



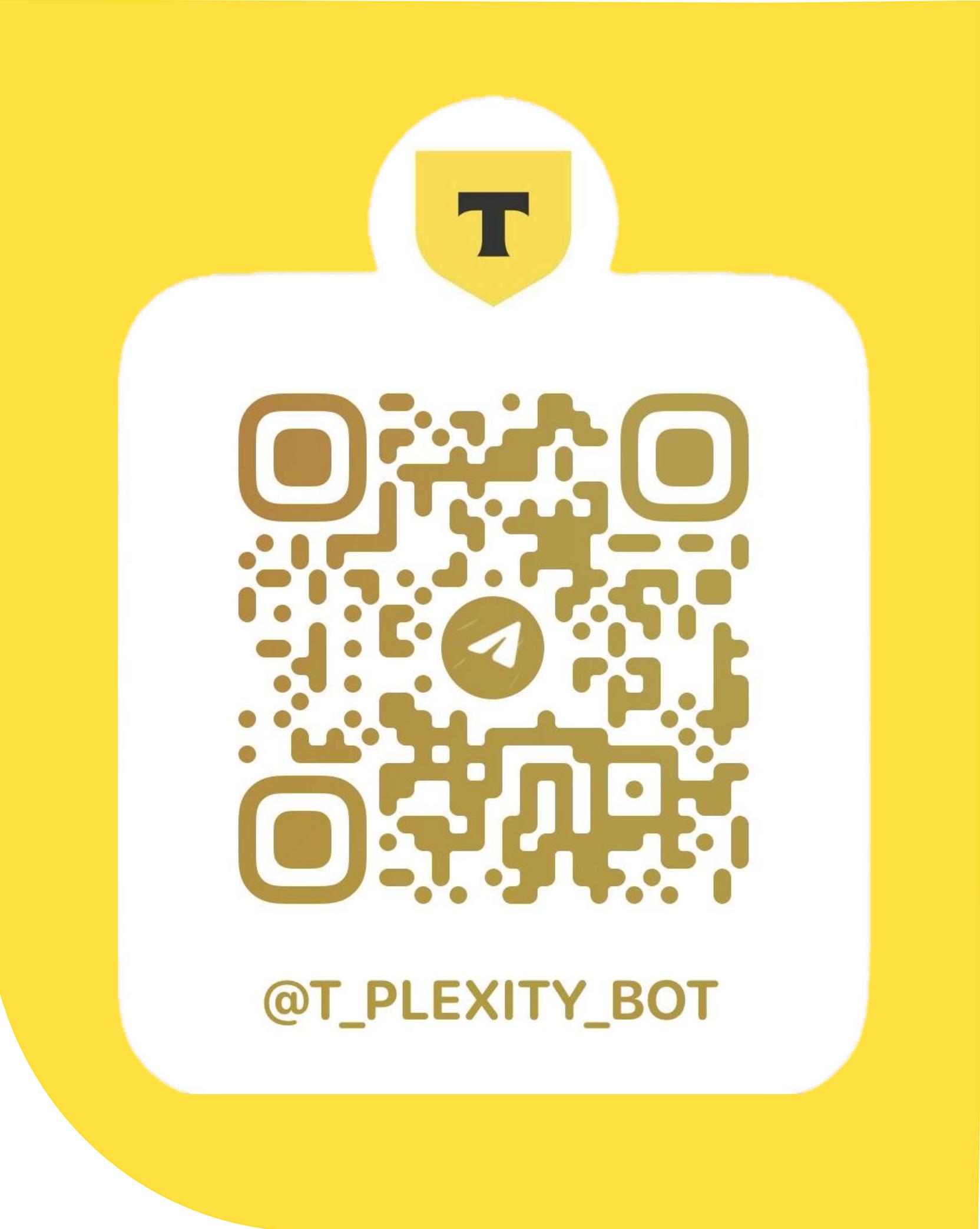
T-Plexity

TBank & Sirius

Идея

Интеллектуальная система, в реальном времени отслеживающая и агрегирует публикации из инвестиционных Telegram-каналов, которая с помощью LLM даёт ответы на пользовательские вопросы о инвестициях.

- **Источники:** инвестиционные Telegram-каналы.
- **Актуальность:** моментальная обработка постов в сервис.
- **Ответы:** формируются LLM, краткие, точные, с рыночным контекстом.
- **Прозрачность:** каждый ответ сопровождается ссылками на первоисточники.



Гипотезы

i

Информационный шум

Пользователи тонут в потоке информации, тратя много времени на фильтрацию шума и недостоверной информации

i

Запаздывание информации

Пользователь вынужден вручную отслеживать множество источников, из-за чего может упустить инвестиционную возможность

i

Проблема доверия и верификация

Пользователь не доверяет информации, если не видит источник этой информации

i

Недостаток персонализации источников

Пользователю важен только определённый список источников, который соответствует их портфелю

Результаты интервью



Недостаток персонализации

6 из 9 респондентов заявили, что им важна любая потенциально полезная информация, он не хочет упустить важную новость из-за того, что не читает определённый источник



Информационная перегрузка

8 из 9 опрошенных заявляют, что сталкиваются с ситуациями, когда пропускают важные новости из-за объема информации



Проблема доверия

7 из 9 респонденты предпочитают не действовать, если не видят источник информации и не доверяют ему

