

Разработка интерактивного дашборда для сравнительного анализа больших языковых моделей (LLM) в среде Power BI

Анишкова Анастасия Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью исследования является разработка интерактивного аналитического дашборда в среде Power BI для визуализации, сравнительного анализа и оценки современных больших языковых моделей (LLM) по показателям производительности, качества, стоимости и функциональных возможностей. Выбор платформы Power BI обусловлен её широкими возможностями в области бизнес-аналитики, интерактивной визуализации данных и поддержки современных инструментов анализа больших массивов информации. Платформа предоставляет удобные средства интеграции с различными источниками данных, позволяет выполнять их очистку и преобразование, а также создавать динамические дашборды с механизмами фильтрации и детализации. Применение такого подхода позволяет обеспечить наглядное представление результатов анализа, автоматизировать процесс сравнительной оценки больших языковых моделей и повысить эффективность принятия решений при выборе наиболее подходящих решений для различных задач. Использование интерактивных фильтров и связанных визуальных элементов предоставляет возможность оперативно исследовать данные по различным критериям, выявлять закономерности и выполнять многомерный анализ характеристик моделей в режиме реального времени.

Ключевые слова: дашборд, Power BI, LLM, анализ, разработка.

Developing an interactive dashboard for comparative analysis of large language models (LLM) in the Power BI environment

Anishkova Anastasia Sergeevna

Sholom-Aleichem Priamursky State University

Student

Abstract

The purpose of the study is to develop an interactive analytical dashboard in the Power BI environment for visualization, comparative analysis, and evaluation of modern large language models (LLM) based on performance, quality, cost, and functionality. The choice of the Power BI platform is based on its extensive capabilities in business analytics, interactive data visualization, and support for modern tools for analyzing large amounts of information. The platform provides convenient tools for integrating with various data sources, cleaning and transforming

data, and creating dynamic dashboards with filtering and drill-down mechanisms. This approach allows for a visual representation of the analysis results, automates the process of comparing large language models, and improves decision-making when selecting the most suitable solutions for various tasks. The use of interactive filters and related

Key words: dashboard, Power BI, LLM, analysis, development.

1 Введение

1.1 Актуальность

Актуальность исследования обусловлена стремительным развитием технологий искусственного интеллекта и постоянным увеличением числа больших языковых моделей (Large Language Models, LLM), различающихся по качеству генерации, функциональным возможностям, стоимости использования и техническим характеристикам. В условиях высокой конкуренции между разработчиками ИИ возникает необходимость в создании инструментов, обеспечивающих наглядное сравнение и комплексный анализ моделей по совокупности критериев. Использование интерактивных аналитических дашбордов позволяет упростить процесс оценки доступных решений, повысить эффективность анализа данных и обеспечить поддержку принятия обоснованных решений при выборе языковых моделей для практического применения.

1.2 Обзор исследований

С. Виджая, Т. Маурициус продемонстрировали разработку визуализации панели мониторинга производительности с использованием power BI в качестве платформы [1]. Управление производительностью цепочки поставок с использованием информационной панели Microsoft Power BI в режиме реального времени методом action design research (ADR) создал Д. Х. Набиль [2]. В. С. Хатувал, Д. Пури описал инструменты бизнес - аналитики для разработки информационных панелей [3]. Исследование инструментов бизнес-аналитики для разработки корпоративных информационных панелей продемонстрировал К. Гоутами, М. Р. П. Кумар [4]. Д. Шивананда предположил панель мониторинга настроек с использованием данных Twitter в power BI [5].

1.3 Цель исследования

Целью исследования является разработка интерактивного аналитического дашборда в среде Power BI для визуализации, сравнительного анализа и оценки современных больших языковых моделей (LLM) по показателям производительности, качества, стоимости и функциональных возможностей.

2 Материалы и методы

В качестве инструментария исследования использовалась платформа Microsoft Power BI, предназначенная для анализа, обработки и визуализации

данных. В качестве источника данных применялся набор характеристик больших языковых моделей, содержащий информацию о производительности, стоимости использования, функциональных возможностях и результатах бенчмарков. Для расчета аналитических показателей использовался язык DAX (Data Analysis Expressions), а для подготовки и преобразования данных — встроенные средства Power Query.

3 Результаты

Ссылка на дашборд <https://disk.yandex.ru/d/BXIcKLrS83z-Mw>.

Ссылка на датасет https://disk.yandex.ru/d/H0_rndu_WH04Bw.

Процесс разработки дашборда в среде Power BI начинается с импорта исходных данных. Для этого в верхнем меню программы необходимо перейти по пути «Главная» → «Получить данные» → «Текст/CSV» и выбрать целевой файл `llm_benchmark_leaderboard_2024_2026.csv`. После визуальной проверки корректности отображения структуры данных в окне предварительного просмотра активируется загрузка файла в модель с помощью кнопки «Загрузить». По завершении импорта первая страница отчета переименовывается в «Обзор», где формируется визуальная иерархия будущего интерфейса. Через вкладку верхнего меню «Вставка» на рабочую область добавляется текстовое поле с заголовком «LLM Benchmark Dashboard 2026».

Следующим этапом является настройка ключевых показателей эффективности (KPI), базовым из которых выступает общий объем анализируемых моделей. Поскольку для корректного вычисления данного параметра требуется создание вычисляемой меры, в меню «Моделирование» выбирается функция «Создать меру». В появившейся строке формул прописывается выражение DAX:

Количество моделей =
`COUNTROWS('llm_benchmark_leaderboard_2024_2026')`,

которое инициализирует автоматический подсчет строк и выводит новую переменную в панель «Данные». Финальная визуализация показателя осуществляется путем добавления элемента «Карточка» из панели «Визуализации» на рабочую область. Перетаскивание созданной меры в поле значений карточки позволяет отобразить итоговое количество исследуемых моделей (рис.1).

