

# Рекомендательные системы и информационный поиск

## Введение

Краснов Александр

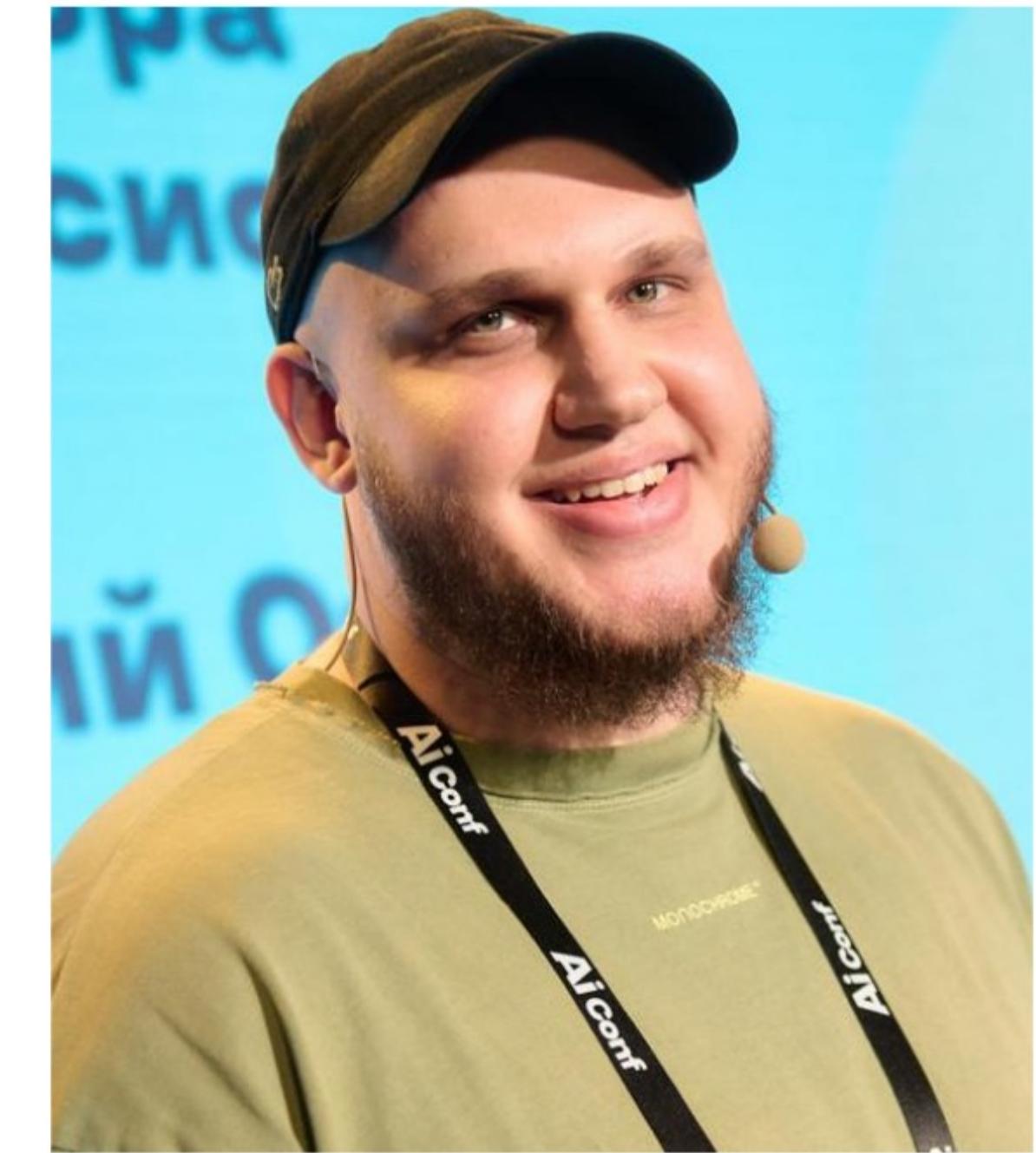
# Команда курса



Илья Осиновсков  
@ossinovskov



Дмитрий Андросов  
@androsovdy



Александр Краснов  
@san9kresh

# План курса

Введение, рекомендательные системы  Introduction to IR	Про рекомендации и поиск, задачи рекомендаций, метрики рекомендаций, простые алгоритмы рекомендаций
Ranked IR and search quality	Введение в информационный поиск, обработка текста, обратный индекс, представление поискового запроса, создание базовой поисковой системы
Методы основанные на статистиках	Поиск и ранжирование, TF-IDF и BM25, векторное представление текста, оценка качества поиска
Контентные рекомендации	Ассоциативные правила, коллаборативные подходы, SVD, ALS
LTR, Многоуровневые ранжирующие системы	Контентные рекомендации, сжатые представления контента, гибридные рекомендации, LightFM, DeepFM, использование ANN/KNN
Проблемы индустриальных рекомендательных систем	Задача LTR, лоссы и метрики, разбор концепций многоуровневых систем ранжирования
	Разнообразие, Холодный старт контента/юзера, popularity/position bias, exploration-exploitation tradeoff

# План курса

Neural IR and language models	Векторное представление текста, векторный поиск, популярные архитектуры нейросетей для поиска, Использование LLM для поиска и рекомендаций
Sequential-based подходы	Трансформеры в рекомендациях v1, SASRec, BERT4REC
Генеративные рекомендации	Semantic IDs, HSTU, OneRec
Графовые сети	Обзор графовых сетей, граф знаний
Search query auto completion	Задача поисковых подсказок, префиксный поиск, ранжирование подсказок, персонализация
Технологии рантайм рекомендаций	Разбор технологий движка персонализаций Ozon и других компаний
Технологии рантайм поиска	Разбор технологий поискового движка Ozon и других компаний

# Оценивание

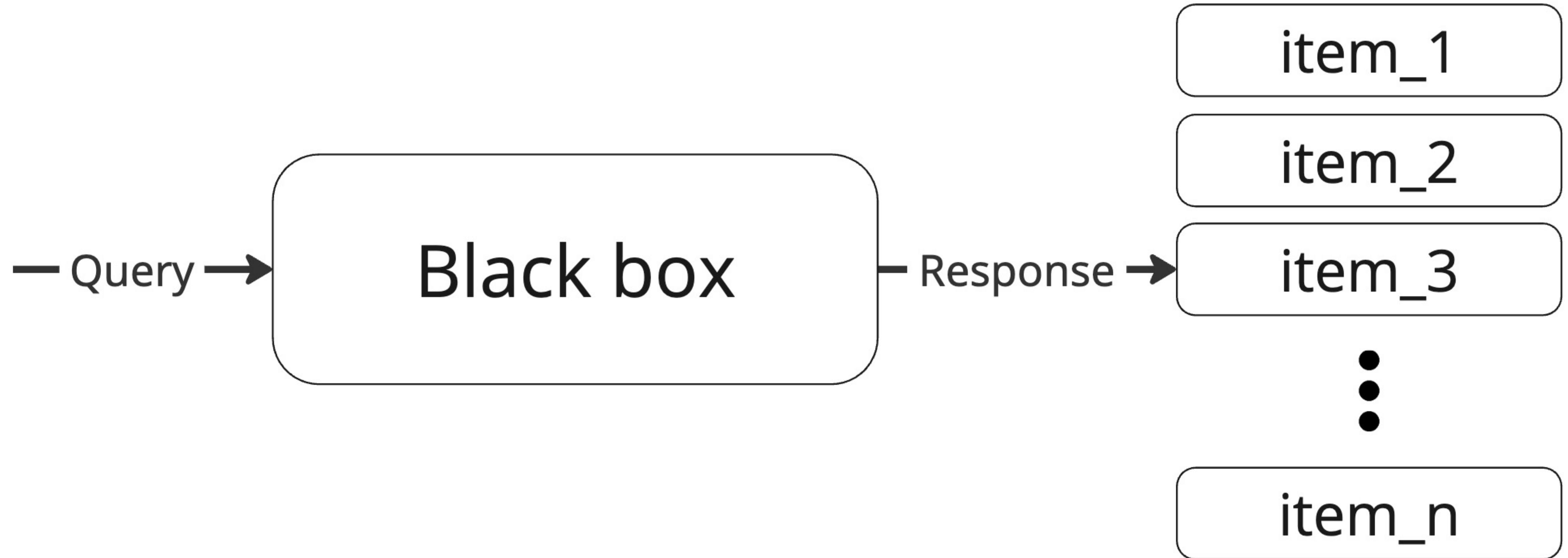
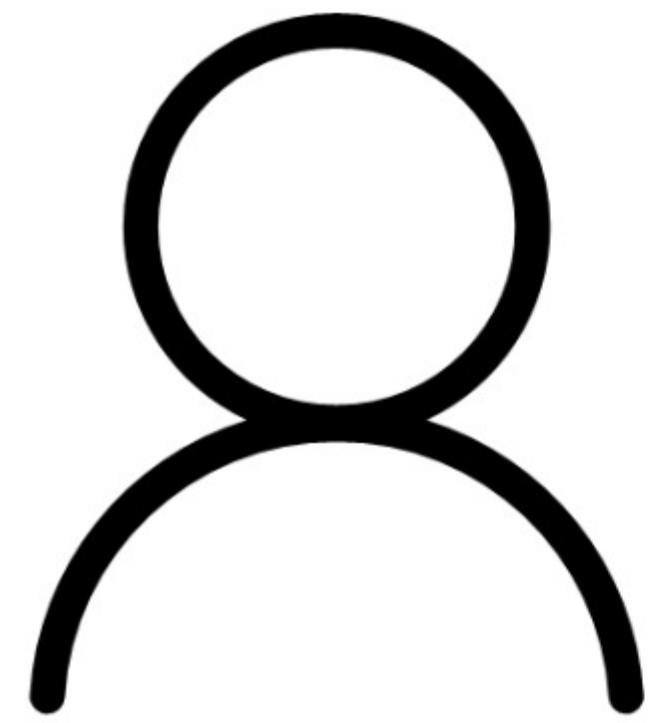
6 домашек, 10 баллов макс за каждую

- собственные реализации алгоритмов и оценка качества
- 1+ контест
- реализации алгоритмов в сервисе

