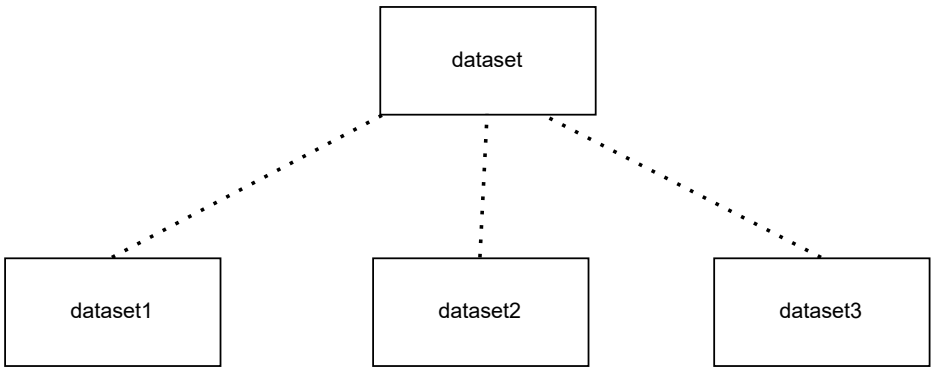
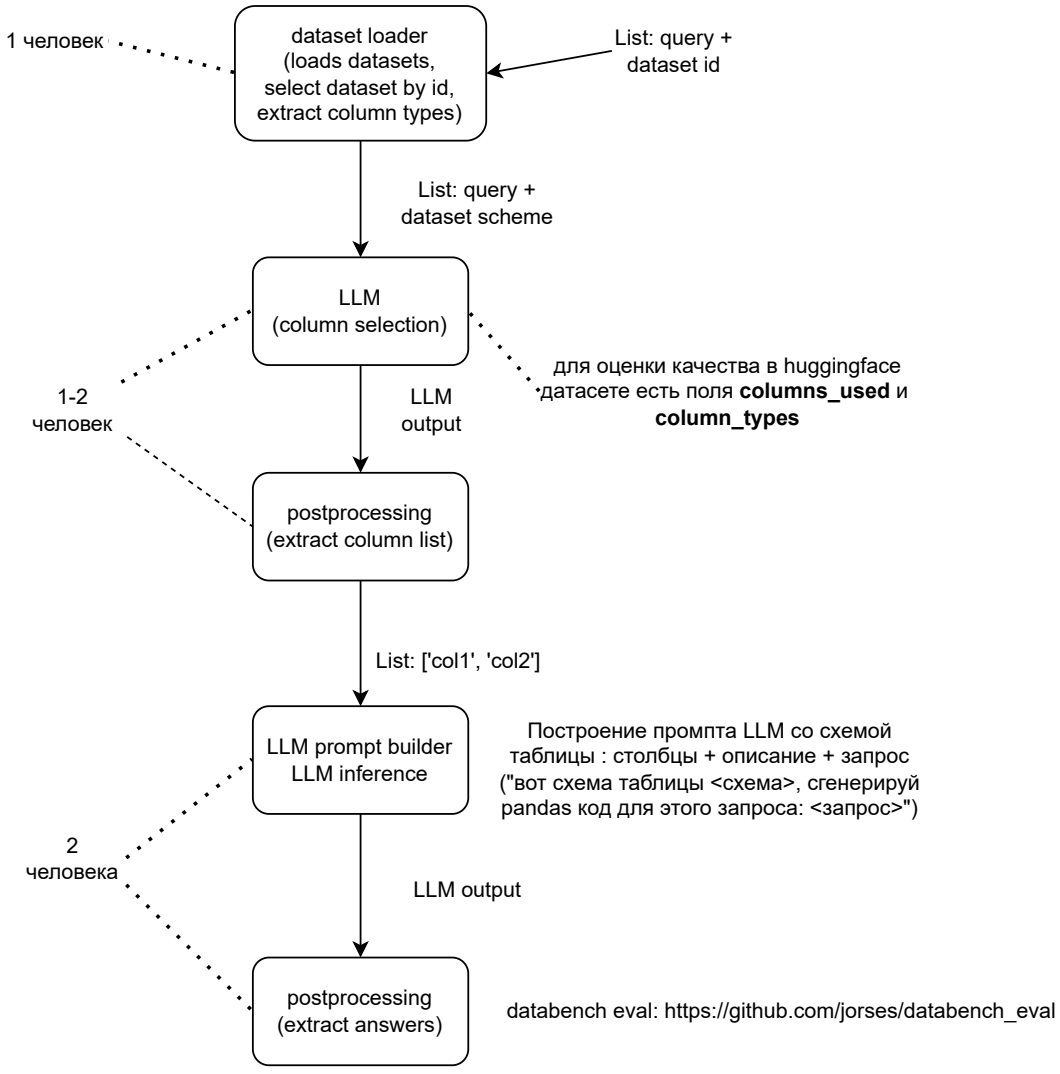