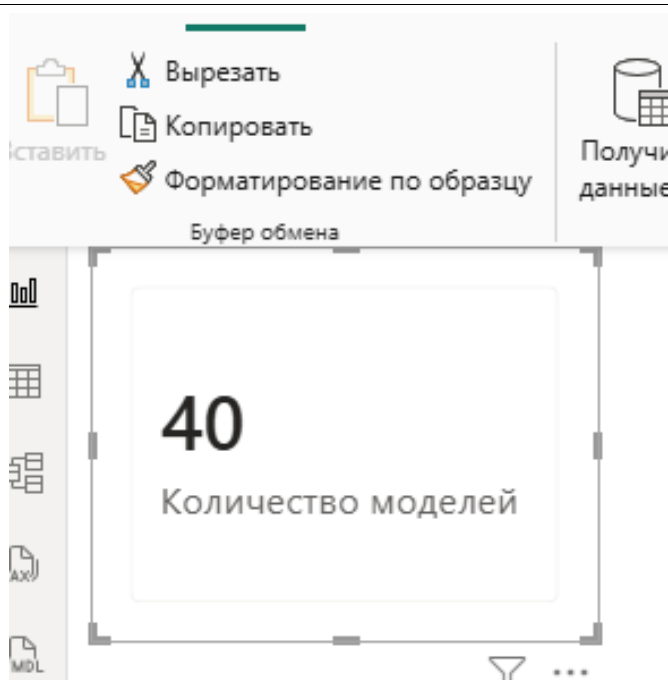


Рисунок 1. Карточка «Количество моделей»

Продолжая настройку показателей, аналогичным образом создается карточка «Количество разработчиков». Для этого во вкладке «Моделирование» выбирается функция «Создать меру», в строке формул прописывается выражение:

Количество разработчиков =
`DISTINCTCOUNT('llm_benchmark_leaderboard_2024_2026'[provider]),`

Эта формула подсчитывает точное число уникальных создателей моделей в таблице. После этого на рабочую область добавляется новая «Карточка», на которую перетаскивается созданная мера, и на экране отображается итоговое количество разработчиков (рис.2).

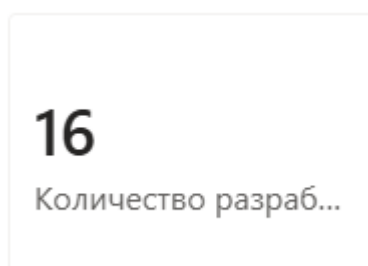


Рисунок 2. Карточка «Количество разработчиков»

Третья карточка предназначена для вывода максимального рейтинга Arena Elo, который отражает наивысший уровень качества и эффективности ответов среди всех представленных в таблице языковых моделей. Данный показатель базируется на системе шахматных коэффициентов Эло и формируется в ходе «слепых» сравнительных тестов реальными

пользователями, что делает его ключевым независимым критерием оценки силы ИИ. Для реализации этого индикатора во вкладке «Моделирование» создается мера со следующим выражением:

Лучший Arena Elo =
`MAX('llm_benchmark_leaderboard_2024_2026'[arena_elo_score])`.

Эта формула автоматически находит максимальный балл в соответствующем столбце. Затем на рабочую область добавляется стандартный элемент «Карточка», на который перетаскивается созданная мера, визуализируя рекордный рейтинг Арены на дашборде (рис.3).

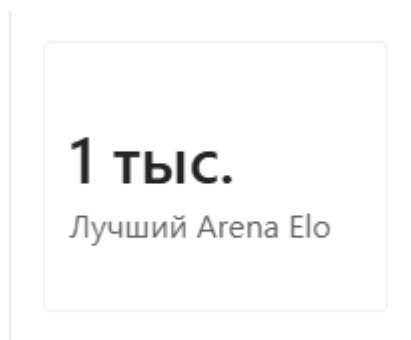


Рисунок 3. Карточка «Лучший Arena Elo»

Четвертая карточка отображает средний показатель MMLU (Massive Multitask Language Understanding). Этот бенчмарк является одним из самых известных академических тестов для искусственного интеллекта. Он оценивает знания моделей и их способность решать задачи в 57 различных областях, включая гуманитарные науки, математику, историю, юриспруденцию и компьютерные дисциплины. Таким образом, средний балл MMLU на дашборде позволяет оценить общую эрудицию и уровень фундаментальных знаний у всей группы анализируемых моделей. Для настройки этого индикатора во вкладке «Моделирование» создается мера со следующим выражением:

Средний MMLU =
`AVERAGE('llm_benchmark_leaderboard_2024_2026'[mmlu_score])`

Эта формула рассчитывает среднее арифметическое значение баллов из соответствующего столбца. После этого, по аналогии с предыдущими шагами, на рабочую область добавляется новая «Карточка», куда перетаскивается созданная мера для вывода итогового среднего результата (рис.4).

83

Средний MMLU

Рисунок 4. Карточка «Средний MMLU»

Пятая карточка отображает количество бесплатных моделей. В отличие от предыдущих индикаторов, здесь используется фильтрация данных, чтобы посчитать только те нейросети, доступ к которым предоставляется бесплатно. Это позволяет быстро оценить объем доступных и бюджетных решений в общем списке. Для настройки этого показателя во вкладке «Моделирование» создается мера со следующим выражением:

Количество бесплатных моделей =
`CALCULATE( COUNTROWS('llm_benchmark_leaderboard_2024_2026'), 'llm_benchmark_leaderboard_2024_2026'[is_free] = TRUE())`.