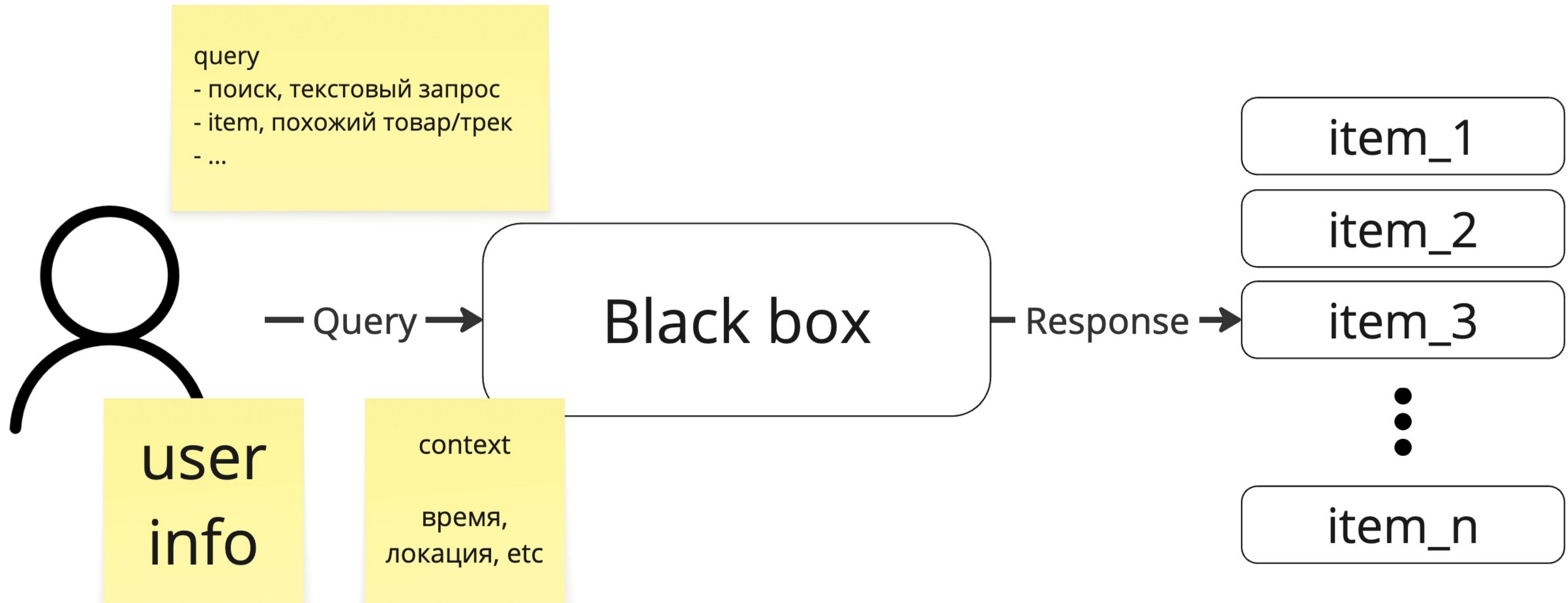
Общая оценка за курс - сумма баллов / кол-во дз

# Вопросы

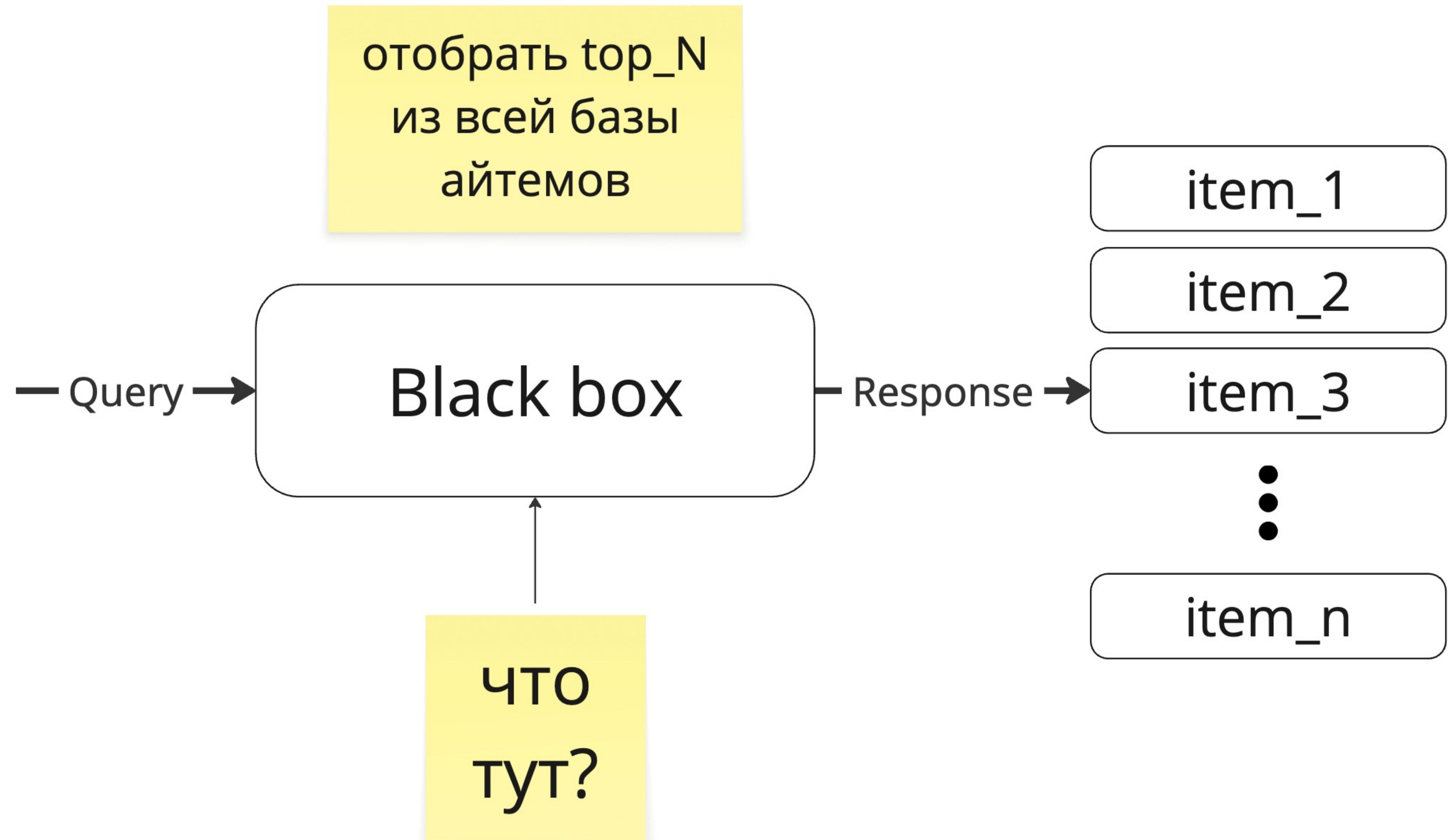
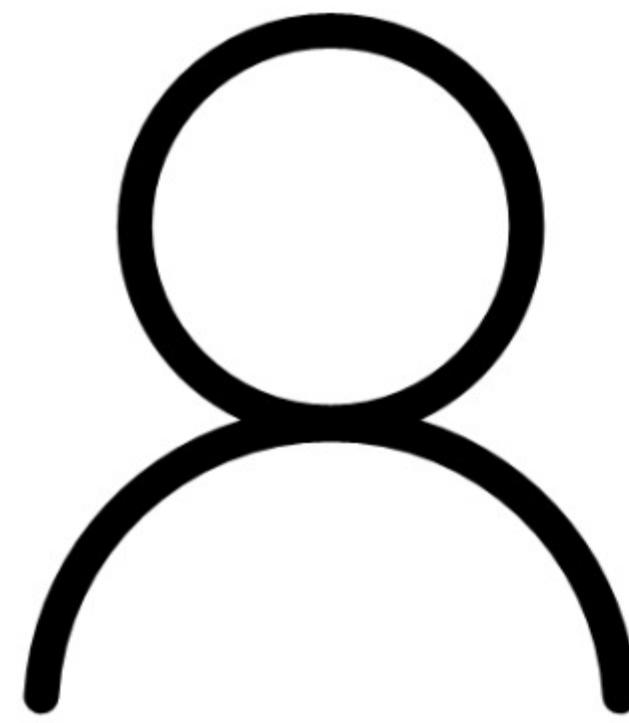
# Контекст



# Контекст



# Контекст



## RS и IR

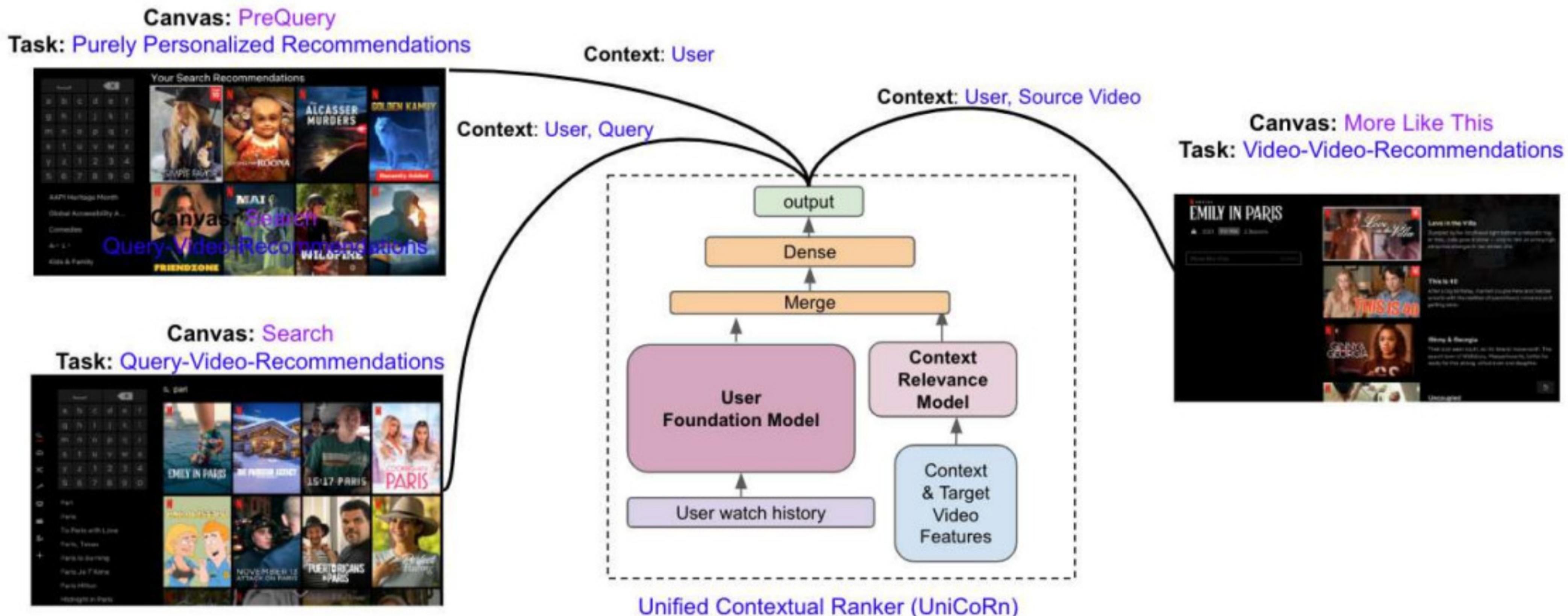
**Рекомендательные системы** (Recommender system, RS) — это программные инструменты и алгоритмы, предназначенные для анализа данных о пользователях, их предпочтениях и поведении, чтобы предлагать наиболее **релевантные** продукты, услуги или контент.

**Информационный поиск** (Information retrieval, IR) — это процесс поиска, отбора и предоставления **релевантной** информации в ответ на запрос пользователя.