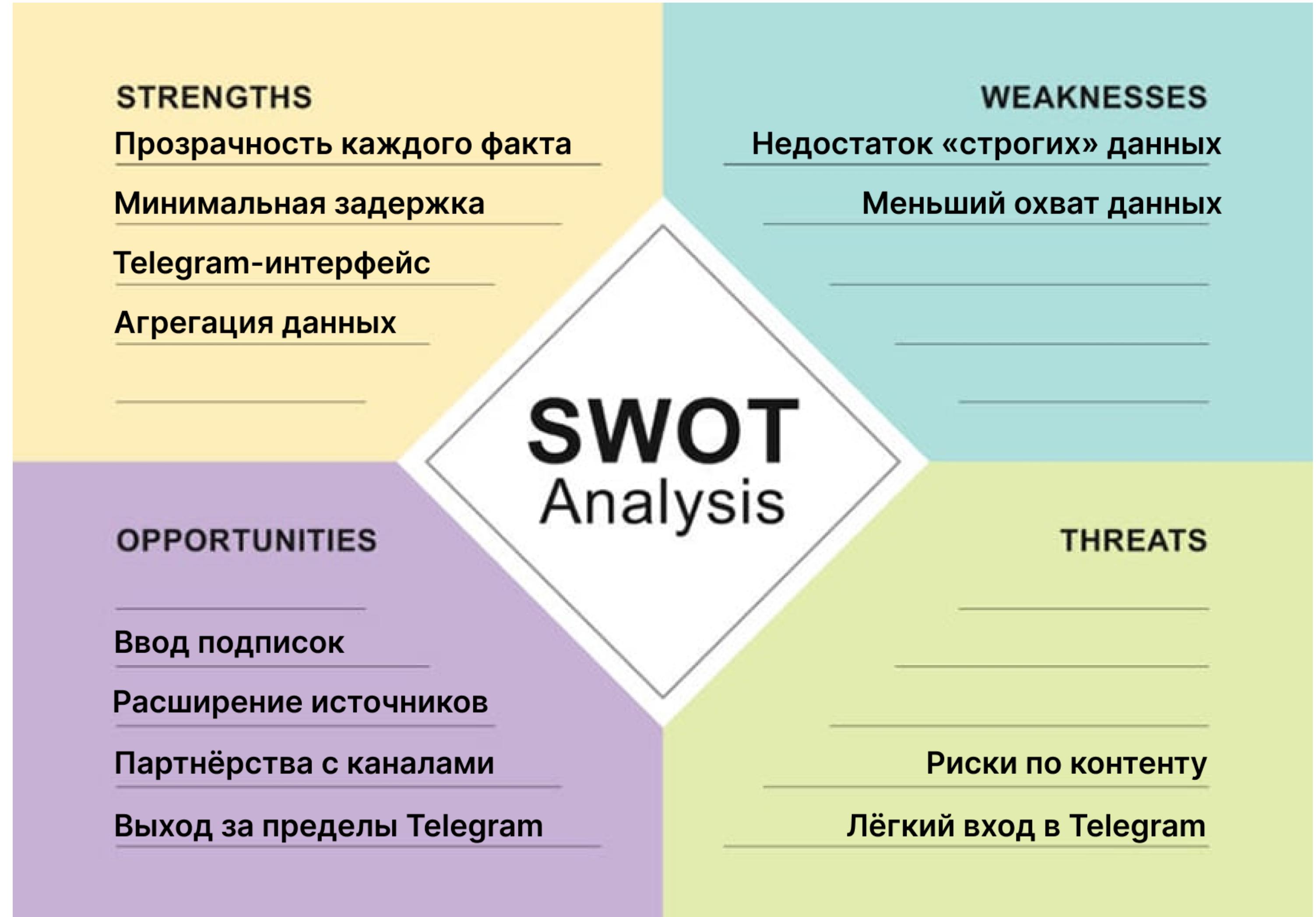


Запаздывание информации

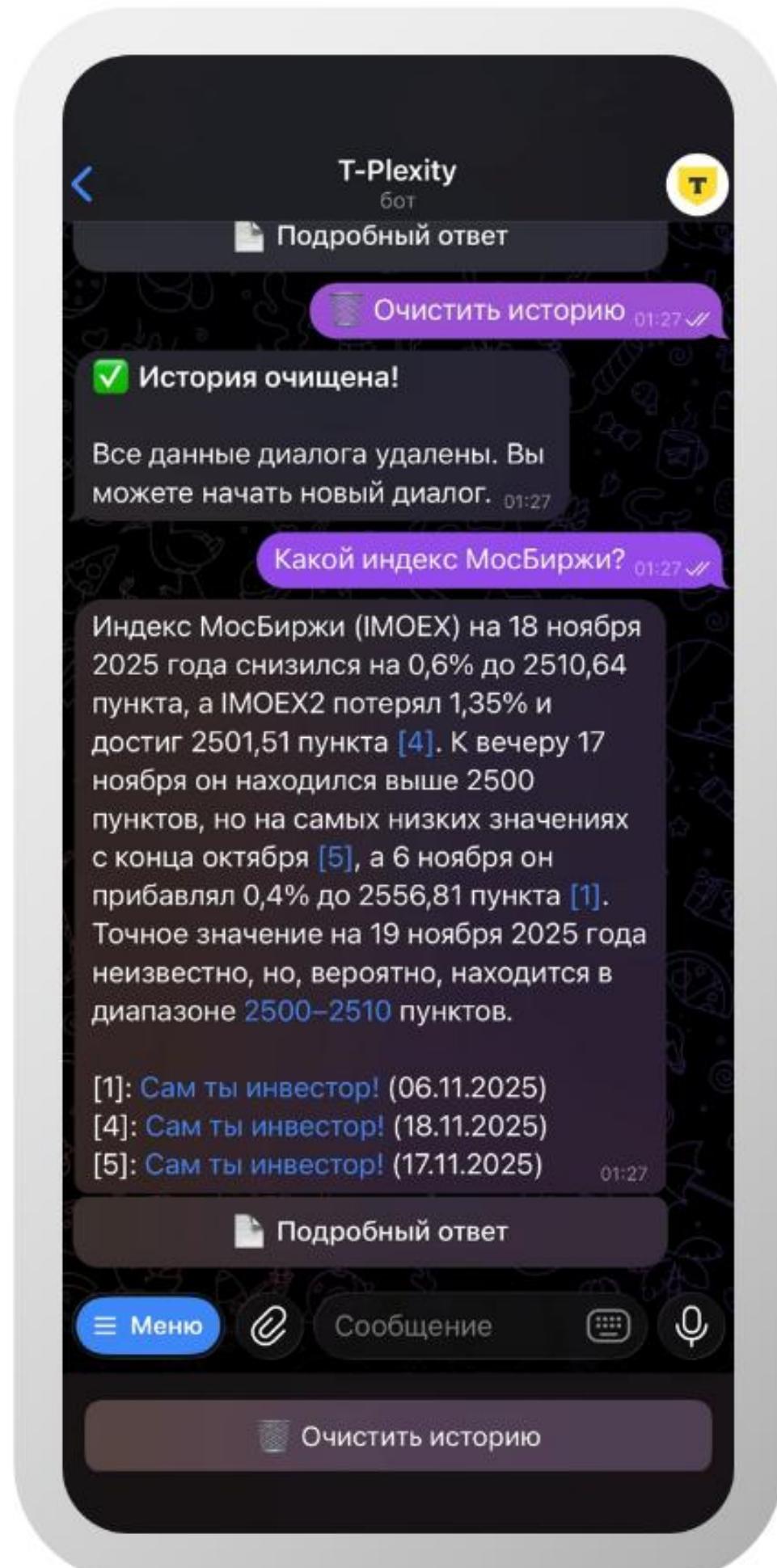
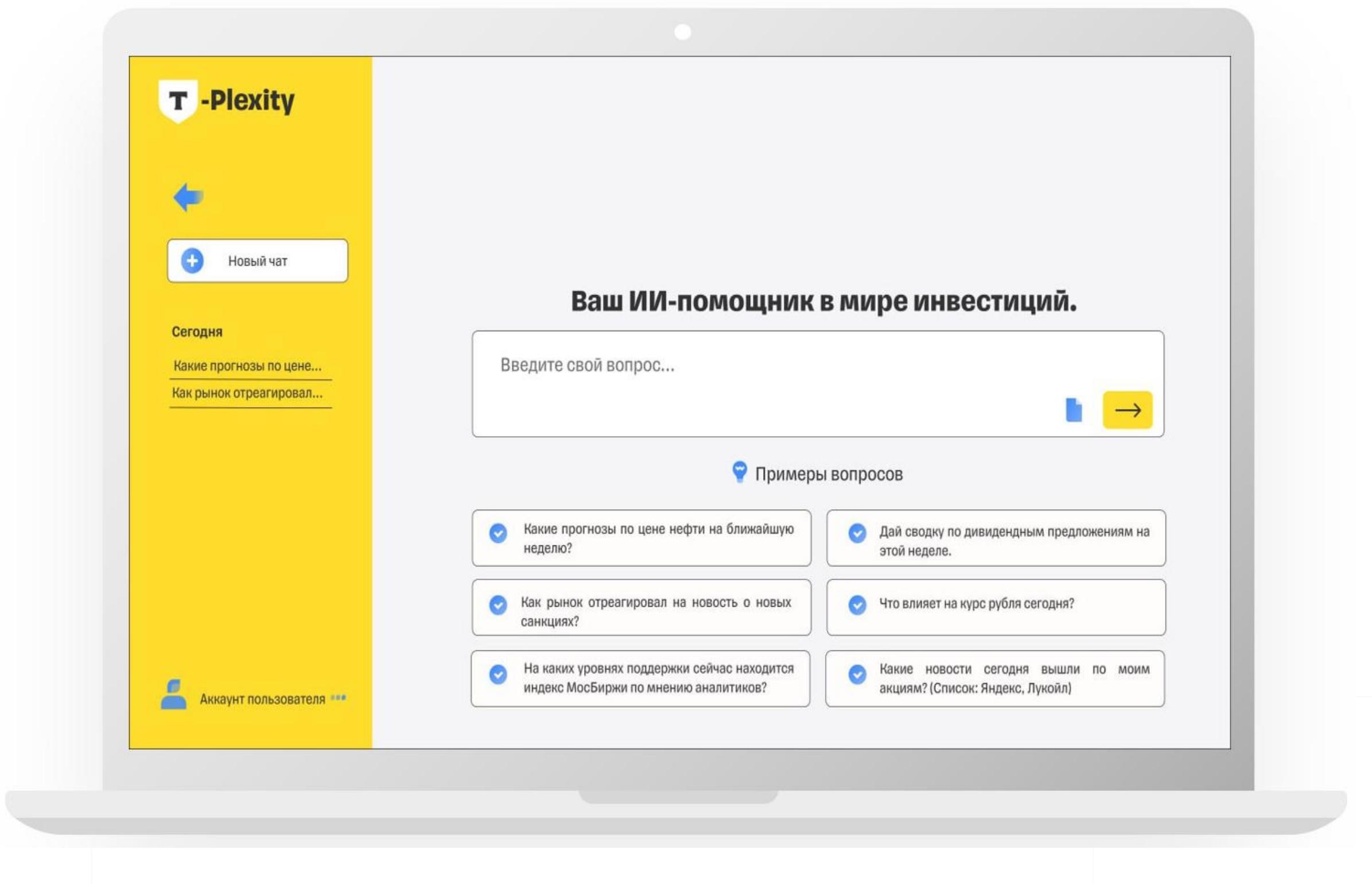
7 из 9 респондентов привели примеры упущеной выгоды
«Задержка в получении и обработке информации - это реальная проблема при моем подходе»