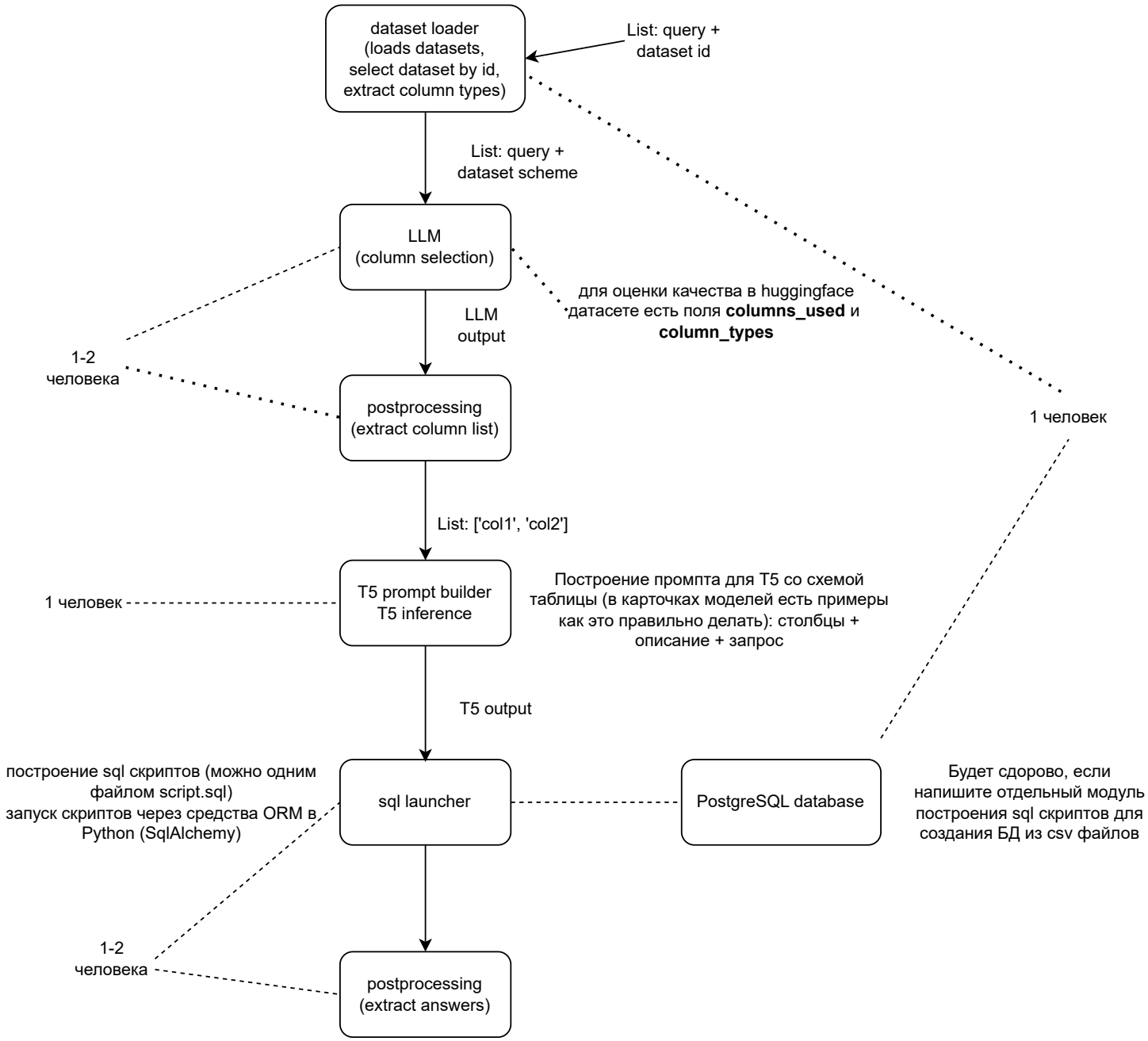
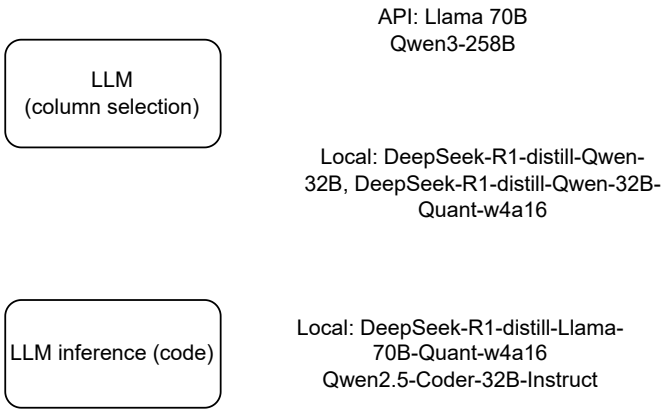

