

# Домашнее задание 2

# Описание базовое

- Задание на LLMOps + «вайб-кодинг», или скорее агентный кодинг
- Ваша задача:
  1. Развернуть локально небольшую LLM – которая помещается
  2. Использовать ее в любом инструменте для реализации фичи

# Поподробнее: LLMOps

- Необходимо развернуть LLM локально, любого размера, главное – с функционалом tool/function calling. Личная рекомендация: qwen-4b / gpt-oss-20b, можно квантированные версии.
- Поднять модель можно как угодно: olama, llama.cpp, lm studio, vllm, etc.
- Главное, чтобы на выходе был OpenAI-Compatible сервис локально (неважно на чем, GPU или CPU)
- Результат выполнения: скринкаст (видео), как вы дергаете модель локально через сваггер или UI (если инструмент с UI идет)

# Подробнее: вайб-кодинг

- Необходимо придумать и реализовать фичу самому для сервиса, который вы создали в первом ДЗ. Если первое ДЗ не делали – можно взять чужое либо просто болванку fastAPI сервиса взять просто.
- Фича должна быть не сильно просто, но и не сложной.
- Эту фичу вы должны реализовать с помощью агента
- Можно выбрать любого: RooCode в VSCode, aider, open code, etc.
- Данного агента вы должны подключить к локальной LLM, развернутой на прошлом шаге.

# Подробнее: вайб-кодинг

- Вы должны реализовать фичу через Plan&Act подход:
  1. План реализации фичи (который нужно сохранить отдельно дополнительно) ->
  2. Работа по плану ->
  3. Запуск линтеров и тестов (они должны быть в плане!!) ->
  4. Правка документации, ченджлога и других md файлов.
- Это может быть долго на локальной модели, но качество не особо важно.

# Подробнее: вайб-кодинг

- По итогу я должен увидеть:
  - CHANGELOG.md
  - Ваш AGENTS.md с правилами для агента
  - API.md – описание вашего API, созданное агентом (для контекста)
  - План реализации вашей фичи в виде md файла
  - Маленький скринкаст, где вы показываете историю чата в вашей системе (где я должен увидеть, как модель вызывает линтеры)

# Оценивание

- *Качество реализации НЕВАЖНО. Пусть будет фигня, с локальной моделькой это нормально. Но практика на большой модельке даст качественно другие результаты.*
- Подняли локально LLM и скинули скрин каст – 4 балла
- Реализовали фичу через агента - суммарно 6 баллов:
  - CHANGELOG.md + API.md – 1 балл
  - AGENTS.md с описанием правил, команд, стиля для вашего проекта – 2 балла
  - Скринкаст с историей чата – 3 балла

# Дедлайн

- Сдать до: **23:59 22.02**
- Сдать можно как в личку ТГ, так и в форме гугл:  
[https://docs.google.com/forms/d/e/1FAIpQLSfOW2o6N7pICM\\_yy6BRmgwugDbBkO2oh\\_GgPVH9aEF87XcEsQ/viewform?usp=publish-editor](https://docs.google.com/forms/d/e/1FAIpQLSfOW2o6N7pICM_yy6BRmgwugDbBkO2oh_GgPVH9aEF87XcEsQ/viewform?usp=publish-editor)