Конкуренты

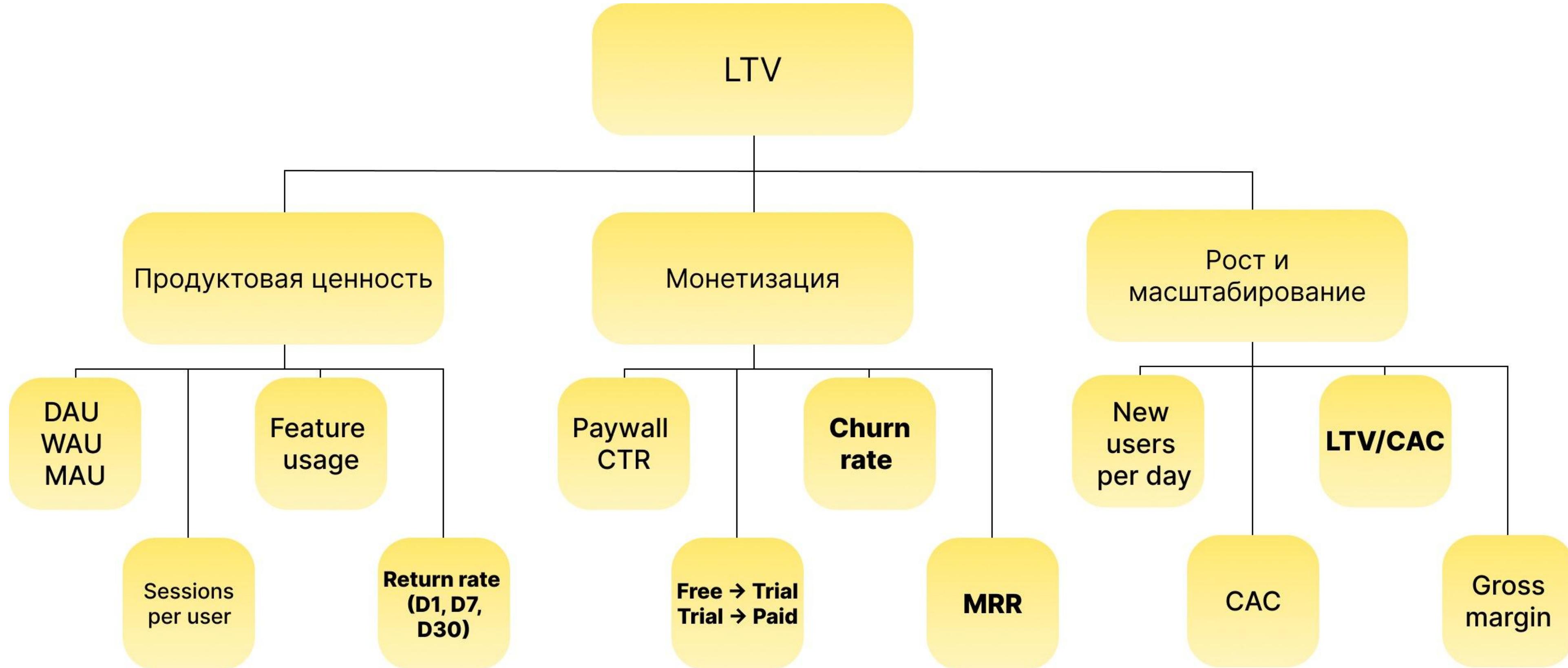
Конкуренты	Суммаризация новостей	Указание источников	Быстрое обновление	Совместимость с ТГ
AlphaSense	✓	✓	✓	✗
AltSignals	✓	✗	✓	✓ (В виде канала)
LevelFields	✗	✗	? (задержка ~40м)	✗
Incite AI	✓	✗	✓	✗
Perplexity	✓	?	✓	✗
T-plexity	✓	✓	✓	✓