# RS и IR, пересечения

- данные
  - текстовые данные, метаданные, историю взаимодействий (просмотры, клики, покупки)
  - неявная (implicit) обратная связь (время просмотра документа/трека)
- ranking
  - LTR
  - формирование финального ответа (разнообразие)
- персонализация
  - учет данных пользователя для формирования ответа
- технические решения
  - обратные индексы (ключ - кандидаты)
  - feature store
- ML
  - векторный поиск
  - схожие идеи в DL подходах (токены текста (IR) - токены действий (RS))
  - объединение контекста в общей задаче
- Метрики ранжирования

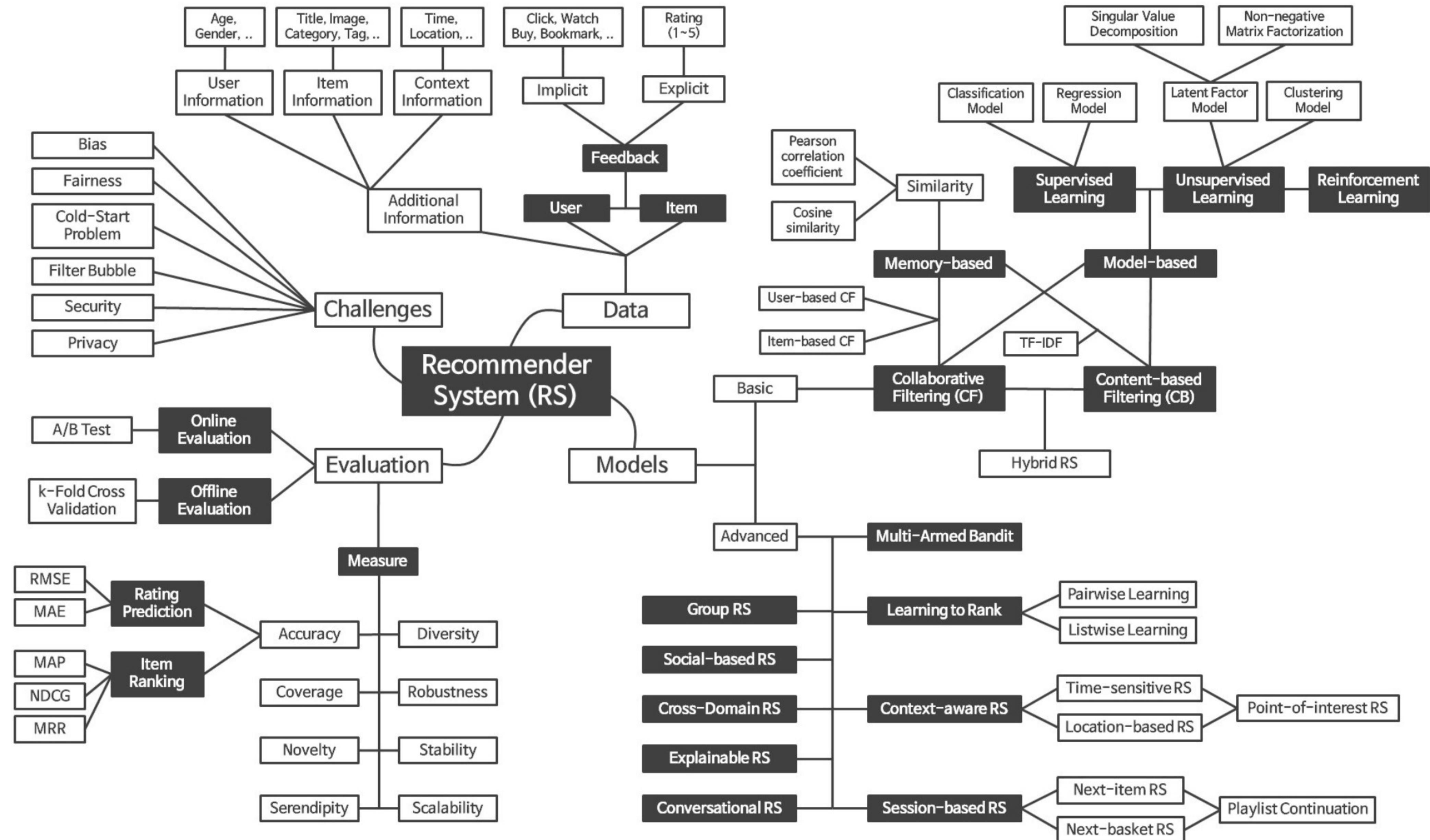
# RS и IR, пересечения



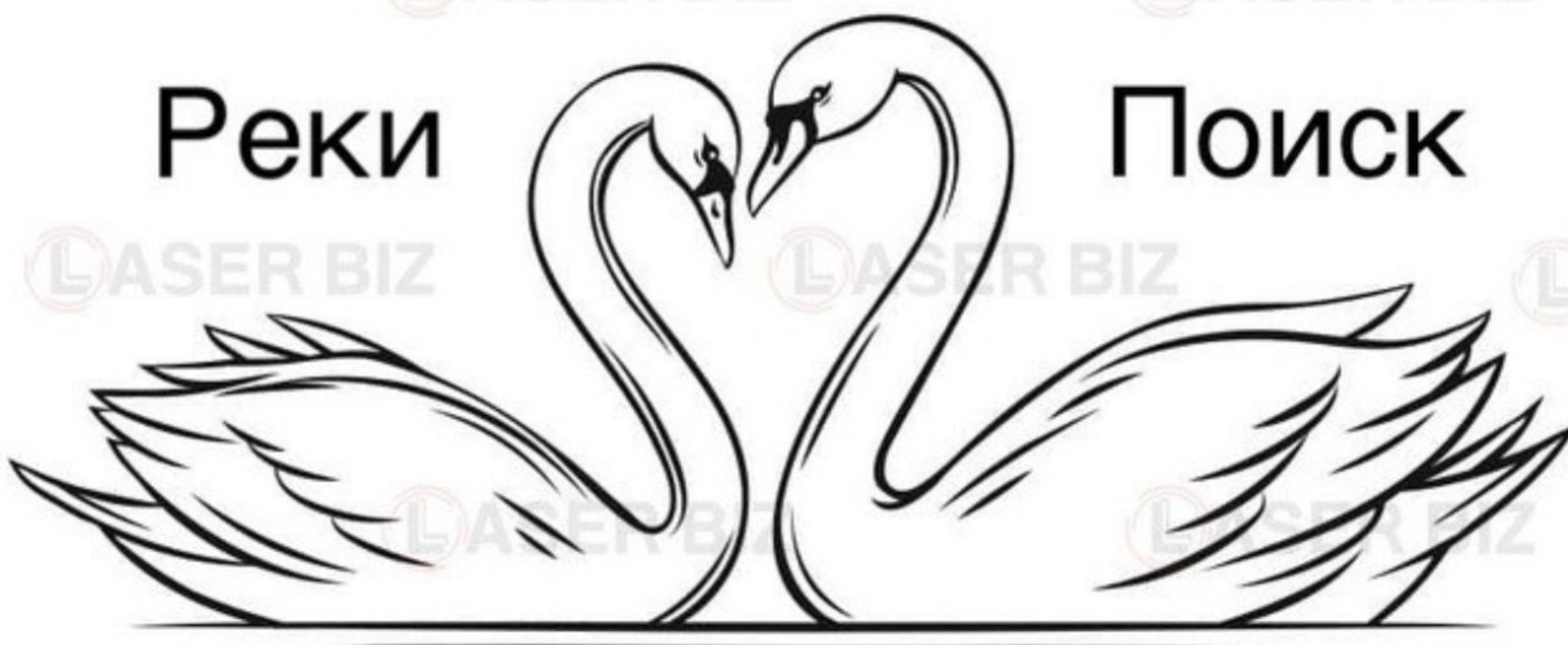
**Figure 1: Unified Contextual Ranker (UniCoRn) powering multiple different search and recommendation tasks**

# RS и IR, отличия

- данные
  - IR - первостепенный акцент на тексте
  - RS - данные о пользователе
- classic ML
  - IR - BM25, TF-IDF и прочее
  - RS - CF, ALS, EASE и прочее
- оценка качества
  - RS - метрики разнообразия (diversity novelty), редко используют оценку асессоров
  - IR - текстовая релевантности - оценка от асессоров



