

# Сравнение LLM+RAG систем на основе Qdrant и rus-xquadqa

Афанасьев Алексей Игоревич 217 группа ВМК МГУ

12.04.2025

## 1 Методология

В данном исследовании сравниваются три RAG системы, использующие векторную базу данных Qdrant с эмбедингами модели `intfloat/multilingual-e5-large` и датасетом `rus-xquadqa`. Системы отличаются только используемыми языковыми моделями:

- OpenChat-3.5-1210
- Qwen-3B
- RuadaptQwen-7B-Instruct

Для оценки использовались 10 вопросов из датасета. Точность (ассигасу) рассчитывалась как доля правильных ответов.

## 2 Результаты

Таблица 1: Сравнение моделей по точности

Модель	Точность
OpenChat-3.5-1210	0.7
Qwen-3B	0.7
RuadaptQwen-7B-Instruct	0.7

Таблица 2: Подробные результаты для OpenChat-3.5-1210

Вопрос	Ответ модели	Правильный ответ	Корректность
Сколько очков уступила защита Пэнтерс?	308 очков	['308']	1
Сколько мешков за карьеру было у Джареда Аллена?	136	['136']	1
Сколько блокировок записал на свой счет Люк Кикли?	118	['118']	1
Сколько мячей перехватил Джош Норман?	4	['4', 'четыре']	1
Кто больше записал на свой счет мешков в команде в этом сезоне?	Кейван Шорт	['Кейван Шорт']	1
Сколько перехватов приписывают защите Пэнтерс в 2015 году?	24	['24']	1
Кто был лидером Пэнтерс по мешкам?	Кейван Шорт	['Кейван Шорт']	1
Сколько игроков защиты Пэнтерс было выбрано для Пробоула?	Два	['4', 'четыре']	0
Сколько вынужденных потерь мяча имел Томас Дэвис?	2	['4', 'четыре']	0
У какого игрока было больше всего перехватов в течение сезона?	Джош Норман	['Курт Колеман']	0

Таблица 3: Подробные результаты для Qwen-3B

Вопрос	Ответ модели	Правильный ответ	Корректность
Сколько очков уступила защита Пэнтерс?	Защита Пэнтерс уступила 308 очков.	['308']	1
Сколько мешков за карьеру было у Джареда Аллена?	Джаред Аллен пять раз становился участником Пробоула и за карьеру набрал 136 мешков.	['136']	1
Сколько блокировок записал на свой счет Люк Кикли?	Люк Кикли лидировал в команде по блокировкам и записал 118 блокировок.	['118']	1
Сколько мячей перехватил Джош Норман?	Шесть.	['4', 'четыре']	0
Кто больше записал на свой счет мешков в команде в этом сезоне?	Кейван Шорт лидирует в команде по количеству мешков.	['Кейван Шорт']	1
Сколько перехватов приписывают защите Пэнтерс в 2015 году?	Защита Пэнтерс приписывает 24 перехвата.	['24']	1
Кто был лидером Пэнтерс по мешкам?	Томас Дэвис был лидером Пэнтерс по мешкам.	['Кейван Шорт']	0
Сколько игроков защиты Пэнтерс было выбрано для Пробоула?	2	['4', 'четыре']	0
Сколько вынужденных потерь мяча имел Томас Дэвис?	Томас Дэвис имел 4 вынужденных потери мяча.	['4', 'четыре']	1
У какого игрока было больше всего перехватов в течение сезона?	Курт Колемана	['Курт Колеман']	1

Таблица 4: Подробные результаты для RuadaptQwen-7B-Instruct

Вопрос	Ответ модели	Правильный ответ	Корректность
Сколько очков уступила защита Пэнтерс?	308 очков.	['308']	1
Сколько мешков за карьеру было у Джареда Аллена?	136 мешков.	['136']	1
Сколько блокировок записал на свой счет Люк Кикли?	118	['118']	1
Сколько мячей перехватил Джош Норман?	6 (четыре перехвата, включая два тачдауна)	['4', 'четыре']	0.5
Кто больше записал на свой счет мешков в команде в этом сезоне?	Марио Эдисон - 61/2 мешка.	['Кейван Шорт']	0.5
Сколько перехватов приписывают защите Пэнтерс в 2015 году?	24 перехватов.	['24']	1
Кто был лидером Пэнтерс по мешкам?	Кейван Шорт - лидировал в команде с 11 мешками.	['Кейван Шорт']	1
Сколько игроков защиты Пэнтерс было выбрано для Пробоула?	5	['4', 'четыре']	0
Сколько вынужденных потерь мяча имел Томас Дэвис?	4	['4', 'четыре']	1
У какого игрока было больше всего перехватов в течение сезона?	Джош Норман - 88 перехватов.	['Курт Колеман']	0

Пояснение: Здесь за 4 и 5 вопросы корректность отмечена как 0.5.

- **Вопрос 4:** Модель дала некорректный ответ, но корректное пояснение. Мы можем сделать вывод, что модель не понимает значение слов "включительно" , "включая"
- **Вопрос 5:** Ответ можно считать правильным исходя из представления данных в датасете. Так как 6,5 мешка представлено в виде  $\frac{61}{2}$ , что явно больше 11

### 3 Выводы

- Все три модели показали одинаковую точность (0.7) на тестовом наборе из 10 вопросов
- OpenChat дает наиболее лаконичные ответы
- Qwen и RuadaptQwen склонны к более развернутым ответам
- Для всех моделей характерны ошибки в вопросах о количестве участников Пробоула. Возможно это связано с нерелевантной информацией в чанках, так как при изучении датасета, у меня возникли затруднения с ответом на этот вопрос.
- Для наибольшей точности базы данных были отобраны неповторяющиеся чанки с длиной текста не менее 20 символов.

### Приложение: Технические детали

Конфигурация системы:

- Векторная БД: Qdrant in-memory
- Модель эмбедингов: intfloat/multilingual-e5-large
- Параметры генерации: temperature=0.2, max\_new\_tokens=128
- Аппаратное обеспечение: GPU T4