Интерфейс



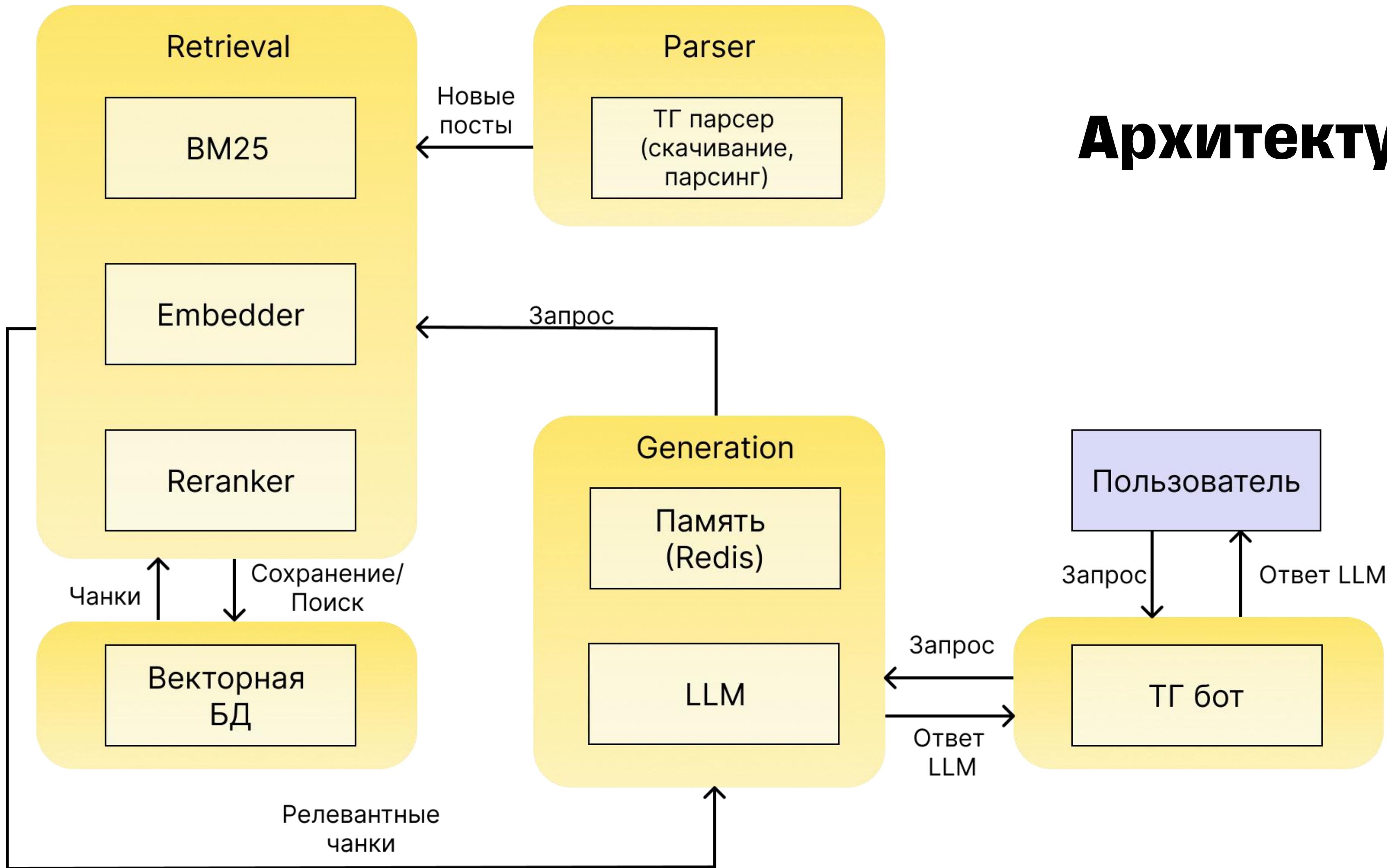
Дерево метрик



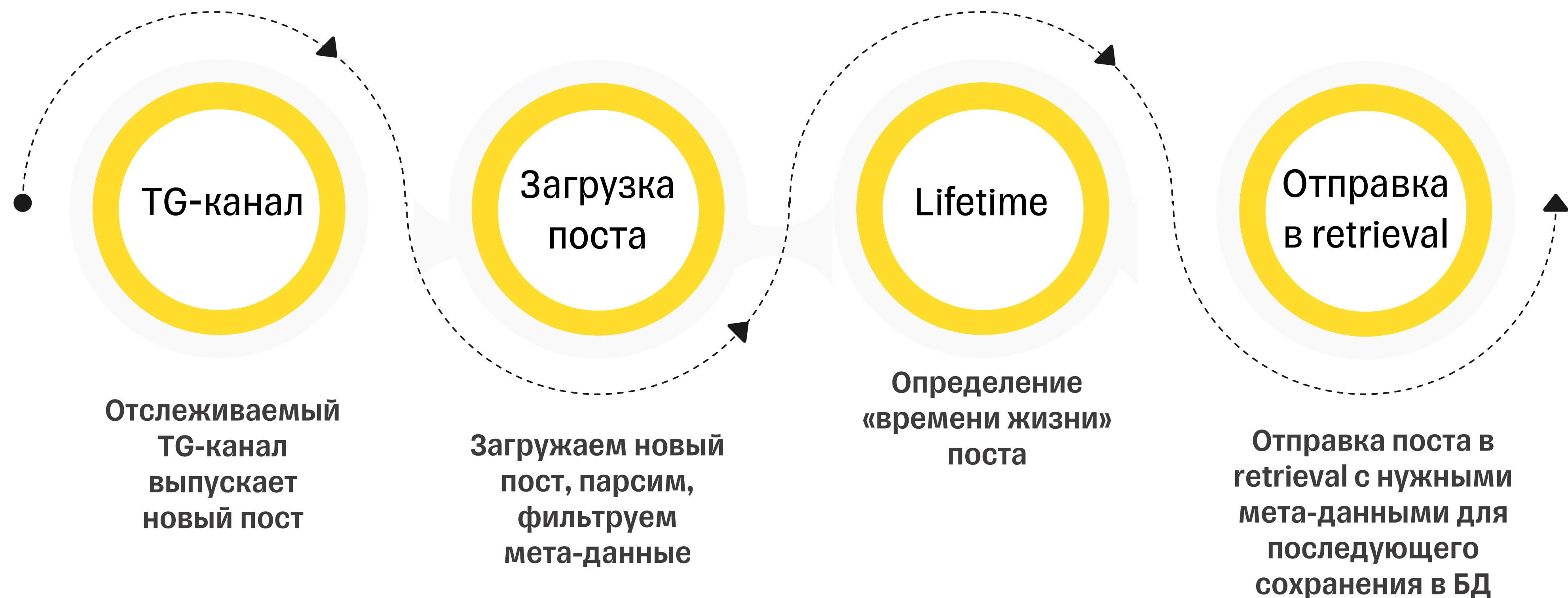


Архитектура

Архитектура



ТГ-парсер



Retrieval

1

2

3

4

Router-агент

Определение необходимости
вызыва retrieval

Предобработка

- Перефразировка
запроса, обогащение
ключевыми словами
- Лемматизация:
rutmorphy3

Гибридный поиск

- Dense: FRIDA
- Sparse: BM25
- Fusion: RRF
- Vector DB: Qdrant

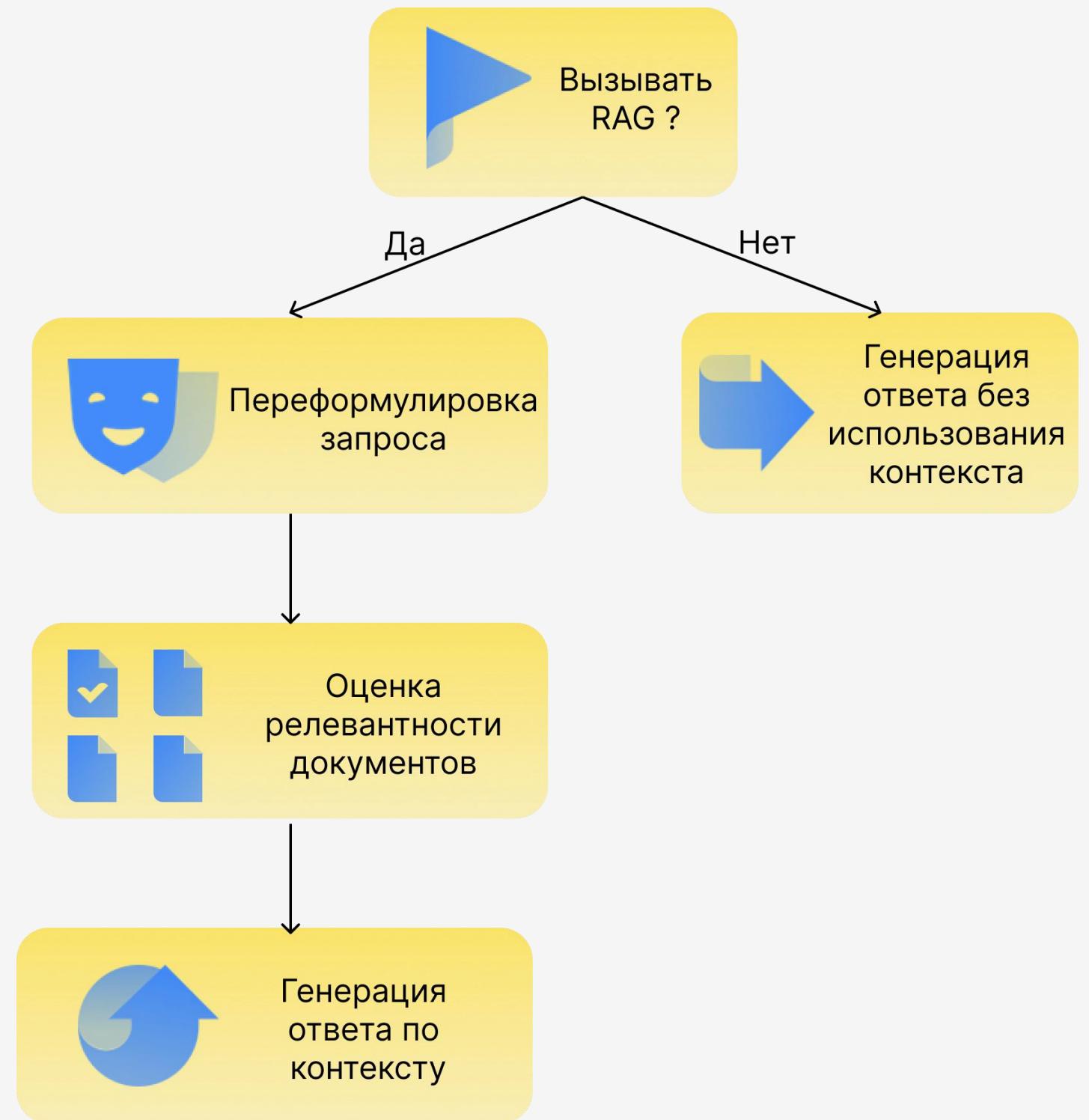
Фильтрация

- Relevance Evaluator
agent

Generation multi-agent

- **Router agent:** решает, нужен ли поиск в базе знаний (RAG) для ответа
- **Query Reformulation agent:** переформулирует запрос для улучшения поиска
- **Relevance Evaluator agent:** бинарная оценка релевантности документов запросу
- **Generation agent:** генерирует финальный ответ на основе контекста