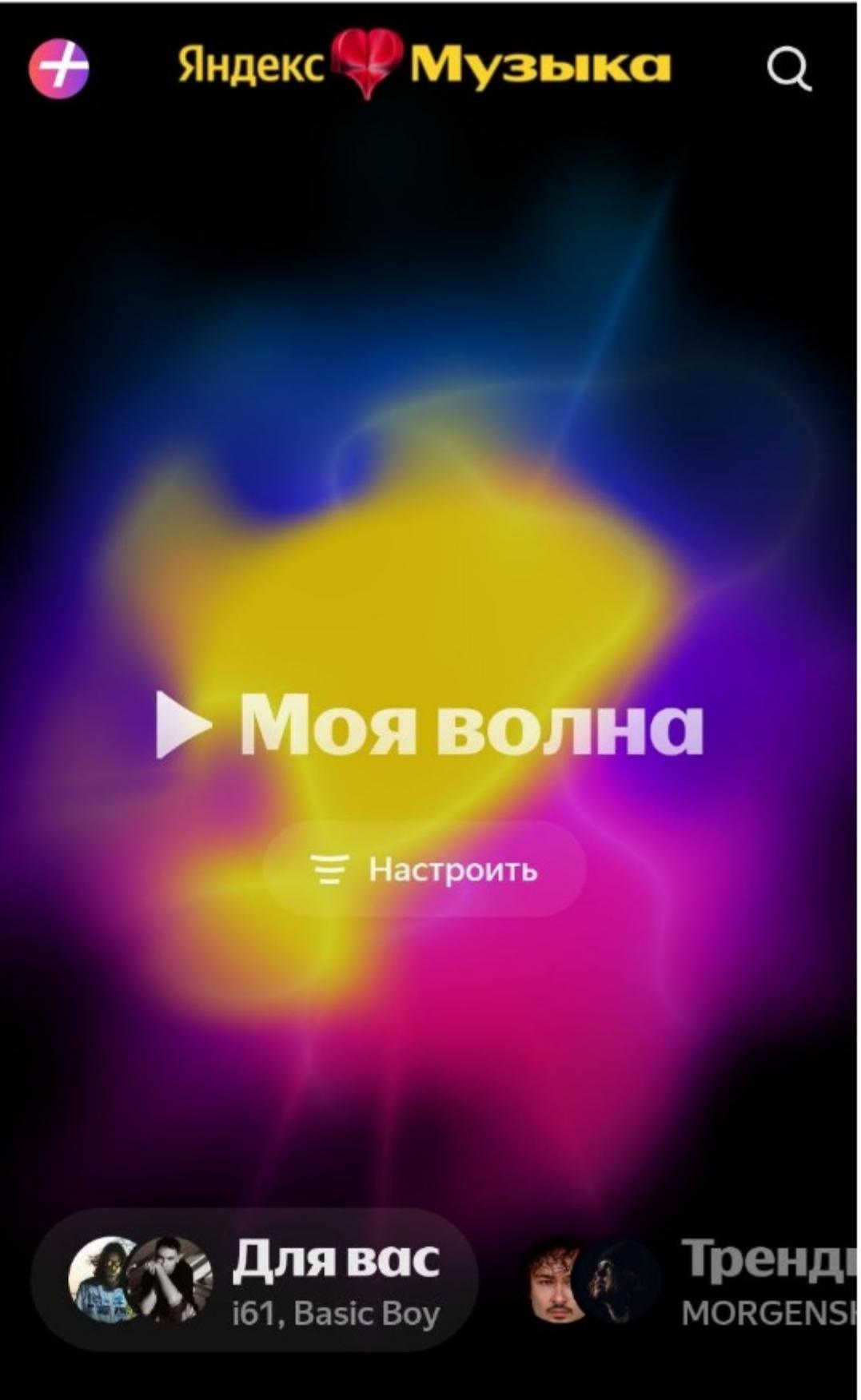
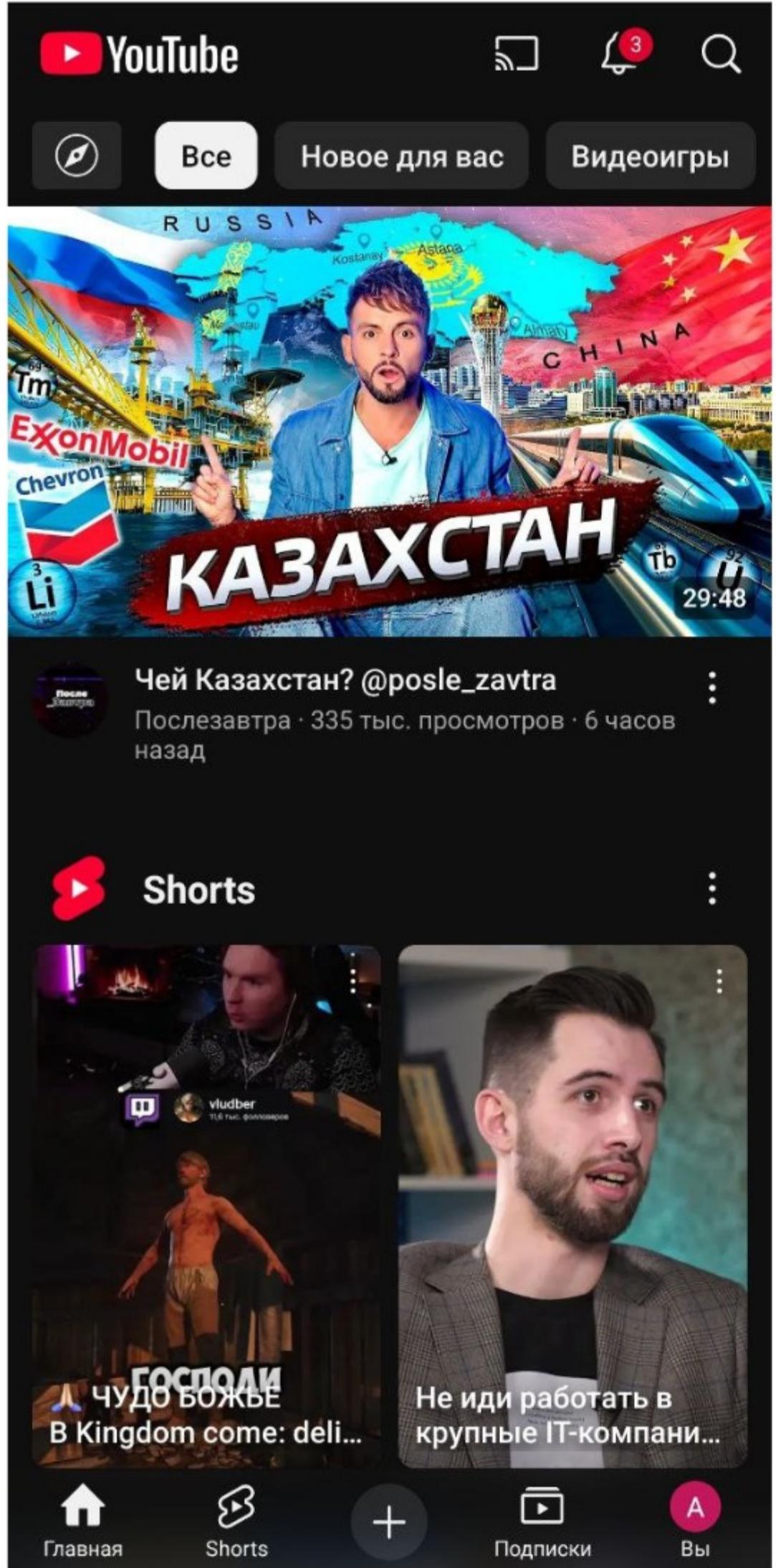
# Вопросы



Реки

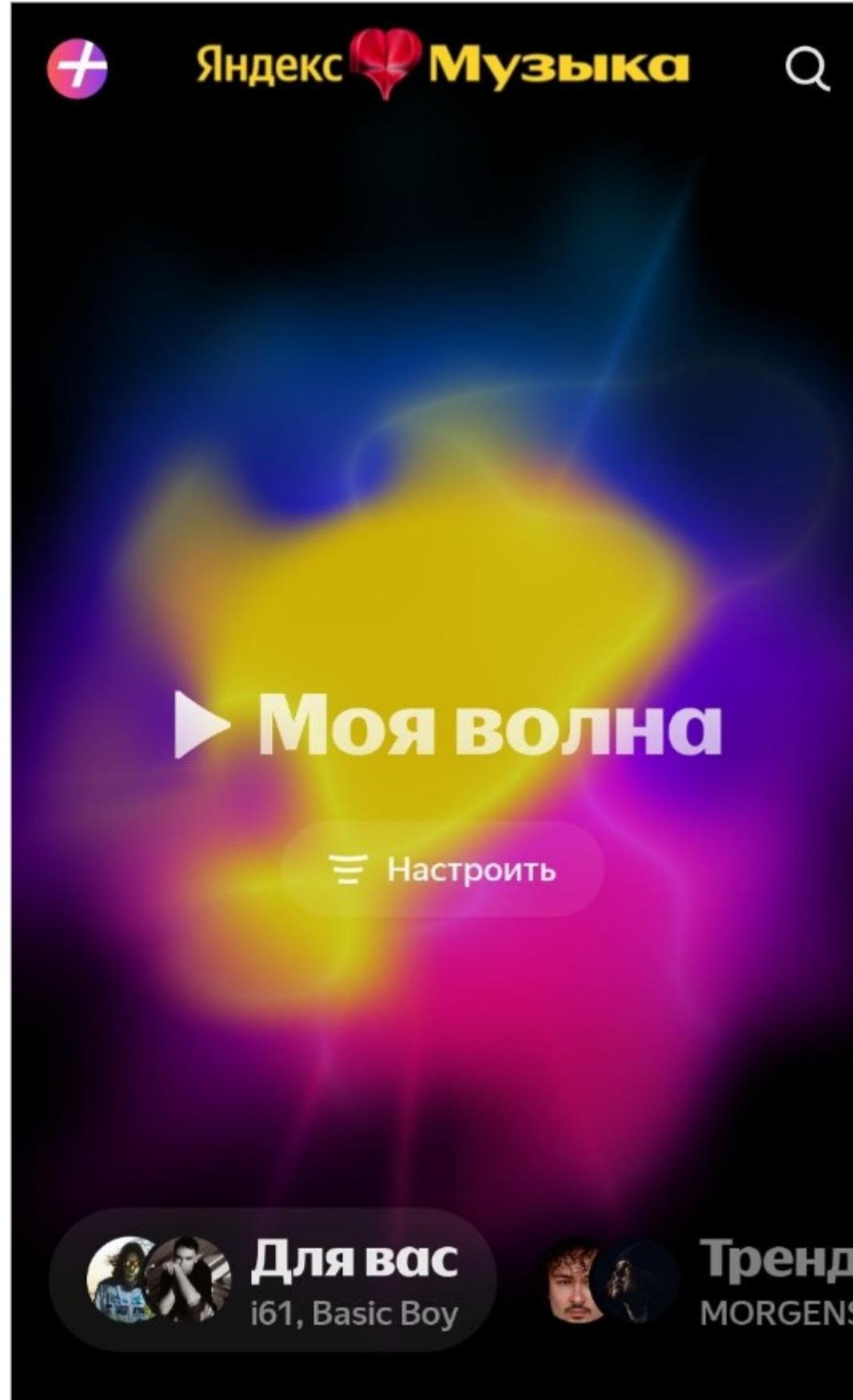
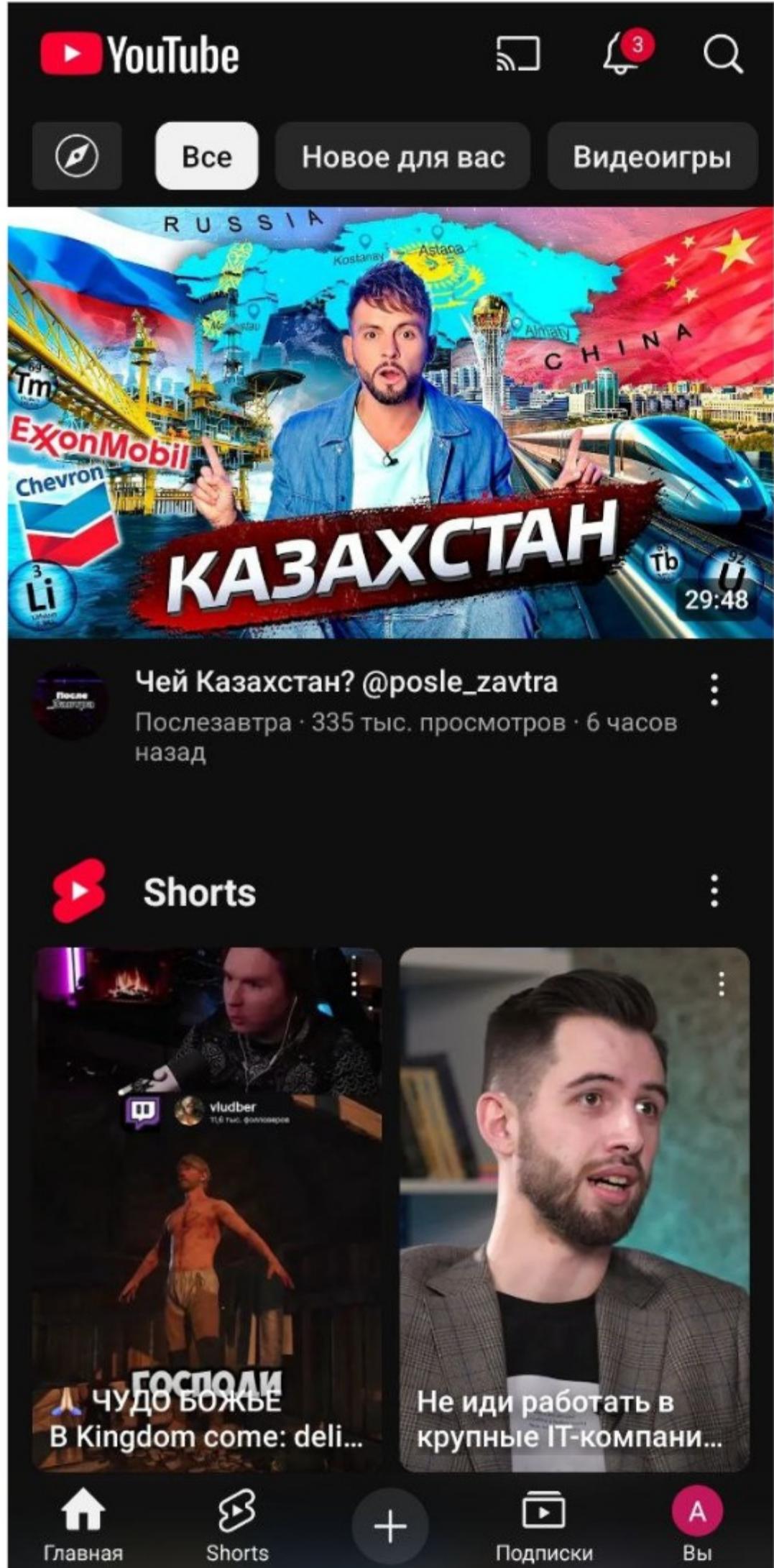
Поиск

# RS примеры



зачем  
рекомендации  
пользователю/  
компании/  
авторам?