Функция `CALCULATE` в данном коде изменяет контекст фильтрации, а условие `[is_free] = TRUE()` отсекает платные модели, оставляя для подсчета строк только бесплатные. На заключительном этапе на рабочую область добавляется стандартный элемент «Карточка», на который перетаскивается созданная мера, завершая формирование блока ключевых KPI-показателей (рис.5).

13

Количество бесплат...

Рисунок 5. Карточка «Количество бесплатных моделей»

Шестая карточка отображает среднюю стоимость API для анализируемых моделей. API (Application Programming Interface) в контексте больших языковых моделей — это интерфейс, который позволяет разработчикам интегрировать нейросеть в свои приложения, сайты или сервисы. Стоимость использования такого API обычно рассчитывается за объем переданных данных — в данном случае показатель `cost_per_1M_combined` отражает комбинированную (суммарную) стоимость обработки 1 миллиона токенов (примерно 750 тысяч слов) как на вход (запрос пользователя), так и на выход (ответ модели). Этот индикатор крайне важен для оценки экономической эффективности коммерческого внедрения ИИ.

Для создания этого показателя во вкладке «Моделирование» создается мера со следующим выражением:

Средняя стоимость API =
`AVERAGE('llm_benchmark_leaderboard_2024_2026(1)'[cost_per_1M_combined])`

Данная формула вычисляет среднее арифметическое значение стоимости среди всех платных запросов к моделям. На заключительном этапе на рабочую область добавляется новая «Карточка», на которую перетаскивается созданная мера, что позволяет наглядно представить средний уровень затрат на интеграцию технологий ИИ (рис.6).

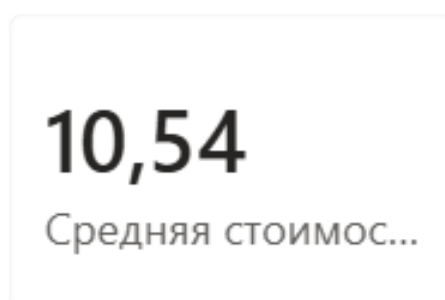


Рисунок 6. Карточка «Средняя стоимость API»

Следующим важным шагом в создании дашборда является добавление элементов интерактивной фильтрации — срезов (Slicers). Срезы позволяют пользователю мгновенно отсеивать лишние данные и анализировать только те модели, которые соответствуют конкретным требованиям.

Для создания первого среза на панели «Визуализации» необходимо нажать на значок «Срез» и расположить его на рабочей области. Затем из панели «Данные» нужно перетащить поле `provider` (разработчик модели) в область значений этого среза.

По аналогии создаются еще 6 срезов (рис.7) для глубокого анализа базы данных. Каждое из этих полей отвечает за ключевые технические и экономические характеристики моделей:

1. `price_tier` (Ценовой уровень): Разделяет модели по категориям стоимости (например, бюджетные, средние, премиальные). Это помогает быстро отсеять слишком дорогие коммерческие варианты.

2. `is_free` (Доступность): Логический фильтр со значениями «Да/Нет» (True/False), который показывает, можно ли использовать модель бесплатно или за нее в любом случае придется платить.

3. `modality` (Модальность): Определяет базовый тип данных, с которыми работает модель (например, исключительно текст или текст и код).

4. `supports_vision` (Поддержка компьютерного зрения): Фильтр, показывающий, способна ли модель распознавать, анализировать и описывать загруженные изображения (мультимодальность по зрению).

5. `supports_audio` (Поддержка аудио): Указывает на способность модели напрямую воспринимать на слух звуковые файлы, голосовые команды или генерировать аудиоответы.

6. `supports_tool_use` (Поддержка использования инструментов / Function Calling): Критически важная функция для разработчиков. Она показывает, умеет ли нейросеть выходить за рамки простого чата — например, самостоятельно вызывать сторонние API, подключаться к базам данных или запускать код для решения задач.

The image shows a web interface for filtering AI models. It consists of several sections, each with a title, a list of options, and a vertical slider control.

- provider**: Options include 01.AI, AI21 Labs, Alibaba Cloud, and Amazon. The slider is positioned at the bottom.
- modality**: Options include Text, Text + Vision, Text + Vision + Audio, and Text + Vision + Audio + Video. The slider is positioned in the middle.
- supports_tool_use**: Options include False and True. The slider is positioned at the top.
- price_tier**: Options include Budget, Free/Open, Mid-Range, and Premium. The slider is positioned in the middle.
- supports_audio**: Options include False and True. The slider is positioned in the middle.
- is_free**: Options include False and True. The slider is positioned at the bottom.

Рисунок 7. Срезы

Для проведения глубокого сравнительного анализа исследуемых моделей на дашборд добавляется точечная диаграмма (рис.8). Этот элемент визуализации позволяет наглядно выявить зависимости между техническими характеристиками нейросетей, их эффективностью и стоимостью. Процесс построения графика начинается с выбора значка «Точечная диаграмма» на панели «Визуализации» и её размещения на рабочей области. Далее переменные из панели «Данные» распределяются по ключевым зонам настройки элемента. В качестве оси X устанавливается поле `parameter_count_B` (количество параметров в миллиардах), а оси Y — `arena_elo_score` (пользовательский рейтинг силы ИИ). Для цветового разделения данных по разработчикам в область «Условные обозначения» переносится переменная `provider`. Масштаб отображаемых точек задается через поле «Размер» с помощью переменной `context_window_tokens` (объем контекстного окна). Чтобы настроить вывод информативных всплывающих окон при наведении курсора, в область «Подсказки» добавляются метрики `model_name`, `mmlu_score`, `humaneval_score` и `cost_per_1M_combined`.