Схема работы

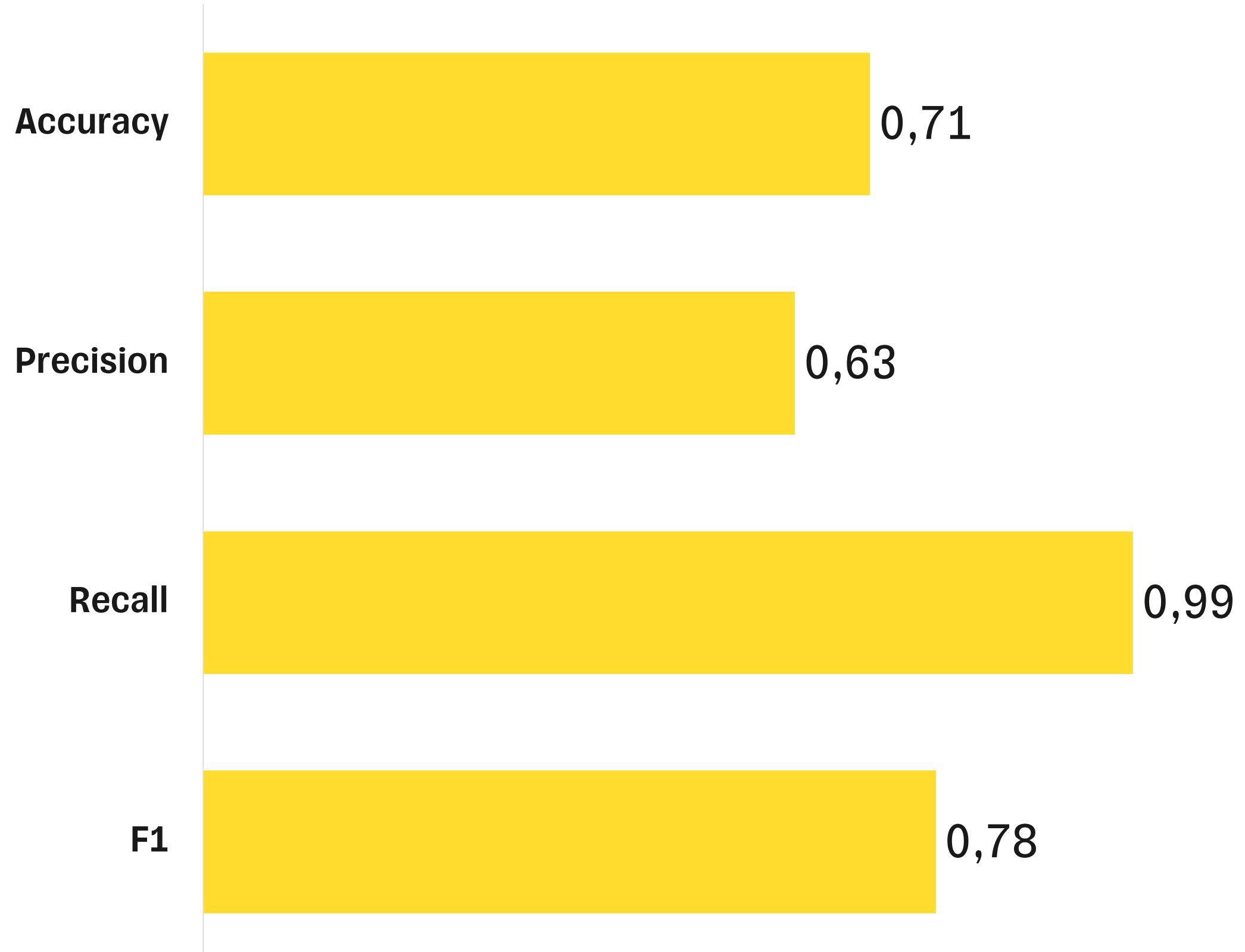




Eval

Binary Classification metrics

Ошибка на старте пайплайна – гарантированная ошибка на финише. Мы должны быть уверены, что этот компонент не становится узким местом, а помогает улучшать производительность и точность всей архитектуры.



Binary Classification metrics

Пример данных:

```
{"query": "Каков курс евро и юаня на 14 ноября 2025 года по да...",  
"doc_id": "fff543e5-99c6-41be-a60...",  
"document_text": "Дата публикации: 14 ноября 2025...  
"expected_router_decision": "YES",  
"requires_rag": true}
```