# RS примеры



## пользователю

- более релевантный контент
- упрощение поиска контента
- exploration

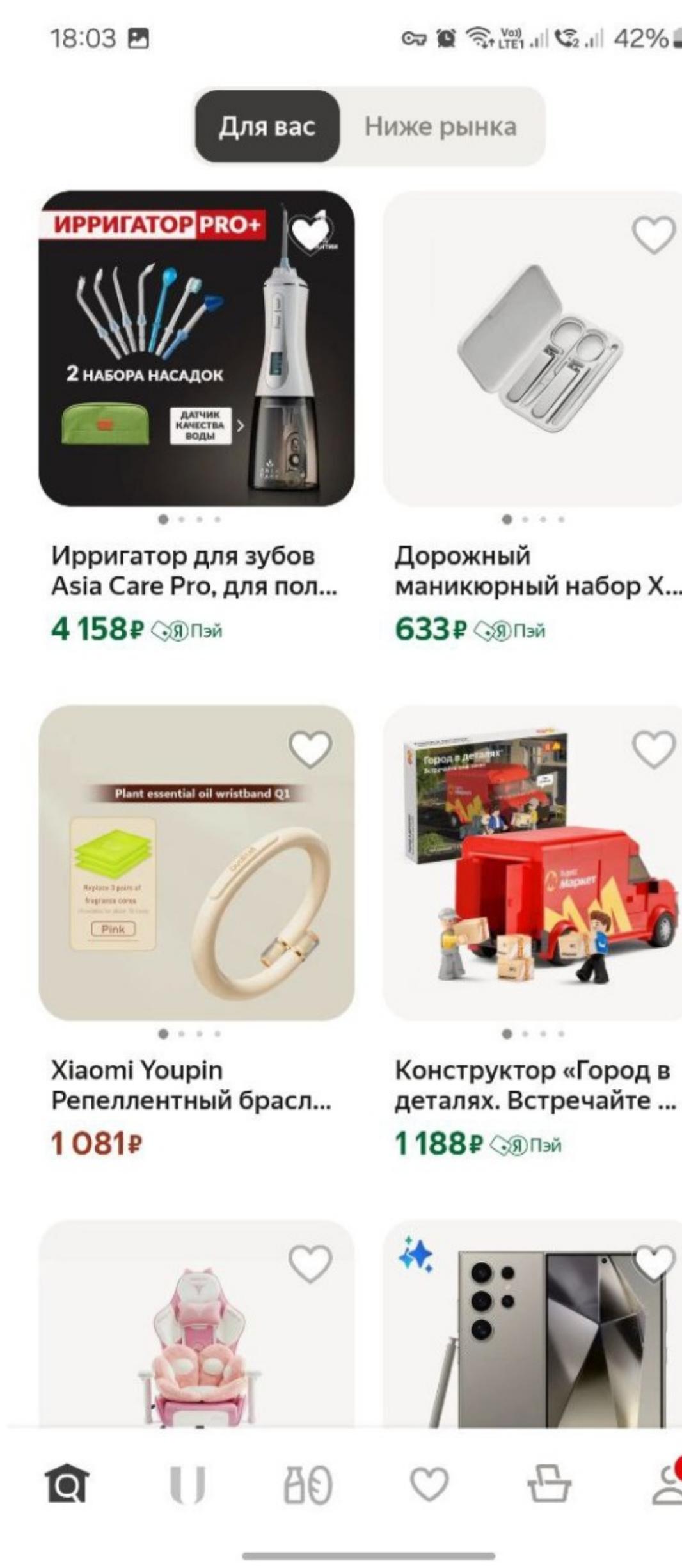
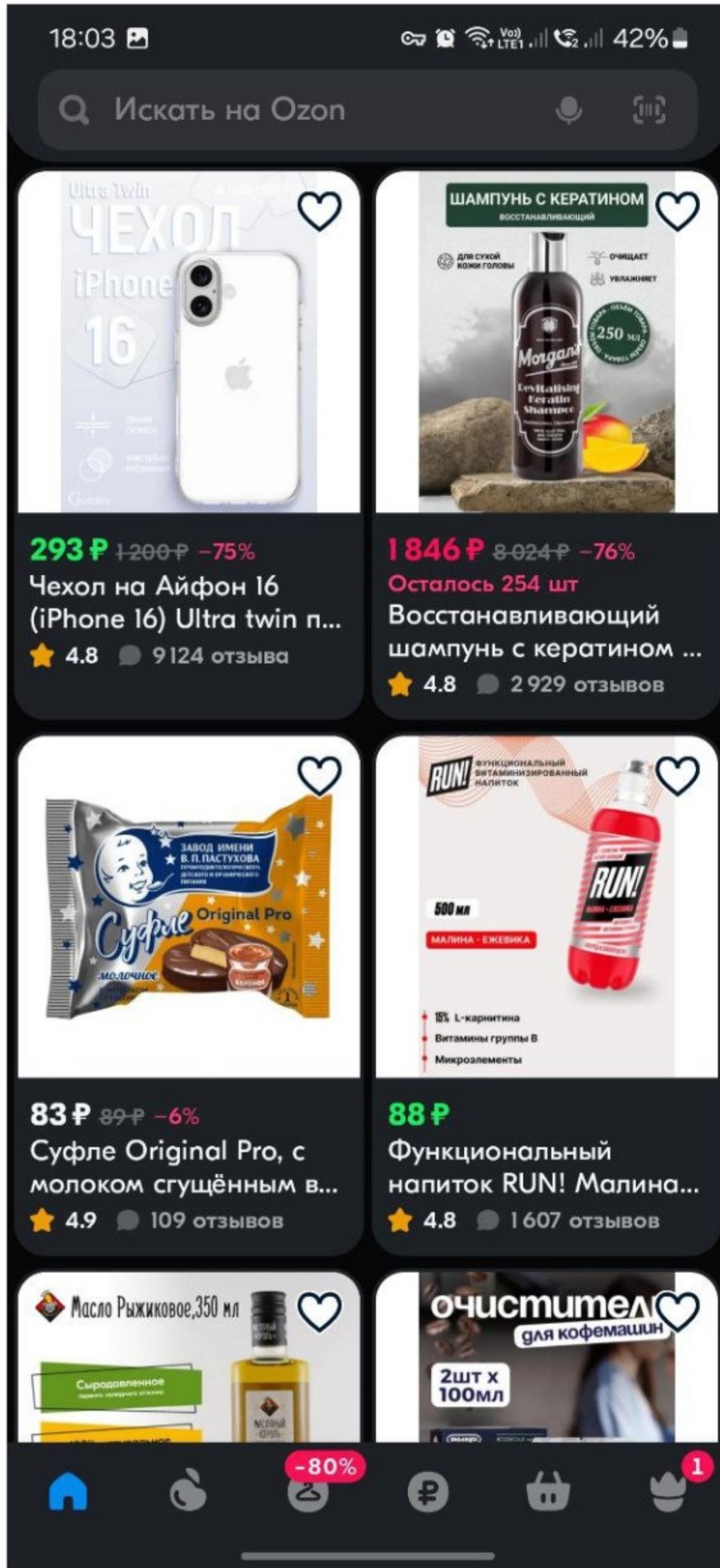
## компании

- увеличение метрик (time spent, adrev)
- увеличение вовлечённости пользователей
- поддержка авторов контента
- увеличение рекламных доходов