dataset
dataset1
dataset2
dataset3

LLM


1 человек
dataset loader
(loads datasets,
select dataset by id,
extract column types)
List: query +
dataset id
List: query +
dataset scheme
LLM
(column selection)
для оценки качества в huggingface
датасете есть поля **columns_used** и
column_types
1-2
человек
LLM
output
postprocessing
(extract column list)
List: ['col1', 'col2']
LLM prompt builder
LLM inference
Построение промпта LLM со схемой
таблицы : столбцы + описание + запрос
("вот схема таблицы <схема>, сгенерируй
pandas код для этого запроса: <запрос>")
2
человека
LLM output
postprocessing
(extract answers)
databench eval: https://github.com/jorses/databench_eval

T5


dataset loader
(loads datasets,
select dataset by id,
extract column types)
List: query +
dataset id
List: query +
dataset scheme
LLM
(column selection)
для оценки качества в huggingface
датасете есть поля **columns_used** и
column_types
LLM
output
1-2
человека
postprocessing
(extract column list)
1 человек
List: ['col1', 'col2']
1 человек
T5 prompt builder
T5 inference
Построение промпта для T5 со схемой
таблицы (в карточках моделей есть примеры
как это правильно делать): столбцы +
описание + запрос
T5 output
построение sql скриптов (можно одним
файлом script.sql)
запуск скриптов через средства ORM в
Python (SqlAlchemy)
sql launcher
PostgreSQL database
Будет сдорово, если
напишите отдельный модуль
построения sql скриптов для
создания БД из csv файлов
1-2
человека
postprocessing
(extract answers)



ссылки на модельки (Все нормально разворачиваются через vllm):

R1 distill Llama 70B int4: https://huggingface.co/RedHatAI/DeepSeek-R1-Distill-Llama-70B-quantized.w4a16

R1 distill Qwen 32B: https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-32B
R1 distill Qwen 32B int4: https://huggingface.co/RedHatAI/DeepSeek-R1-Distill-Qwen-32B-quantized.w4a16
Qwen 2.5 32B Coder: https://huggingface.co/Qwen/Qwen2.5-Coder-32B-Instruct

API (Qwen3, Llama 70B): https://yandex.cloud/ru/docs/foundation-models/concepts/yandexgpt/models