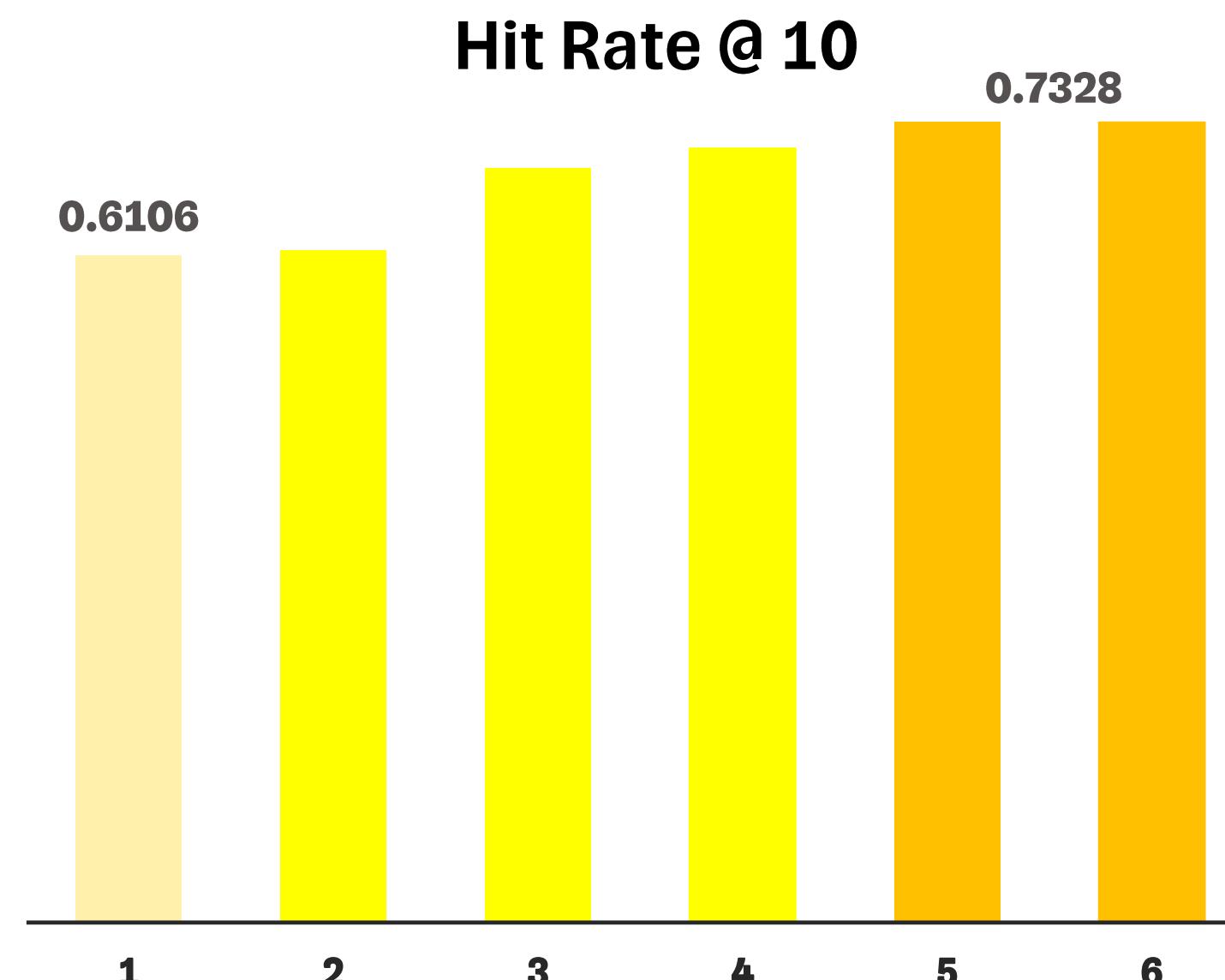
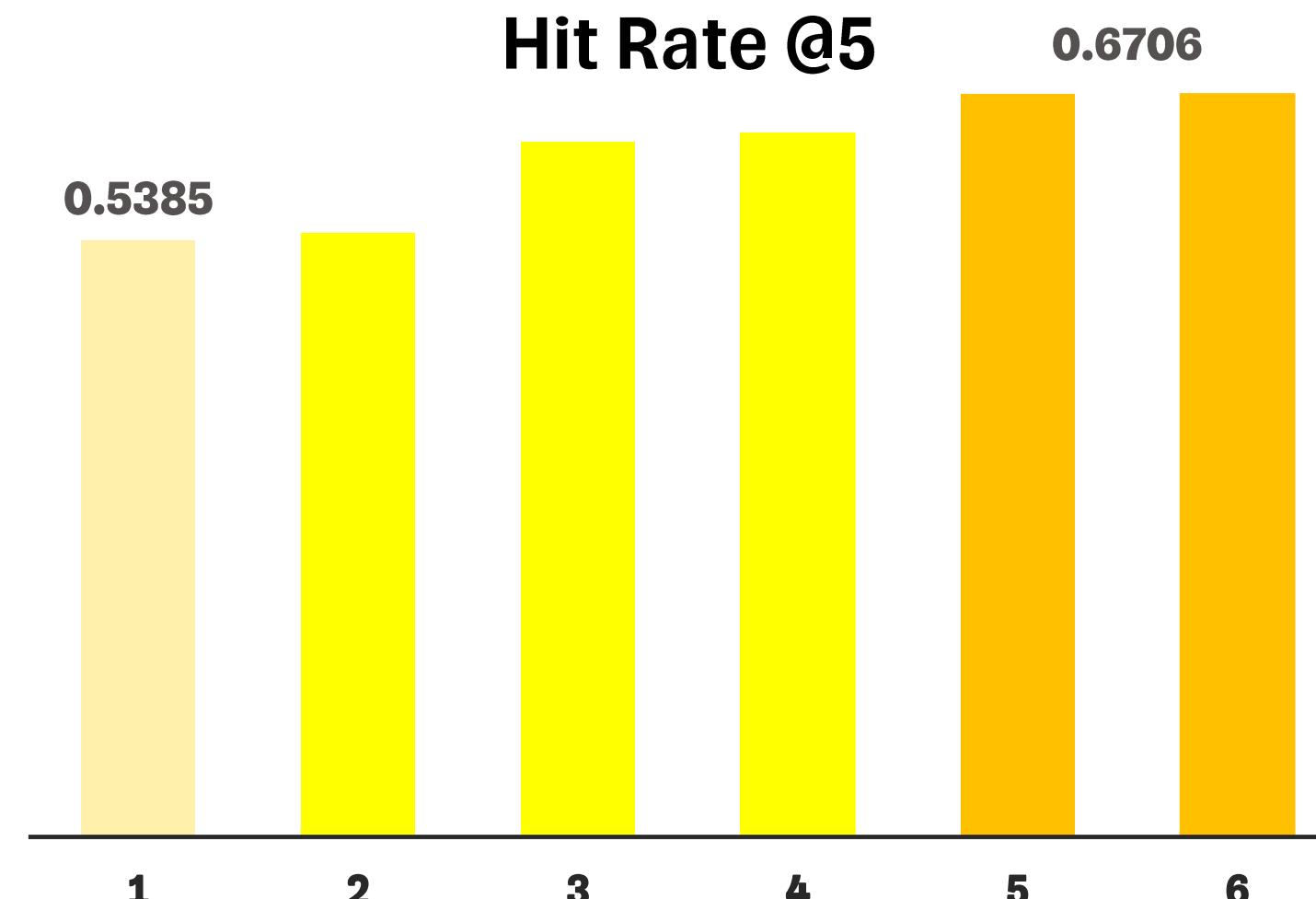
```
{"query": "Что такое дивидендная политика компании?",  
"doc_id": "0065e9a4-858b-4976-877f-d9f1d463c295",  
"document_text": "Дата публикации: 10 ноября 2025...  
"expected_router_decision": "NO",  
"requires_rag": false}
```

Оценка Retrieve

Hit Rate @ k

Данная метрика считает попал ли якорный документ, на котором основан запрос в первые k ответов, которые извлек Rerieve

- 1 – Frida
- 2 – BM25
- 3 – BM25 + Lemma
- 4 – Jina + BM25 + Lemma
- 5 – Frida + BM25 + Lemma**
- 6 – Frida + BM25 + Lemma+Reranker