Важным техническим нюансом является предотвращение автоматического агрегирования данных со стороны Power BI. По умолчанию система может попытаться рассчитать сумму показателей, искажая логику графика до значений вида «Сумма по parameter_count_B». Чтобы скорректировать отображение, необходимо нажать на стрелку рядом с полями parameter_count_B, arena_elo_score и context_window_tokens непосредственно в зонах настройки осей и активировать опцию «Не суммировать» (Don't summarize). Окончательное распределение моделей в виде дискретных точек на плоскости происходит после переноса переменной model_name в область «Детализация». В заключение, для повышения читаемости интерфейса технические наименования осей и легенд адаптируются под конечного пользователя: через клик правой кнопкой мыши по соответствующему полю выбирается команда «Переименовать для этого визуального элемента», после чего вводятся русскоязычные названия.

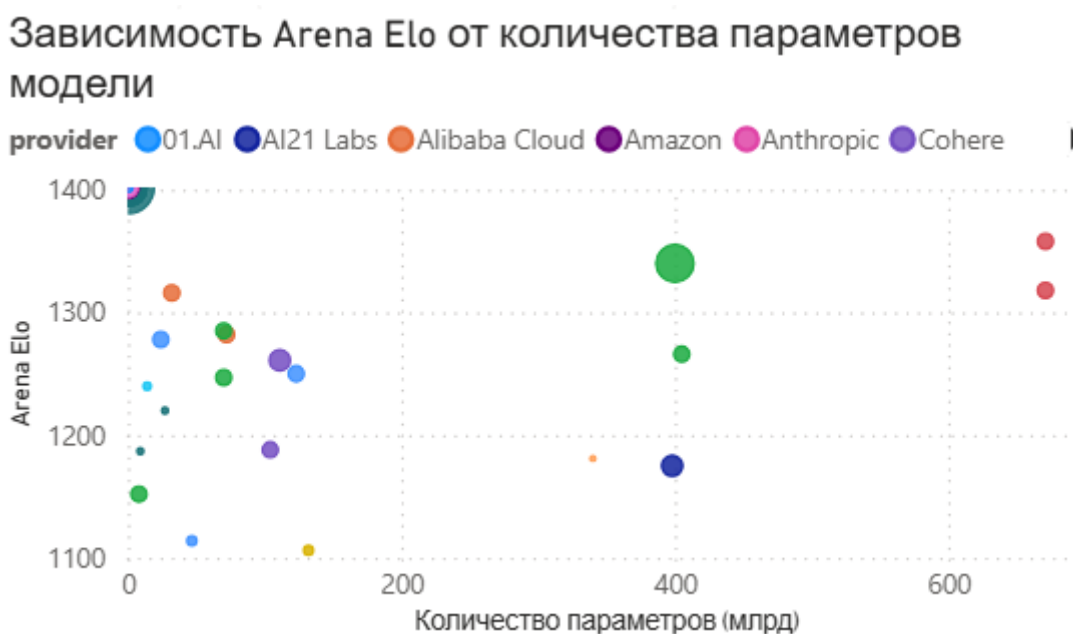


Рисунок 8. Точечная диаграмма

Заключительным элементом рабочей области выступает структурированная таблица (рис.9), предназначенная для детального изучения и прямой сверки числовых и текстовых характеристик моделей. Процесс её создания начинается с выбора значка «Таблица» на панели «Визуализации» и её размещения на свободном пространстве дашборда. Настройка детализированного представления осуществляется путем последовательного перетаскивания ключевых переменных в область значений элемента. Для формирования полной аналитической картины в таблицу добавляются такие параметры, как model_name (название модели) и provider (разработчик), показатели эффективности в виде arena_rank (место в глобальном рейтинге Арены), arena_elo_score (балл Elo), а также результаты академических тестов mmlu_score и humaneval_score. Экономические и функциональные

характеристики отражаются через переменные `price_tier` (ценовая категория), `cost_per_1M_combined` (комбинированная стоимость миллиона токенов) и `is_free` (индикатор бесплатного доступа). Подобная конфигурация позволяет пользователю не только видеть обобщенные графические зависимости, но и переходить к точечному анализу конкретных технологических решений.

model_name	provider	Сумма arena_rank	Сумма arena_elo_score	Сумма mmlu_score	Сумма humaneval_score	price_tier	Сумма cost_per_1M_combined
Mixtral 8x7B Instruct	Mistral AI	39	1114	70,60	40,20	Free/Open	0,0
Gemma 2 9B	Google DeepMind	33	1187	71,30	54,30	Free/Open	0,0
Gemma 2 27B	Google DeepMind	30	1220	75,20	63,40	Free/Open	0,0
Reka Core	Reka AI	35	1178	73,70	65,90	Mid-Range	18,0
Всего		820	50471	3 320,80	3 275,60		421,4

Рисунок 9. Таблица

Затем в дашборд добавляется вторая страница, которая переименовывается в «Представление». На ней размещаются три графика для быстрого визуального сравнения моделей. В качестве основного элемента на панели «Визуализации» выбирается «Гистограмма с группировкой». Настройка первого графика выполняется путем переноса переменной `model_name` в поле оси Y и показателя `arena_elo_score` в поле оси X.

Для экономии времени и сохранения единого оформления два оставшихся графика создаются обычным копированием, после чего в получившейся копии поле оси X заменяется с `arena_elo_score` на `mmlu_score` для оценки эрудиции моделей. Затем этот новый график снова копируется в третьей гистограмме значение оси X заменяется с `mmlu_score` на `humaneval_score` для оценки навыков программирования. В результате на странице формируются три удобных рейтинга, которые позволяют наглядно сравнить возможности моделей по трем разным независимым направлениям (рис.10).

