Домашнее задание

Урок 22. Локальная ЛЛМ и ее интеграция с внешними системами через API

https://github.com/qw662008/llm-tlg

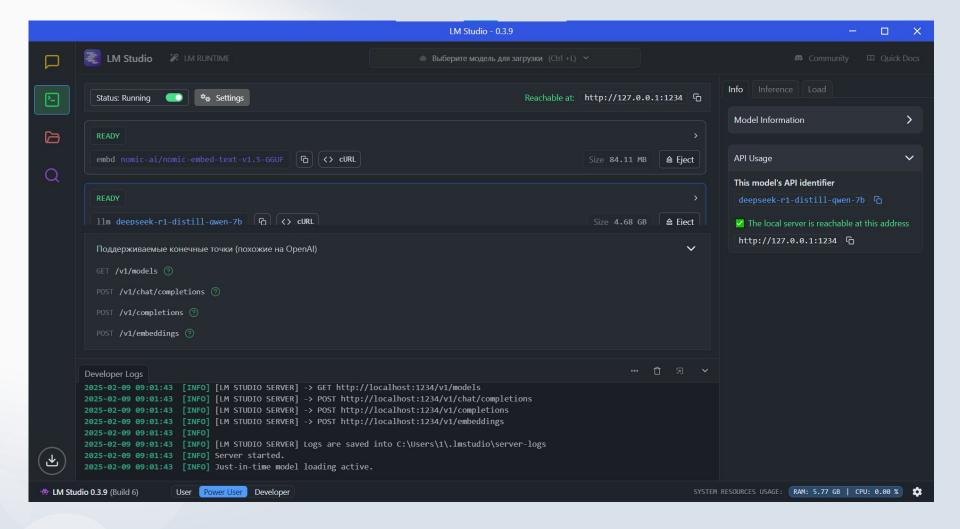
сервер LM Studio загружены:

- nomic-ai/nomic-embed-text-v1.5-GGUF
- · DeepSeek R1 distilled into Qwen 7B

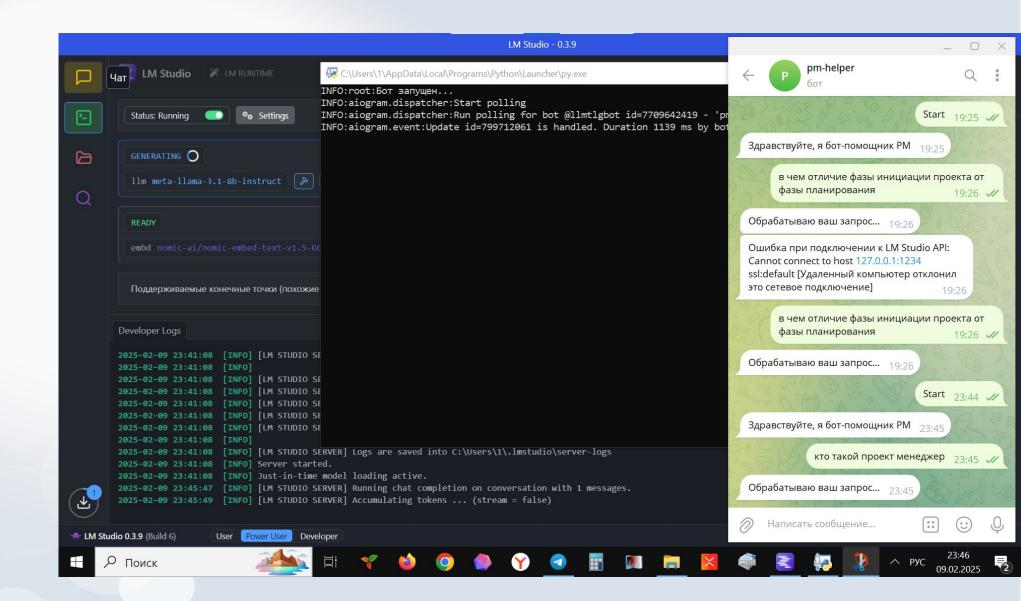
Один из первых вариантов. Рассуждения модели выбирали практически все окно контекста и в некоторых случаях превышали лимит сообщений Telegram.

в качестве материала для базы знаний использовалась методичка в формате PDF.