## авторам

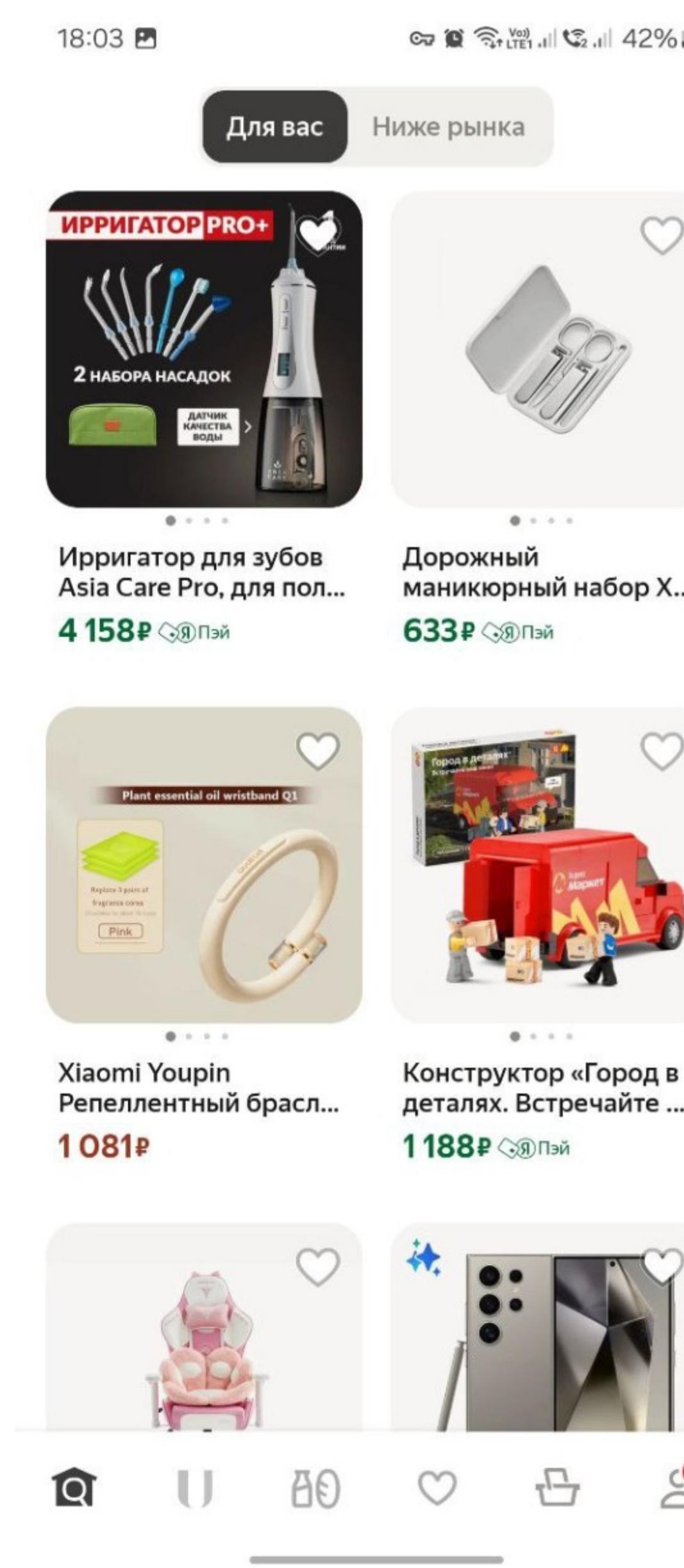
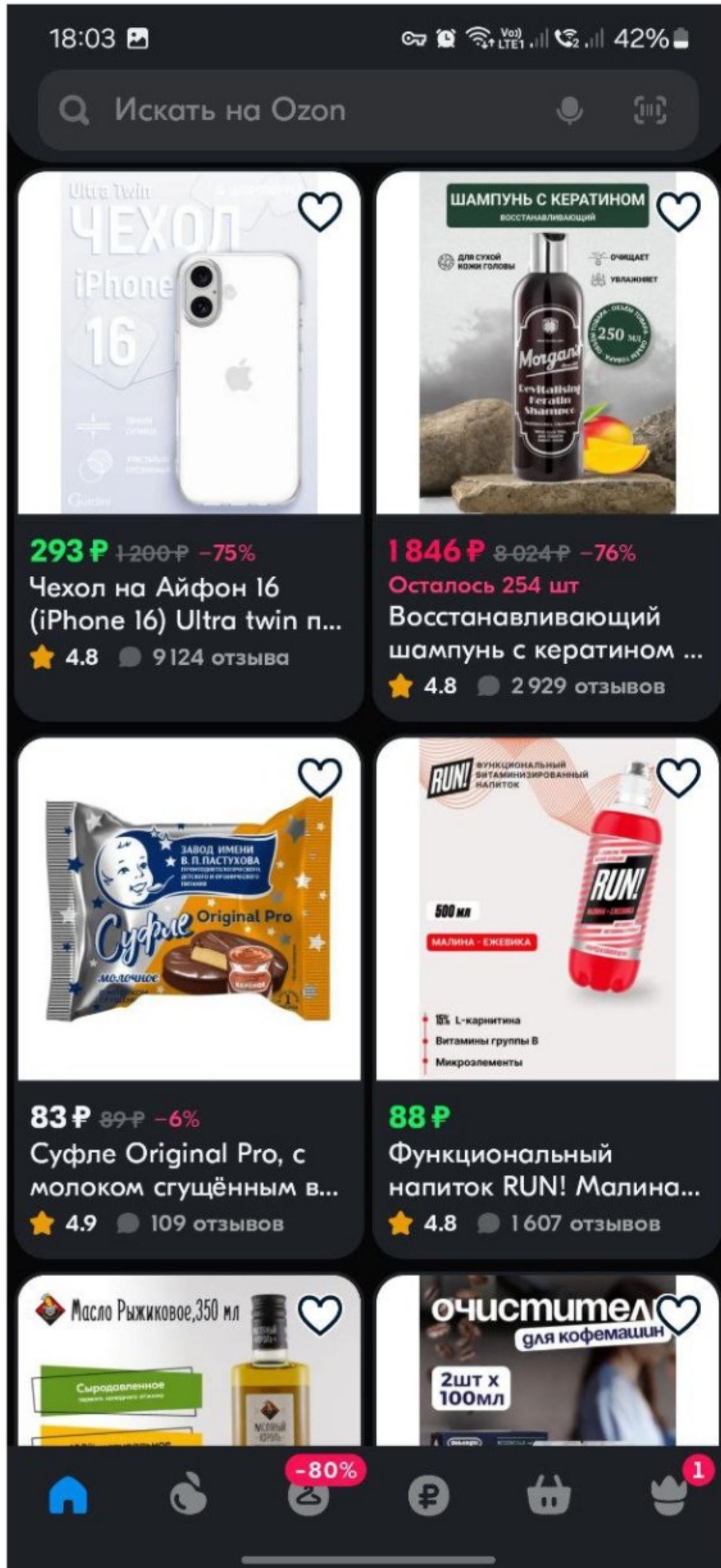
- поиск аудитории
- поиск новой аудитории
- партнерские программы (adrev)

# RS примеры



зачем  
рекомендации  
пользователю/  
компании/  
продавцам?

# RS примеры



## пользователю

- более релевантные товары
- exploration
- выгодные предложения
- выдача, учитывающие "интересы"

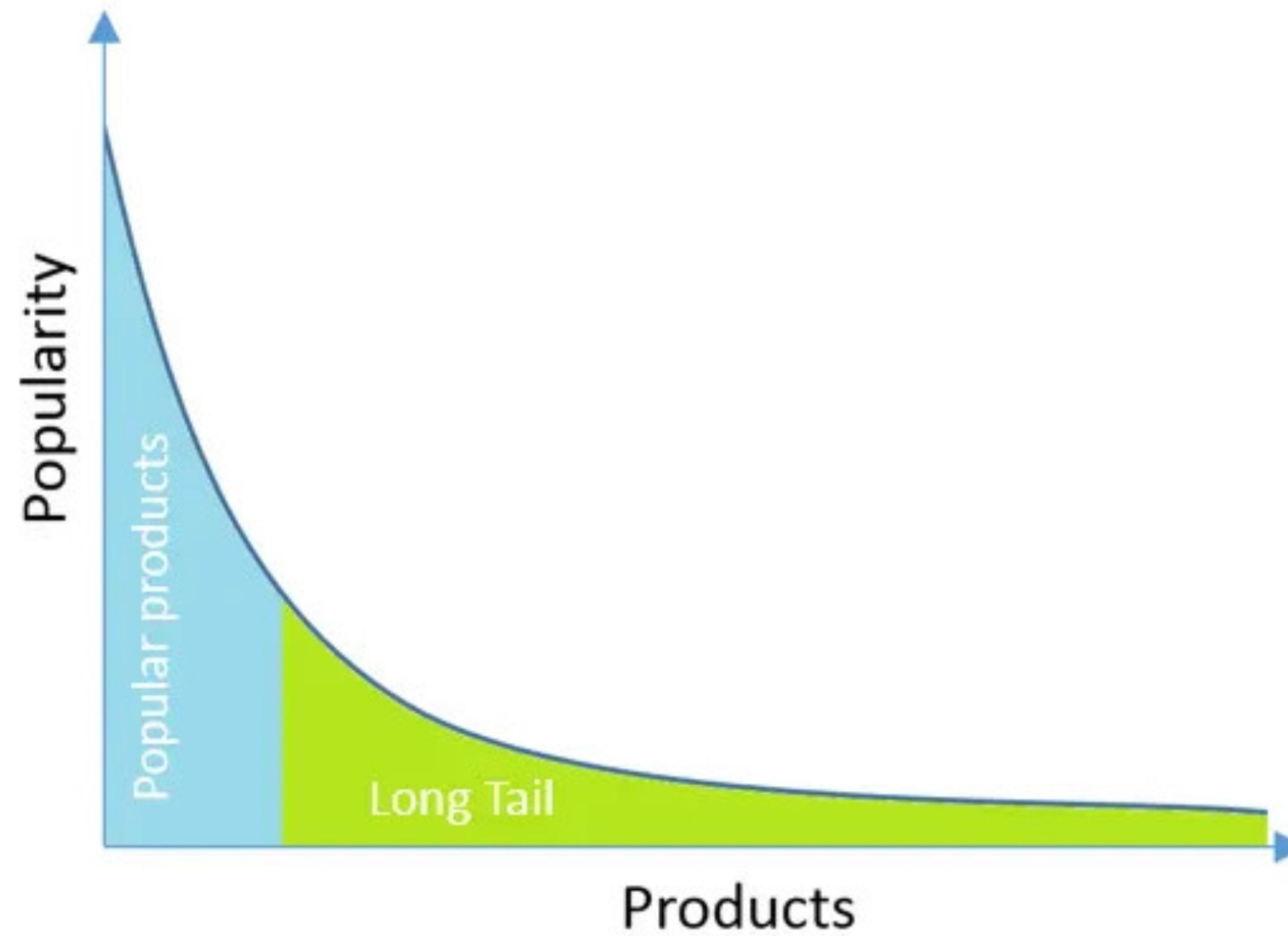
## компании

- увеличение метрик (продажи, gmv, adrev)
- поддержка продавцов

## продавцам

- поиск аудитории
- дополнительные продажи

# Long tail



<https://medium.com/@kyasar.mail/recommender-systems-what-long-tail-tells-91680f10a5b2>

## RS домены

- Стиминговые платформы (Netflix, Spotify, Youtube, Я.Музыка, VK видео :), ...)
- Электронная коммерция (Ozon, WB, Amazon ...)
- Социальные сети (Instagram, X, LinkedIn ...)
- Новостные агрегаторы и медиа (Google News, Дзен ...)
- Путешествия и туризм (Booking.com, Airbnb...)
- Еда и доставка (Лавка, Uber Eats, ...)
- Игры, Искусство и дизайн, Образование и многое другое

# База документов

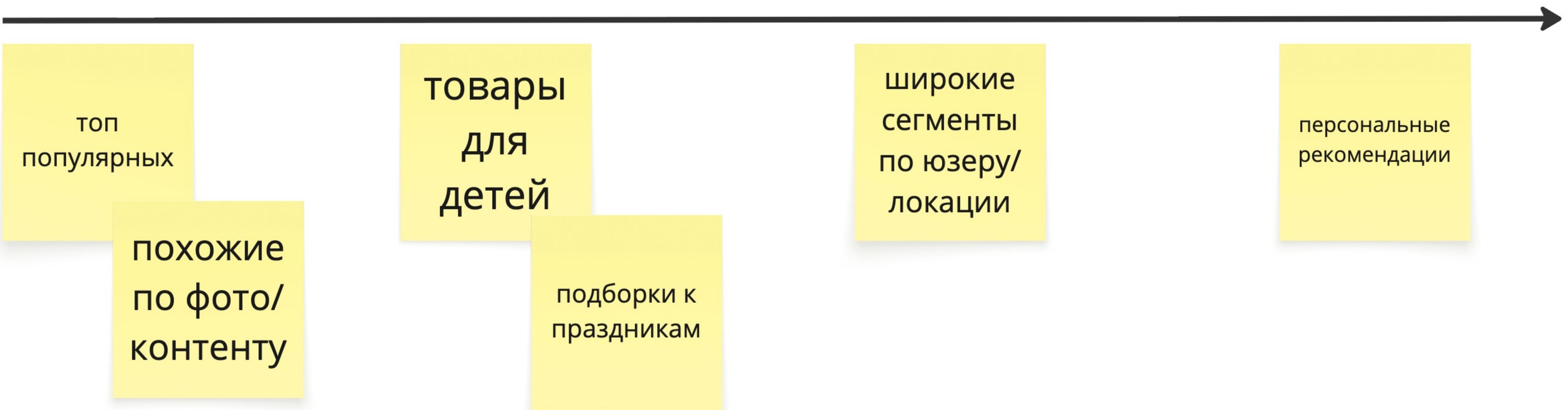
**База документов** — структурированная коллекция объектов, которые система должна уметь анализировать, индексировать, **ранжировать** и выдавать в ответ на запрос пользователя/для формирования рекомендации.

- Текстовый документ (статья, книга, описание товара)
- Веб-страница
- Медиа (видео, трек, фото)
- Карточка товара

item=document

# Персонализация

**Персонализация** — процесс адаптации объектов под индивидуальные предпочтения, интересы, поведение и характеристики конкретного пользователя.