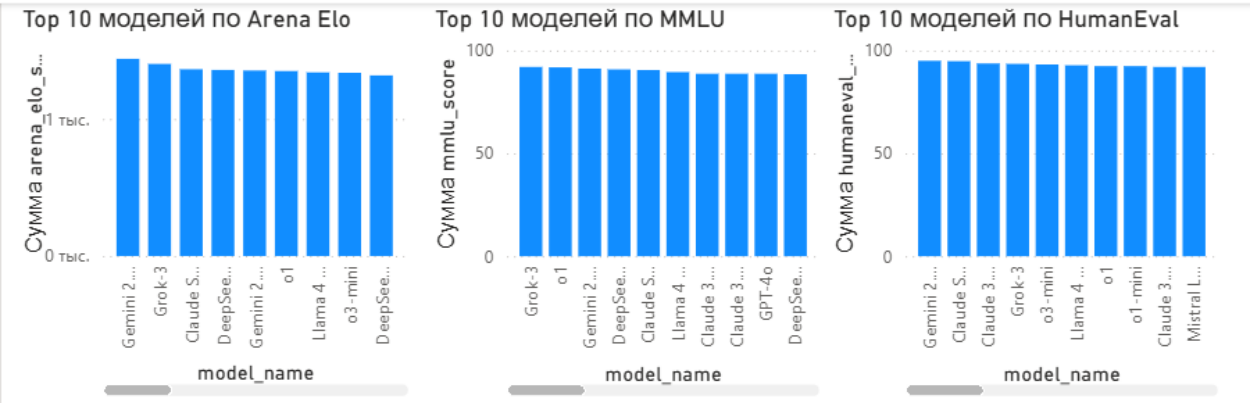


Рисунок 10. Графики

На следующем этапе в дашборд добавляется новая страница под названием «Анализ затрат», предназначенная для оценки экономической эффективности моделей. На этой странице размещается точечная диаграмма, которая позволяет сопоставить финансовые затраты на ИИ с его фактической производительностью.

Настройка визуального элемента выполняется путем распределения переменных по соответствующим осям и областям. В качестве оси X устанавливается поле `cost_per_1M_combined`, отображающее комбинированную стоимость миллиона токенов, а по оси Y задается рейтинг `arena_elo_score`. Для разделения моделей по ценовым категориям в область «Легенда» переносится переменная `price_tier`. Чтобы детализировать общие данные до уровня конкретных нейросетей, в поле «Детализация» добавляется параметр `model_name`. Масштаб или размер точек на графике задается с помощью переменной `parameter_count_B`, что позволяет визуально оценить объем параметров каждой модели. Как и при настройке предыдущих графиков, для корректного отображения данных необходимо отключить автоматическое агрегирование в меню полей осей, выбрав команду «Не суммировать» (рис.11).

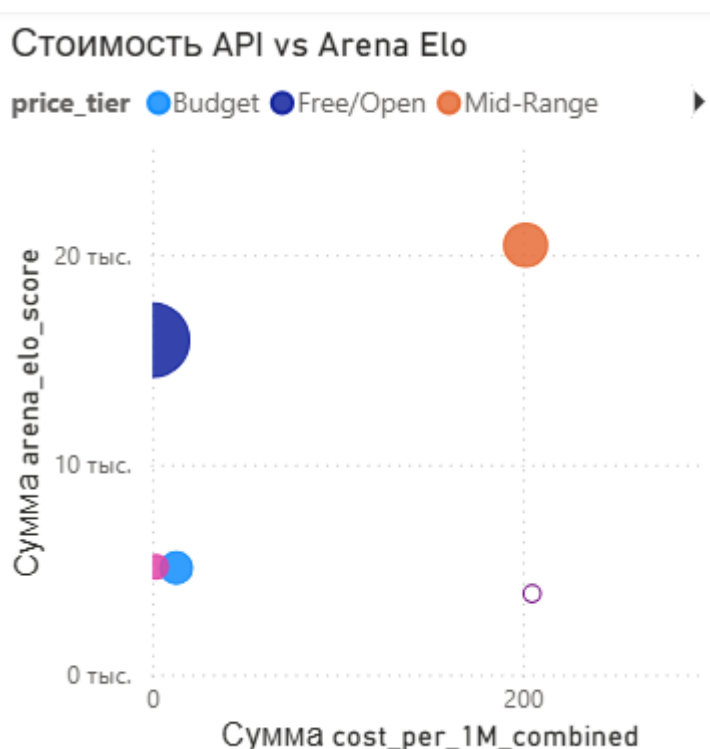


Рисунок 11. Точечная диаграмма

На этой же странице размещается элемент «Кольцевая диаграмма» (Donut Chart), который наглядно показывает распределение моделей по ценовым категориям. При настройке этого визуального элемента в поле «Легенда» (Legend) перетаскивается переменная `price_tier`, которая отвечает за группировку по уровням стоимости. В поле «Значения» (Values) добавляется параметр `model_name`. Чтобы диаграмма отображала именно числовые пропорции, необходимо нажать на стрелку рядом с полем `model_name` в области настройки и выбрать тип агрегации «Количество» (Count). В результате график автоматически подсчитает и отобразит количество

моделей, вошедших в каждый ценовой уровень, позволяя быстро оценить соотношение дорогих, средних и бюджетных решений на рынке (рис.12).

