claude qween codex opencode)
 - Процессы (что спрашиваем и в каком порядке)
-
- Что с блокировками? OpenRouter, CCR, прокси, z.ai
 - Что с комплаенсом/NDA? CCR, свои модели
 - Что с деньгами? Будет дорожать. Локальное, бесплатное.

Наш выбор

- Qween Code + OpenRouter (дорого, быстро)
- Claude + z.ai max
- Спецификация → План → Код → Верификация

Процесс — новый проект

- Пишем, чего хотим достичь. Обсуждаем с людьми и правим
- Просим LLM уточнить и задать вопросы
- Обсуждаем с людьми и правим
- Просим LLM написать — как будем проверять
- Обсуждаем с людьми и правим
- Просим реализовать
- Смотрим, уточняем задачу
- Катим в стейдж, тестируем, общаемся
- Обвешиваем метриками и логами
- Катим и СМОТРИМ СМОТРИМ СМОТРИМ ААААААААА!

Изоляция

- На VPS
- В докере
- Отдельный пользователь linux
- Отдельный пользователь git
- Доступы для чтения в сервисы
- Умеет сходить на сервер по ssh — отдельный пользователь!
- Как с человеком, который иногда курит траву на рабочем месте

Смещение акцентов

- Вайбкодим ТЗ
- Тестируем ТЗ
- Кодирование дешево
- Постановка ТЗ дорого, но быстрая обратная связь
- Тестирование дорого — но оно и так дорого
- Становится заметнее то, что не связано с кодированием
- Выгораем быстрее — постоянно переключать контекст

Просите, и воздастся Вам

- Приведи спецификацию в соответствие
- Проверь покрытие тестов
- Проведи аудит проблем безопасности
- Где скорее всего возникнут проблемы
- Как мы справляемся с

AGENTS.MD

- QWEEN.md
- CLAUDE.MD
- Верхнеуровневой бриф
- Не мельтешите
- Несколько уровней
- Кстати [промнты](#)

Почитать и посмотреть

- [AI Coding Playbook](#)
- [OpenCode School](#)
- [Подводка про LLM в разработке](#)
- [Claude Code System Prompts](#)