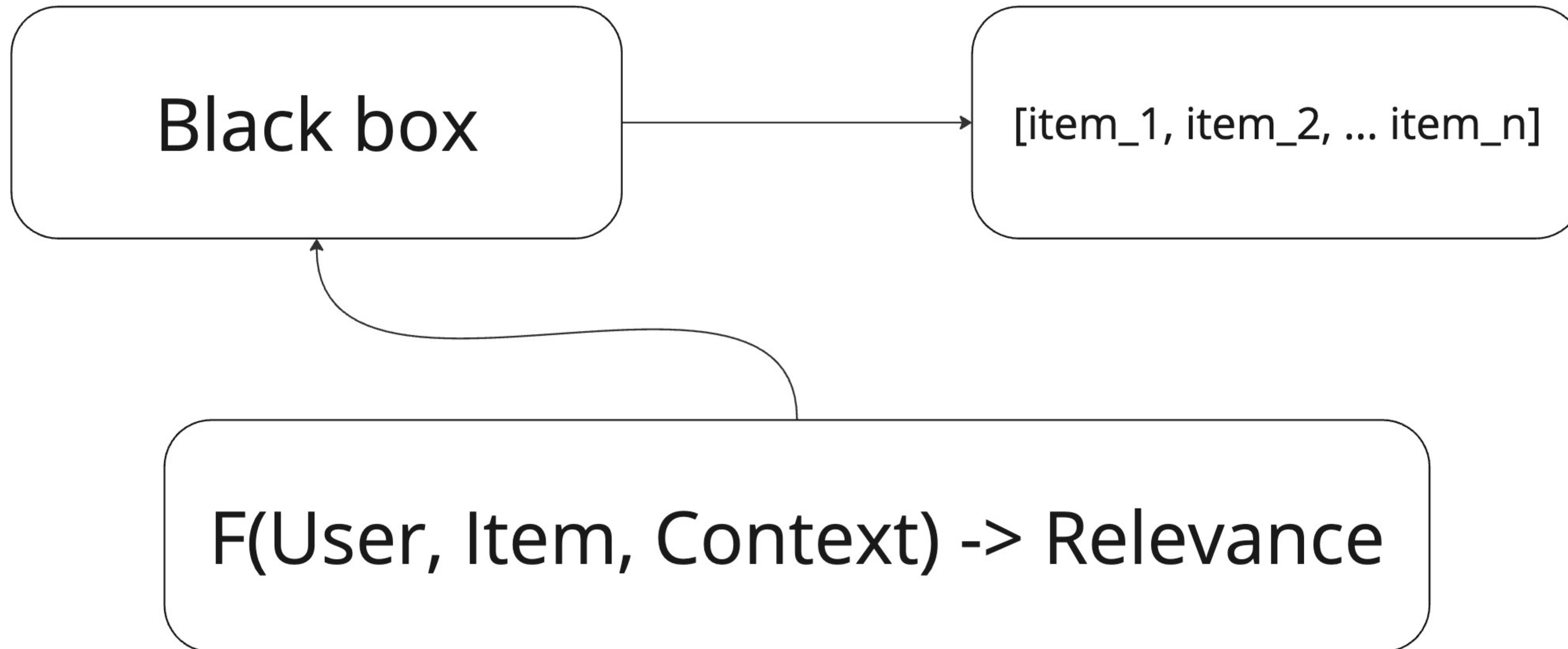


# Постановка задачи



хотим упорядочить  
базу документов  
(размер N) и взять  
топ (n)

# Постановка задачи



# Context

поиск	текстовый запрос
персональные рекомендации	???
похожие/аксессуары/корзина	карточка товара, товары в корзине
общее	Локация, время/дата, продукт

# Релевантность

Для item'ов проставить метку {0, 1}

товар на маркетплейсе/доске объявлений	Факт покупки/клика/добавления в корзину/добавления в избранное
контент	Полный просмотр/прослушивание, просмотр >= X%, лайк, рейтинг > X, share, сохранение

# Релевантность

Для item'ов проставить метку из R

товар на маркетплейсе/доске объявлений	Взвешенная сумма действий с весами, gmv, оценка от асессоров
контент	процент просмотра, общее время просмотра/взаимодействия с контентом, оценка от асессоров

# Релевантность

## Scoring and Ranking

The Phoenix Grok-based transformer model predicts probabilities for multiple engagement types:

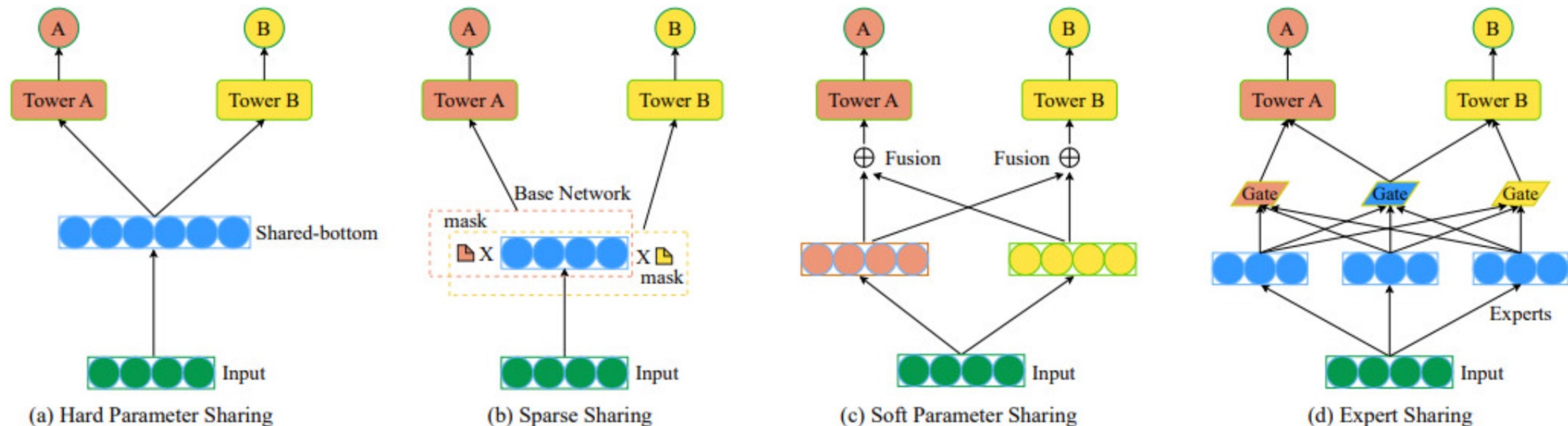
```
Predictions:  
└── P(favorite)  
└── P(reply)  
└── P(repost)  
└── P(quote)  
└── P(click)  
└── P(profile_click)  
└── P(video_view)  
└── P(photo_expand)  
└── P(share)  
└── P(dwell)  
└── P(follow_author)  
└── P(not_interested)  
└── P(block_author)  
└── P(mute_author)  
└── P(report)
```

The **Weighted Scorer** combines these into a final score:

$$\text{Final Score} = \sum (\text{weight}_i \times P(\text{action}_i))$$

<https://github.com/xai-org/x-algorithm?tab=readme-ov-file#scoring-and-ranking>

# Релевантность



# Вопросы

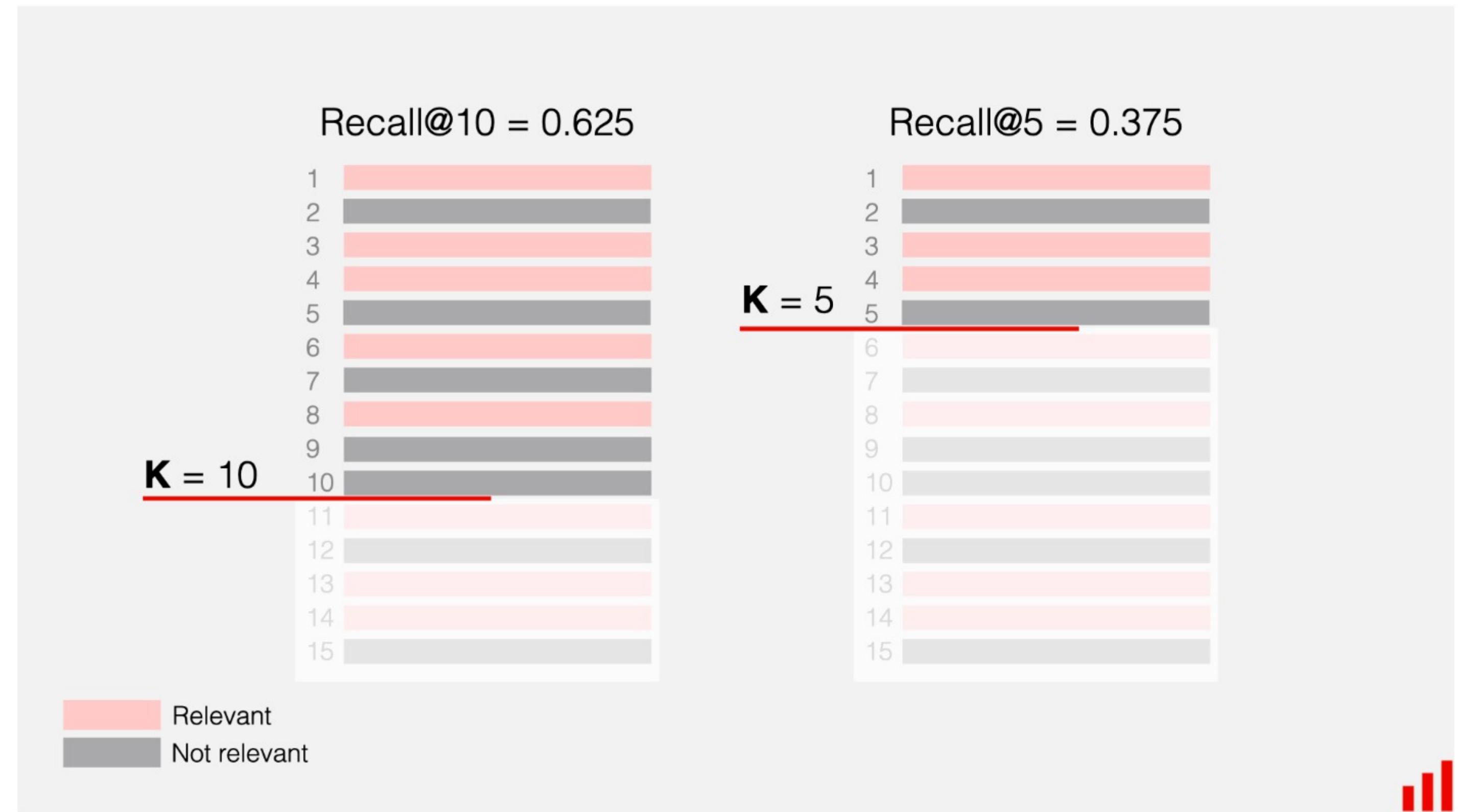
# Метрики, Precision@K

Precision at  $K = \frac{\text{Number of relevant items in } K}{\text{Total number of items in } K}$



# Метрики, Recall@K

Recall at  $K = \frac{\text{Number of relevant items in } K}{\text{Total number of relevant items}}$