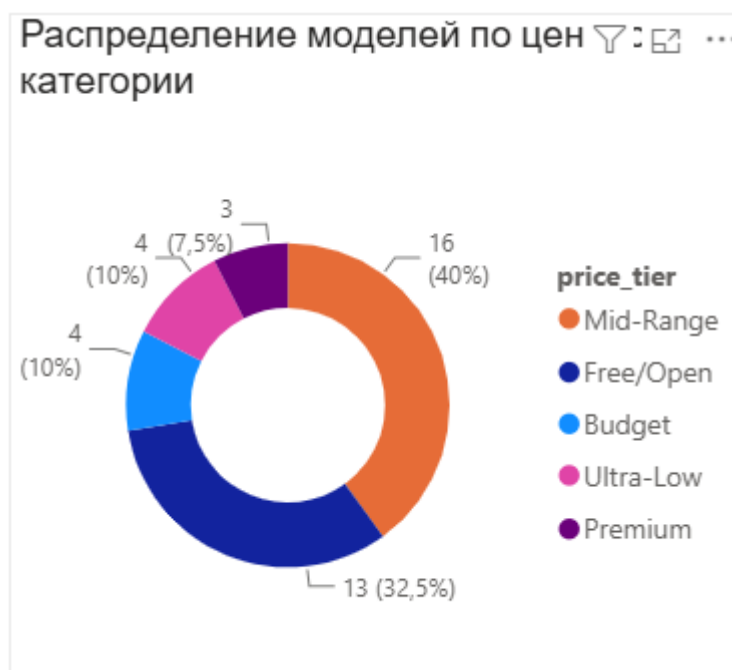


Рисунок 12. Круговая диаграмма

Затем на рабочем поле размещается элемент «Линейчатая диаграмма с накоплением» (Stacked Bar Chart), который позволяет визуализировать финансовую нагрузку в разрезе компаний-создателей ИИ. При настройке этого графика в поле оси Y перетаскивается переменная `provider`, группирующая данные по разработчикам. В поле оси X добавляется параметр `cost_per_1M_combined`, для которого по умолчанию оставляется функция агрегации «Сумма». В результате формируется информативный график «Совокупная стоимость API по разработчикам», отражающий общую сумму стоимости тарифов за миллион комбинированных токенов для всех продуктов каждого конкретного провайдера, что помогает быстро определить авторов самых дорогостоящих линеек моделей (рис.13).

Совокупная стоимость API по разработчикам

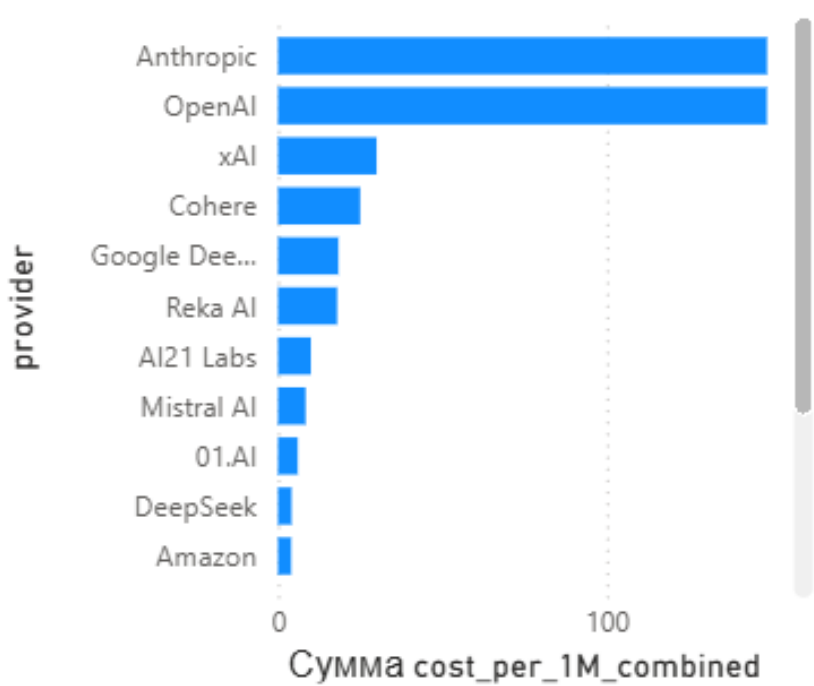


Рисунок 13. Линейчатая диаграмма

Дашборд собран приведен в единый стиль (рис.14,15,16).