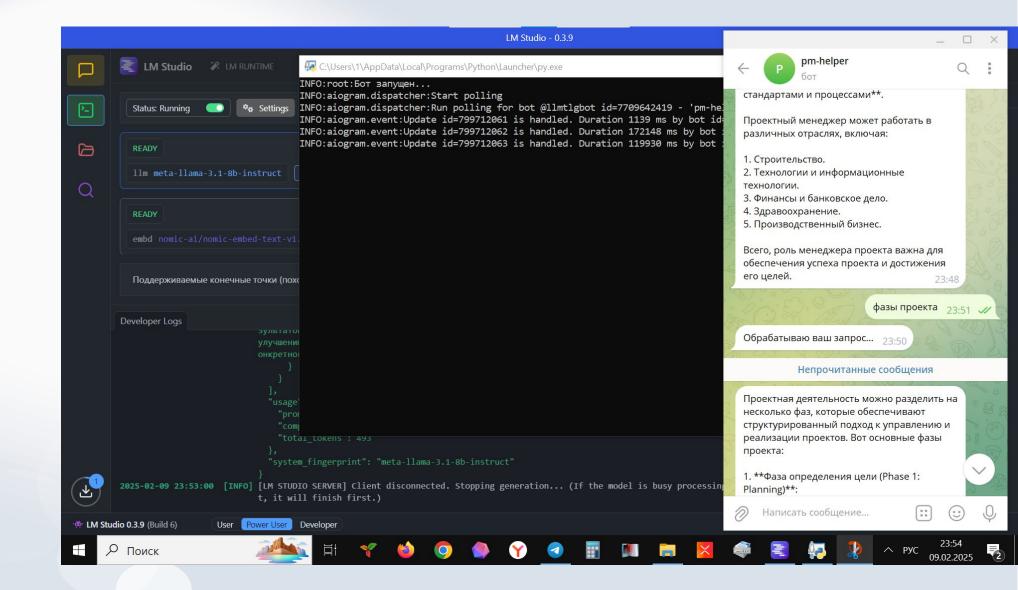
После перебора доступных по ресурсу баз пришлось остановиться на metallama-3.1-8b-instruct часть моделей путала русский китайский и английский, часть просто не справлялась с обработкой базы знаний.



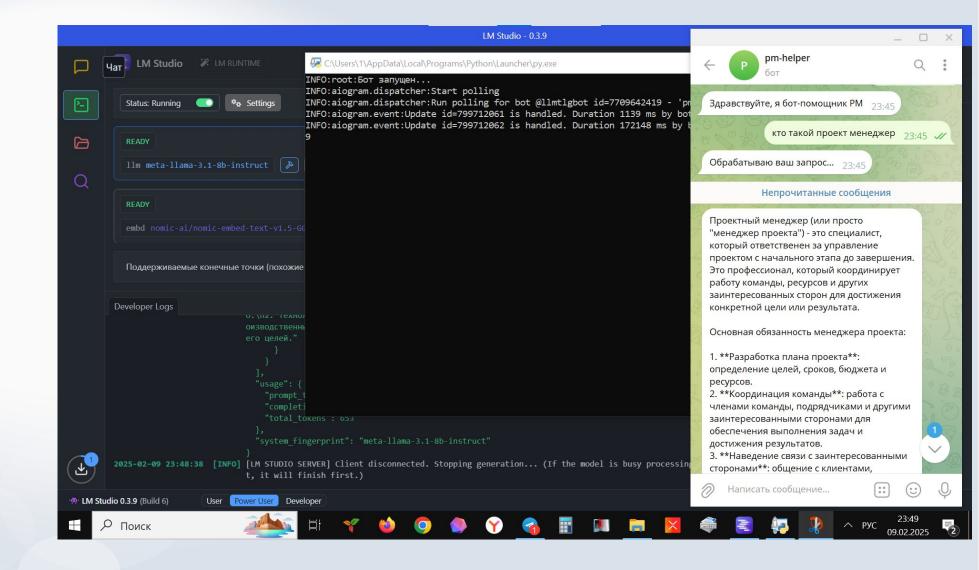
скрин сервера, запущенной программы и телеграм-бота.



скрин диалога

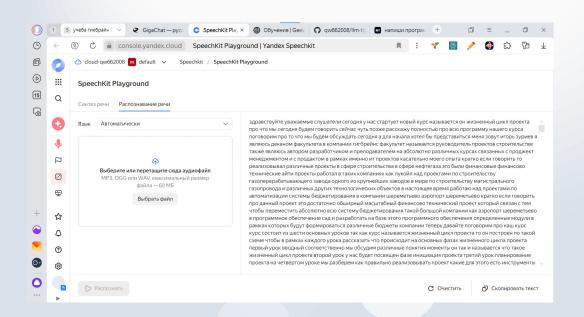


скрин диалога



комментарии:

первоначальная идея сделать базу знаний из текста лекции провалилась из-за превышения лимита токенов. SpeechKit Playgroud преобразовал mp3 в текст, но отказался работать дальше, типа больше 30000 знаков. Пришлось использовать для базы знаний методичку.



попытка написать код с помощью GigaChat тоже провалилась. Он упорно пытался использовать старую версию Aiogram и не смог переключиться на текущую версию. Пришлось использовать Claude Sonnet-3.5. С GitVerse решил не экспериментировать.