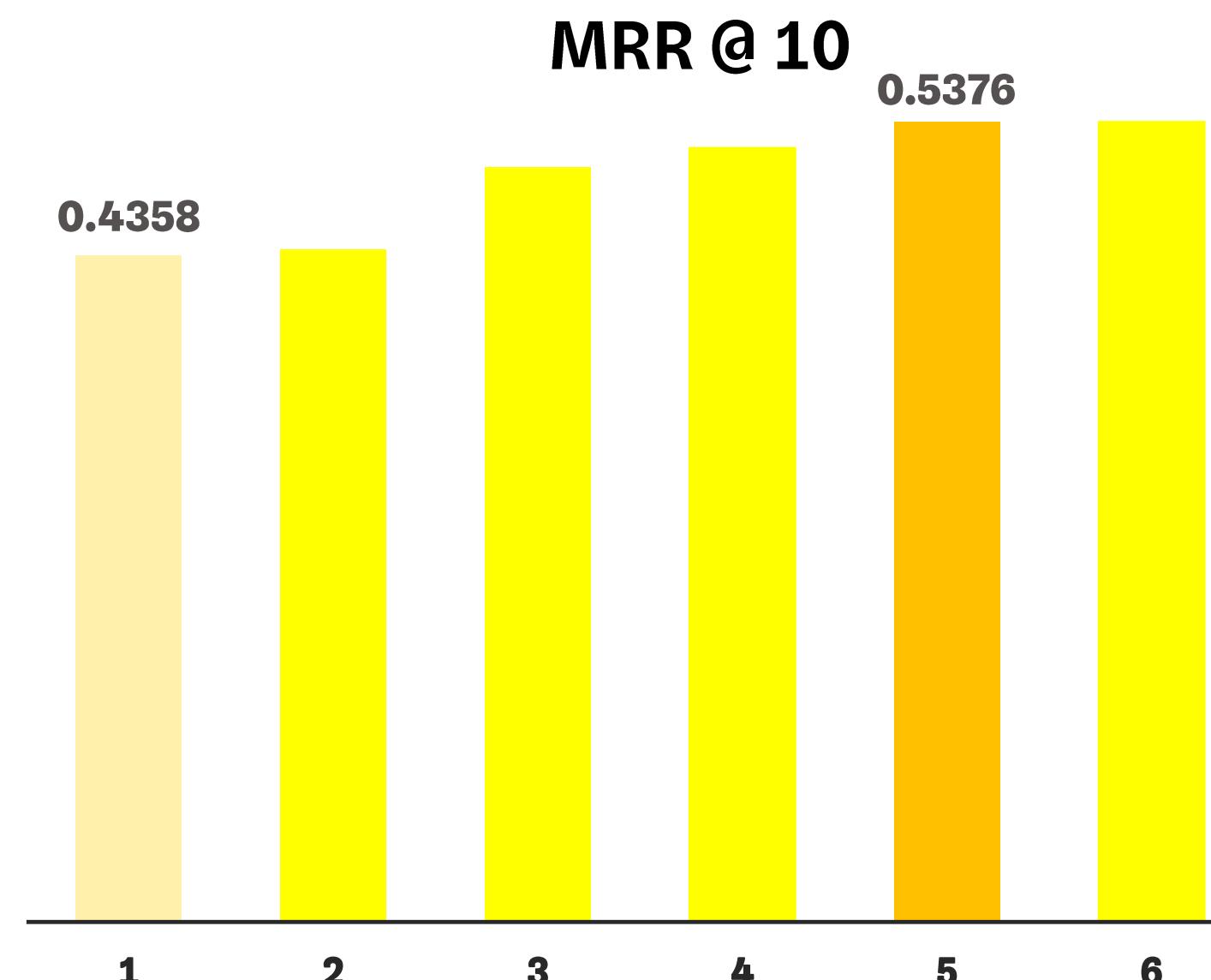
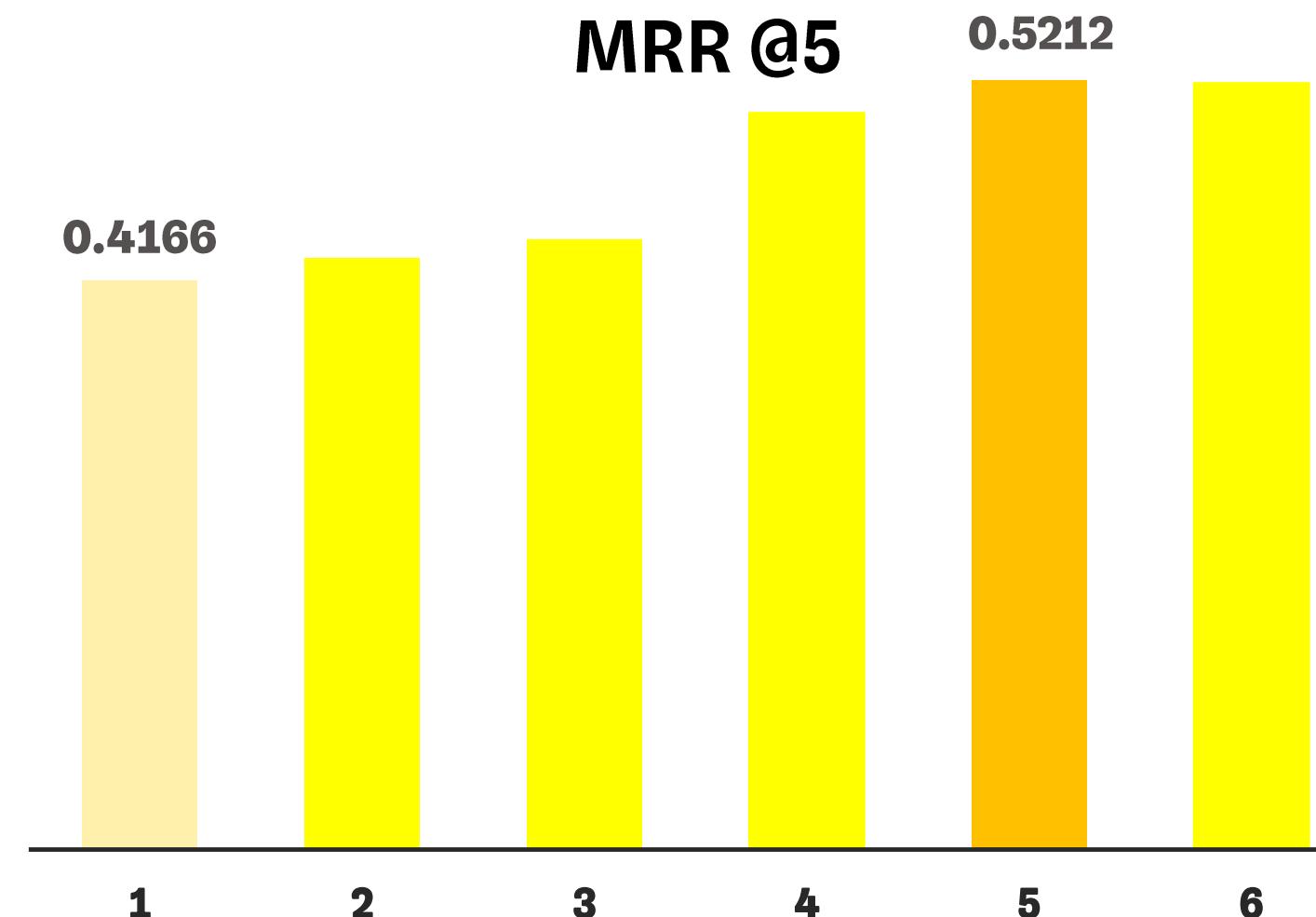


Оценка Retrieve

MRR @ k

Данная метрика показывает
насколько высоко в списке
результатов поиска находится
якорный документ

- 1 – Frida
- 2 – BM25
- 3 – BM25 + Lemma
- 4 – Jina + BM25 + Lemma
- 5 – Frida + BM25 + Lemma**
- 6 – Frida + BM25 + Lemma+Reranker



Пример Retrieve

MRR x Hit Rate

Пример данных

```
{ "id": 7077,  
  "channel_id": 1418181070,  
  "link": "https://t.me/alfa_investments/7077",  
  "text": "***Стартовал сбор заявок на ..."}
```

Пример запроса

```
{ "query": "Какие текущие тренды в области ...  
  "id_channel": 1418181070,  
  "id_message": 7078}
```