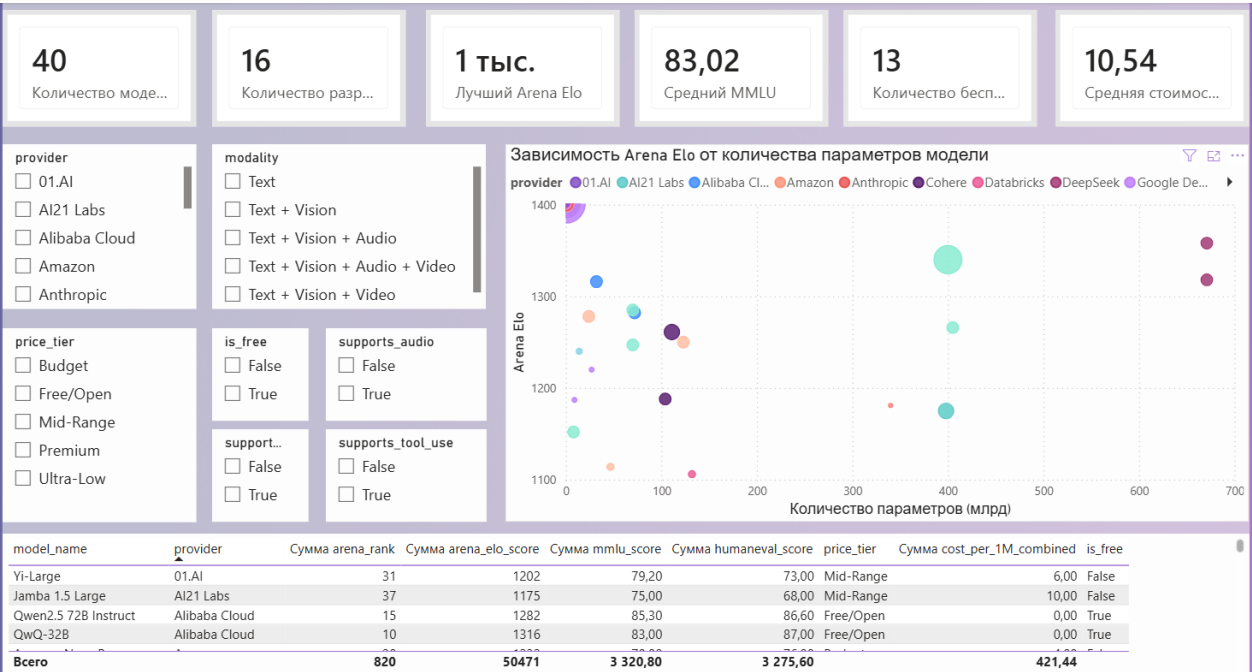


Рисунок 14. Первый лист

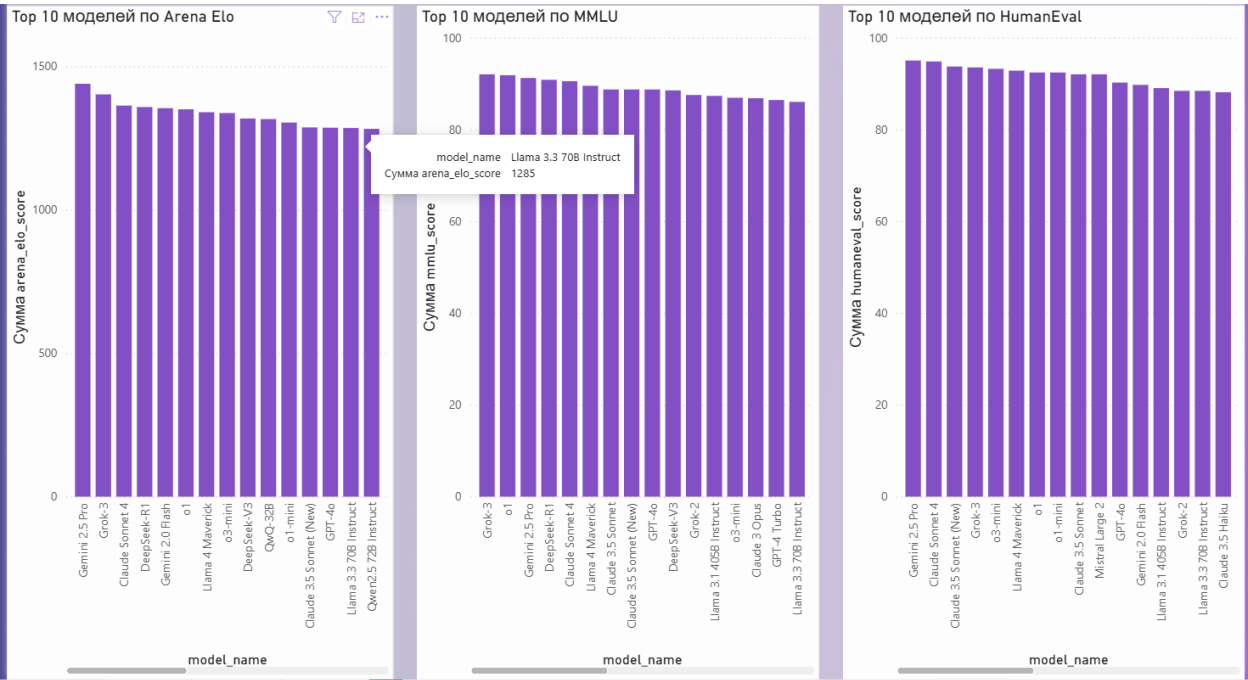


Рисунок 15. Второй лист

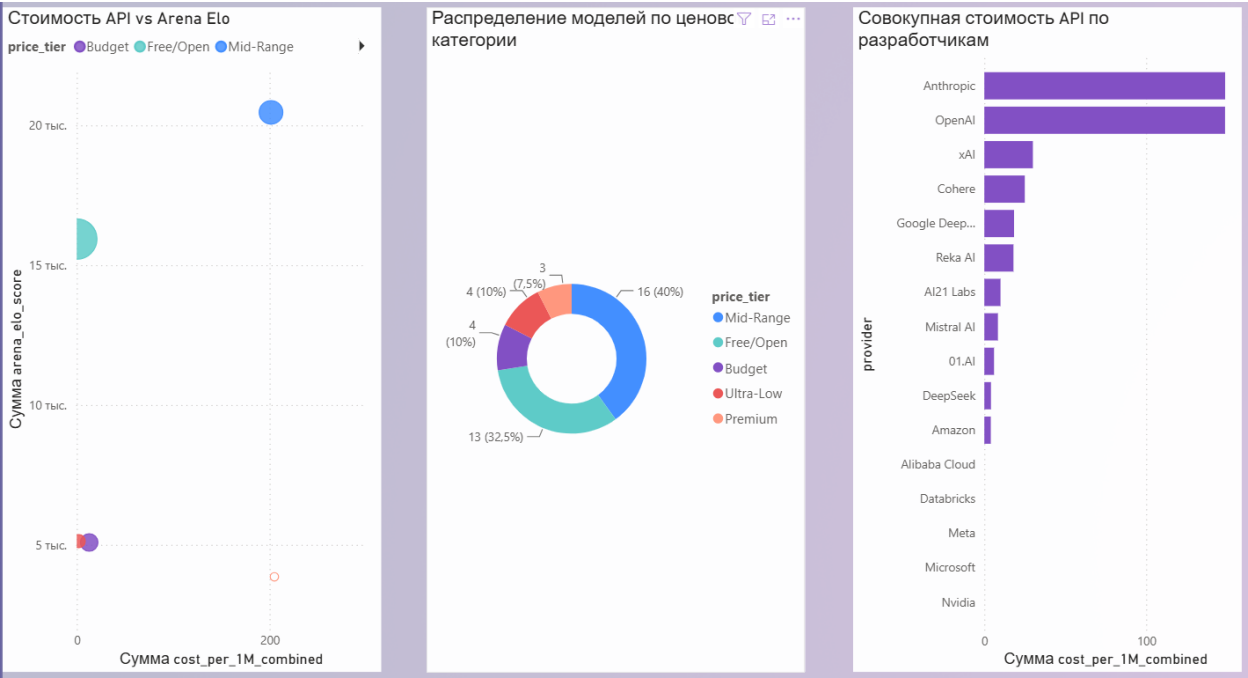


Рисунок 16. Третий лист

Все визуальные элементы связаны между собой. При выборе разработчика, ценовой категории, модальности или других параметров происходит динамическая фильтрация данных на всех графиках и таблицах, что позволяет проводить сравнительный анализ моделей в режиме реального времени (рис.17).

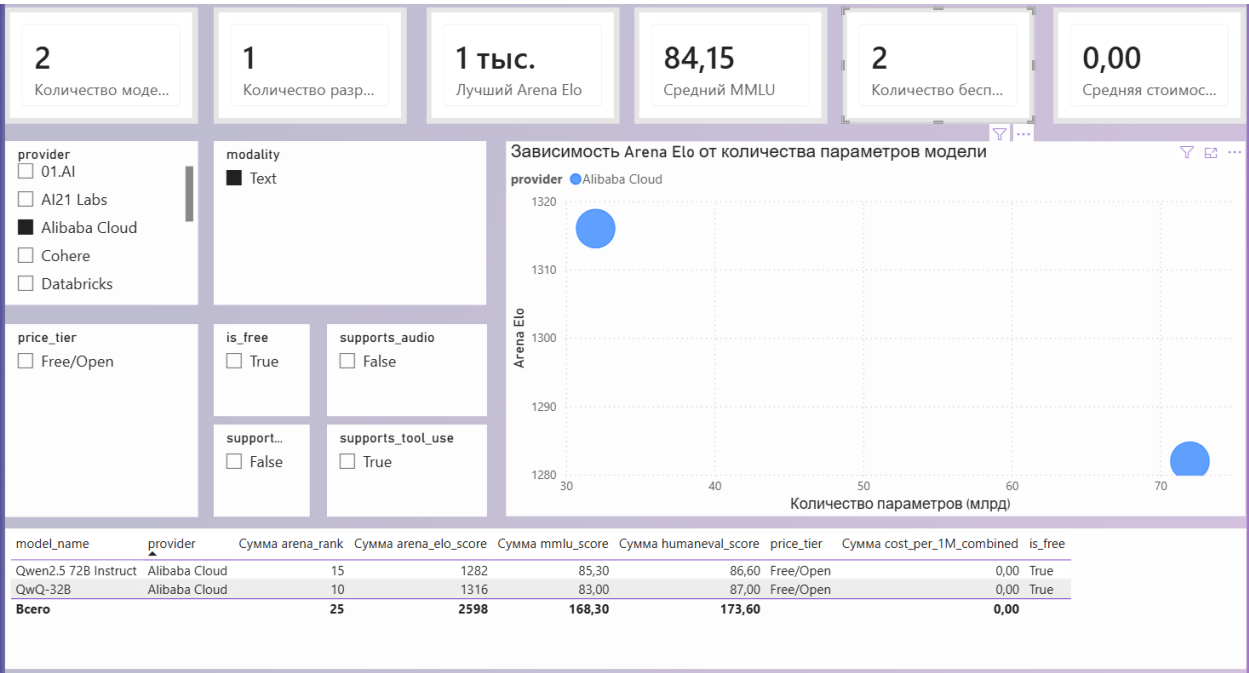


Рисунок 17. Дашборд

4 Выводы

В результате исследования был разработан интерактивный дашборд в среде Power BI для анализа и сравнения больших языковых моделей. Проведенный анализ показал, что лидирующие позиции по ключевым показателям занимают модели Gemini 2.5 Pro, Grok-3, Claude Sonnet 4 и DeepSeek-R1. При этом наиболее высокие показатели качества и программирования продемонстрировала Gemini 2.5 Pro, а лидером по уровню общих знаний стала модель Grok-3. Разработанный дашборд обеспечивает наглядную визуализацию данных, поддерживает интерактивную фильтрацию и позволяет эффективно выполнять многокритериальный анализ современных LLM-моделей по показателям производительности, стоимости и функциональных возможностей.

Библиографический список

1. Widjaja S., Mauritsius T. The development of performance dashboard visualization with power BI as platform //Int. J. Mech. Eng. Technol. 2019. T. 10. №. 5. С. 235-249;
2. Nabil D. H. et al. Managing supply chain performance using a real time Microsoft Power BI dashboard by action design research (ADR) method //Cogent Engineering. 2023. T. 10. №. 2. С. 2257924
3. Khatuwal V. S., Puri D. Business intelligence tools for dashboard development //2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). IEEE, 2022. С. 128-131
4. Khatuwal V. S., Puri D. Business intelligence tools for dashboard development //2022 3rd International Conference on Intelligent Engineering and Management

(ICIEM). IEEE, 2022. C. 128-131

5. Sivananda D. et al. Sentiment analysis dashboard using twitter data in power BI //Manuscript Editor. 2021. T. 2021. C. 61.