Оценка Generation

LLM is Judge

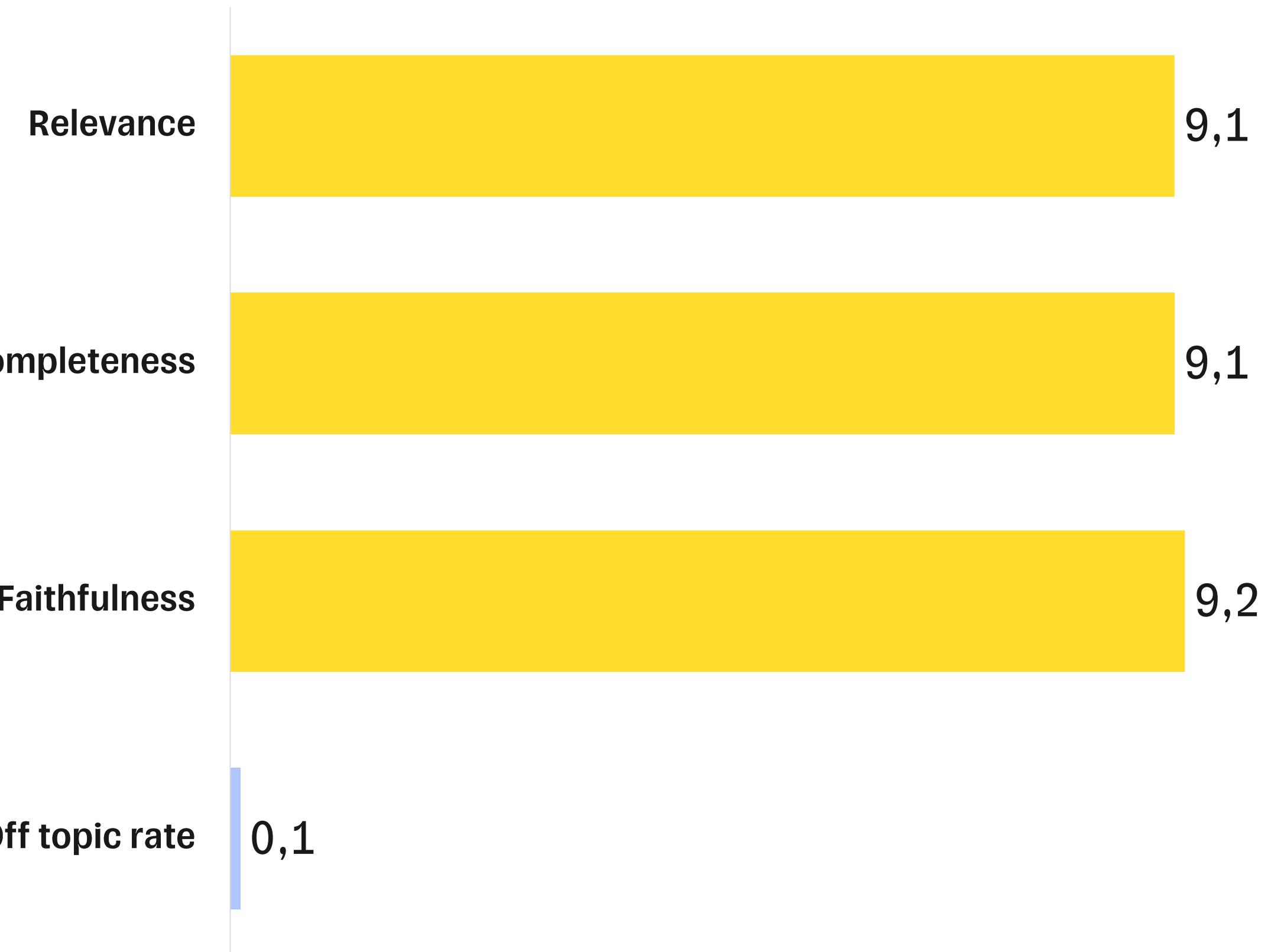
Оценивали решение при помощи LLM
через критерии:

Relevance - релевантность ответа
запросу

Достоверность - насколько факты
соответствуют контексту

Completeness - насколько ответ
покрывает важные пункты

Off-topic - доля ответов не по теме
запроса



Пример Generation

LLM is Judge

```
{"query": "Почему акции Северстали и ИКС 5 показали рост, несмотря на общее давление на нефтяной сектор и геополитическую неопределенность?",  
"id_channel": 1418181070,  
"id_message": 7076,  
"query_num": 4}
```

Пример ответа модели:

Акции Северстали и ИКС 5 продемонстрировали рост...

Пример ответа судьи:

```
"relevance": 0.95,  
"faithfulness": 0.9,  
"completeness": 0.9,  
"off_topic_rate": 0.01,
```

Планы на будущее

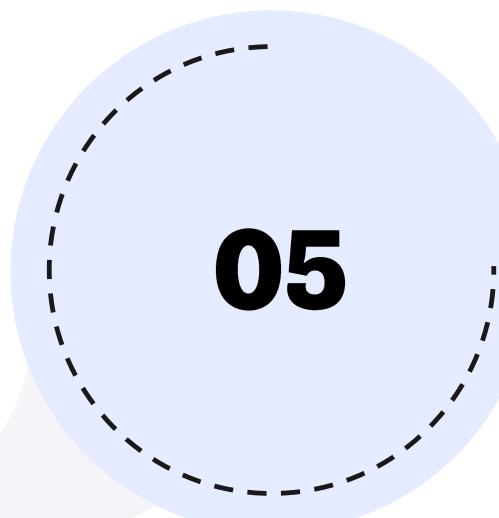
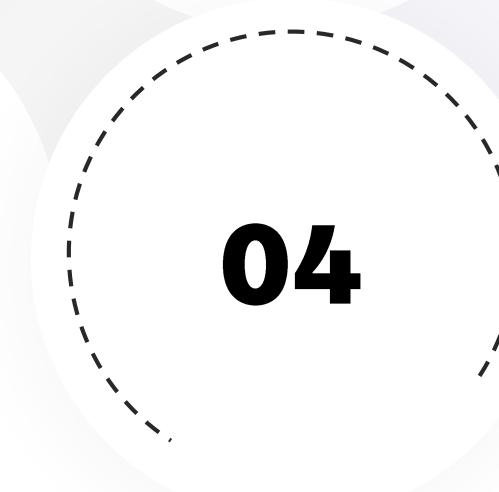
Телеграм бот
T-plexity



Реализация
рабочего веб
интерфейса.
Исследование UI/UX



Расширение
доменов
источников



Переход на
подписочную систему
монетизации

Использование
temporal-RAG

Команда разработки

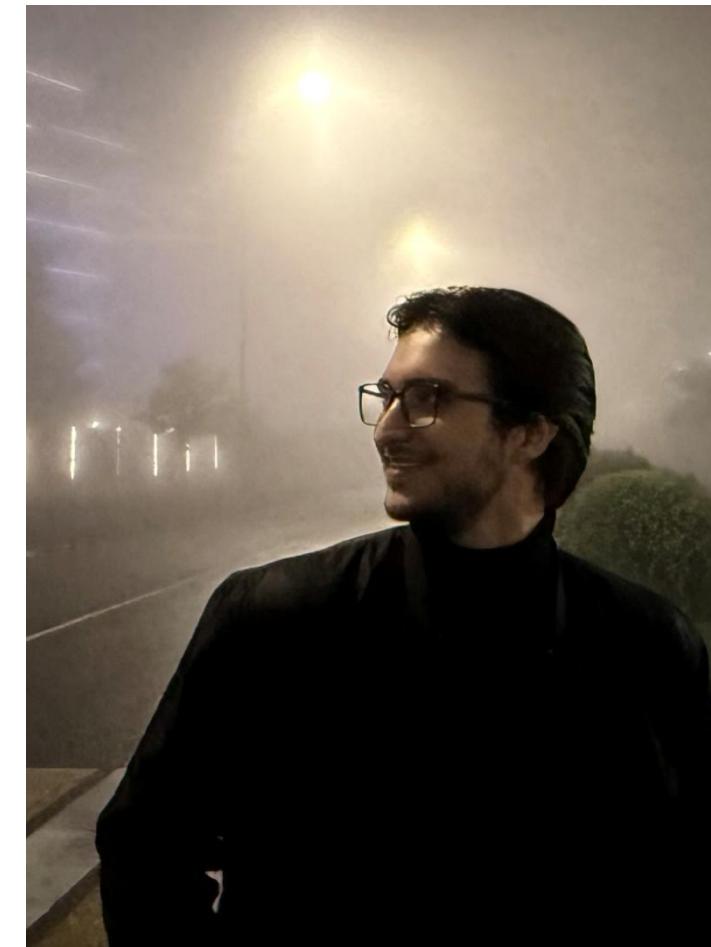
Липилин Матвей
Ментор



Цымбалов Александр
Ментор



Кладиенко Иван
NLP Engineer



Елдашов Владислав
NLP Engineer



Лопатин Иван
NLP Engineer



Убушаева Яна
Product Manager



Спасибо!