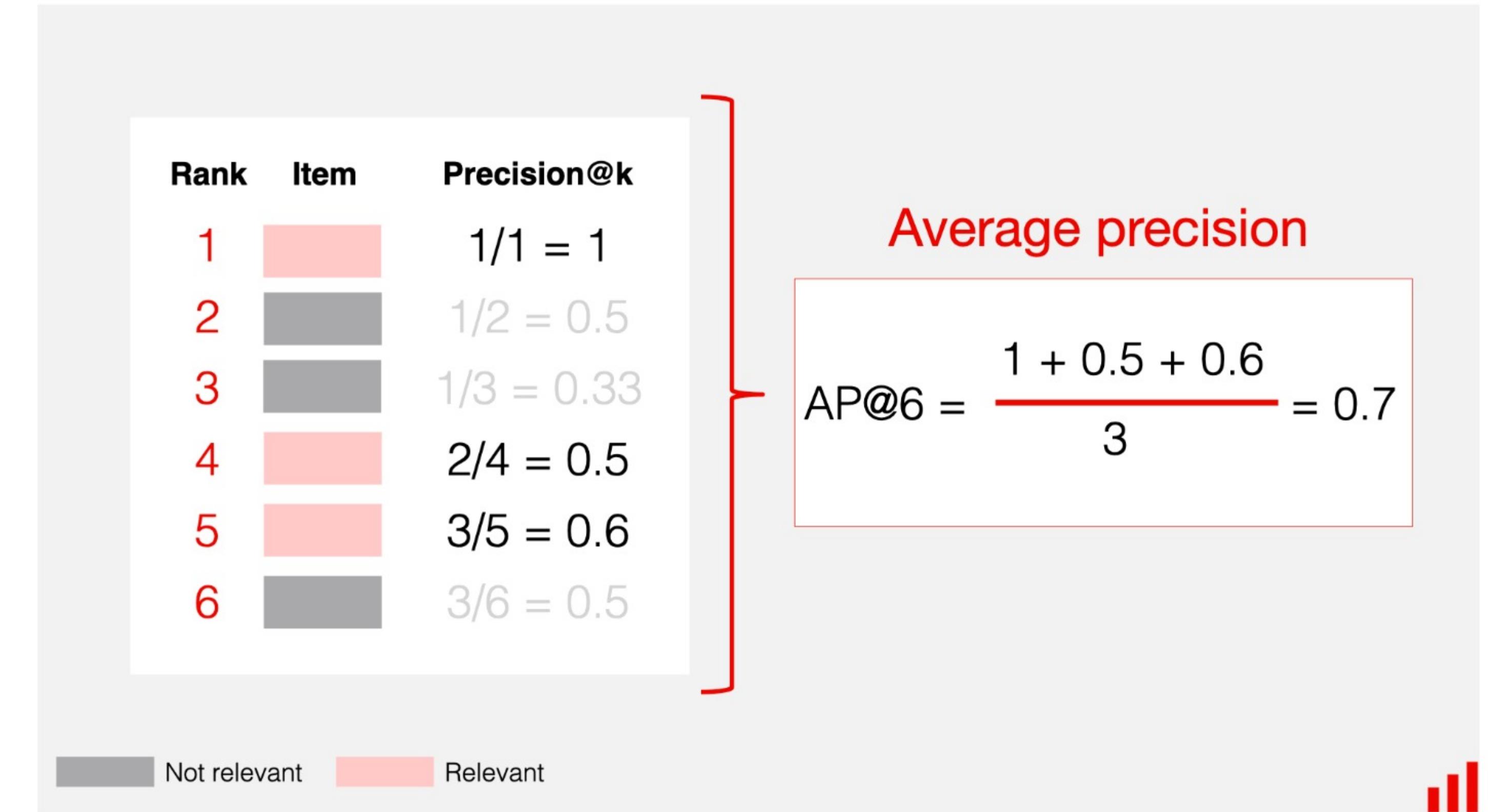


# Метрики, Precision@K и Recall@K

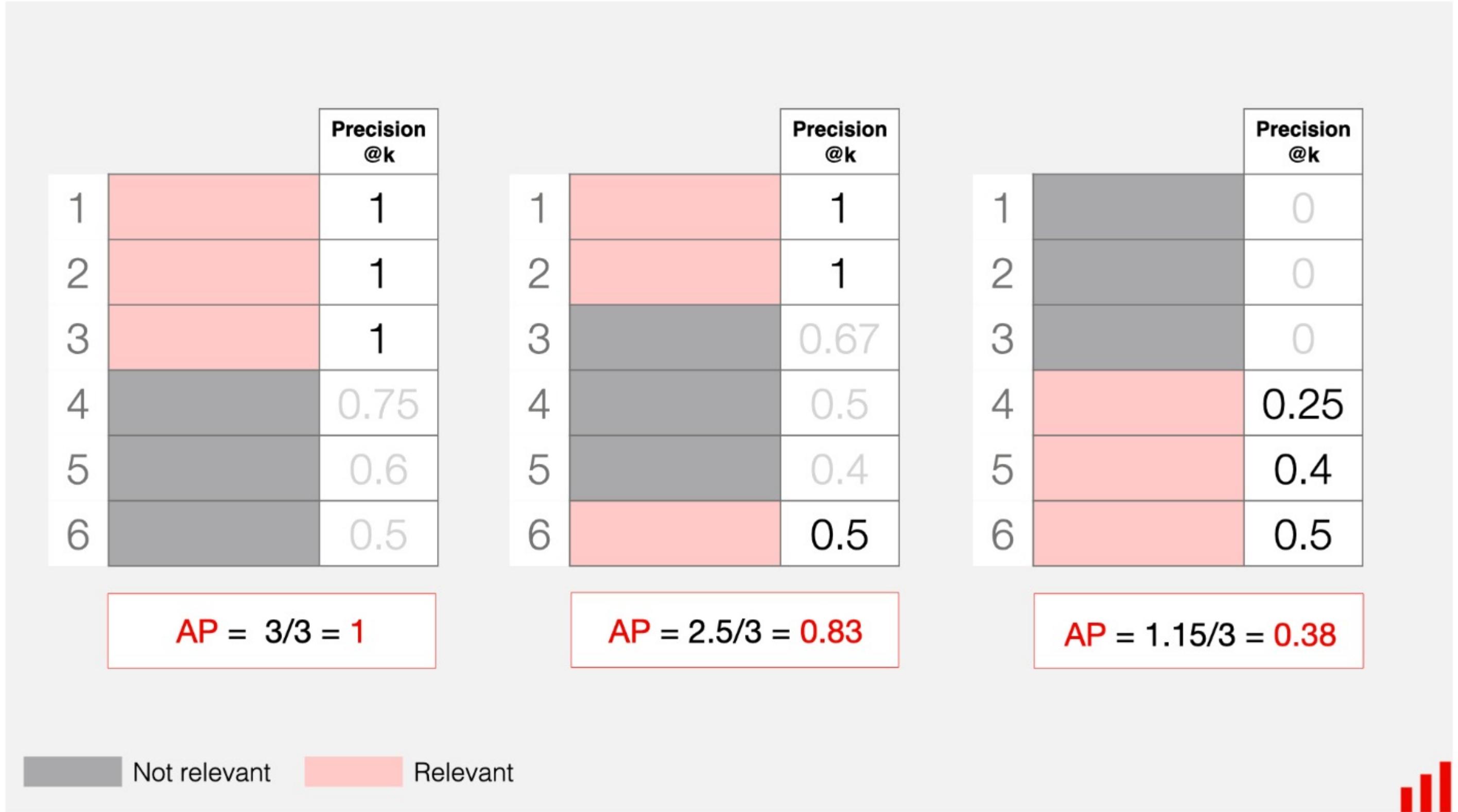
	Precision@K	Recall@K
Плюсы	<ul style="list-style-type: none"><li>интерпретируемость</li><li>легкость вычисления</li></ul>	<ul style="list-style-type: none"><li>интерпретируемость</li><li>используется в задачах с большим количеством релевантных объектов (кандидатогенерация)</li><li>легкость вычисления</li></ul>
минусы	<ul style="list-style-type: none"><li>не учитывает порядок</li><li>есть нюанс с k (об этом на семинаре)</li><li>зависит от k</li></ul>	<ul style="list-style-type: none"><li>не учитывает порядок</li><li>зависит от k</li></ul>

# Метрики, AP@K

$$AP@K = \frac{1}{N} \sum_{k=1}^K Precision(k) \times rel(k)$$



# Метрики, MAP@K



$$\text{MAP}@K = \frac{1}{U} \sum_{u=1}^U \text{AP}@K_u$$



# Метрики, DCG@K

$$\text{DCG}@K = \sum_{k=1}^K \frac{\text{rel}_i}{\log_2(i + 1)}$$



1    2    3    |    4    5

**K = 3**

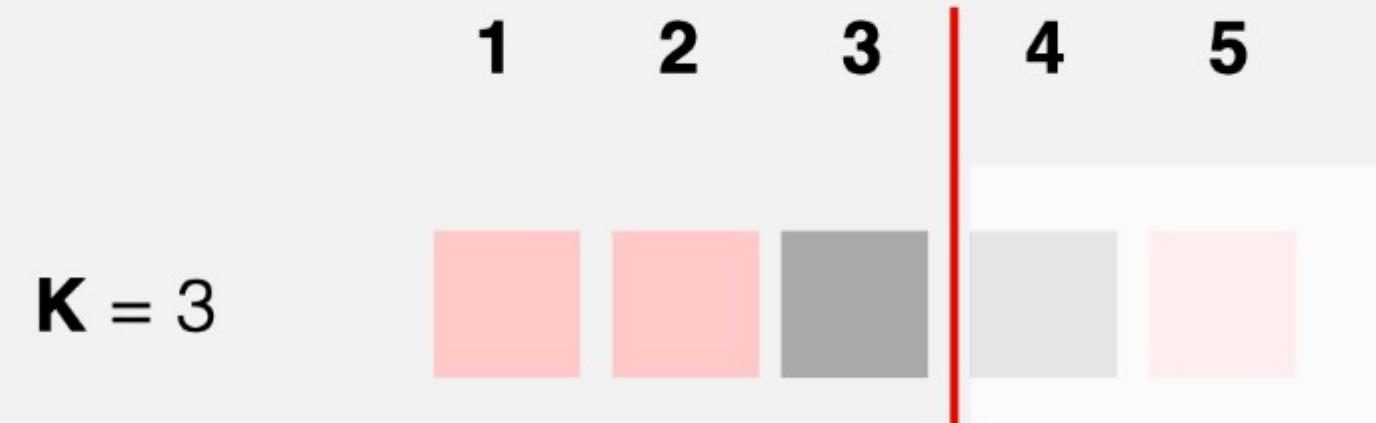


$$\text{DCG}@3 = \frac{1}{\log_2(1 + 1)} + \frac{0}{\log_2(1 + 2)} + \frac{1}{\log_2(1 + 3)}$$

$$\text{DCG}@3 = 1 + 0 + 0.5 = 1.5$$



# Метрики, NDCG@K



$$DCG@3 = \frac{1}{\log_2(1+1)} + \frac{1}{\log_2(1+2)} + \frac{0}{\log_2(1+3)}$$

$$DCG@3 = 1 + \frac{1}{1.585} + 0 \approx 1.63$$

$$NDCG@K = \frac{DCG@K}{IDCG@K}$$



# Метрики, MAP@K и NDCG@K

	MAP@K	NDCG@K
Плюсы	<ul style="list-style-type: none"><li>учёт порядка релевантных элементов</li><li>формально "баланс" между precision и recall</li></ul>	<ul style="list-style-type: none"><li>учёт порядка релевантных элементов</li><li>учет степени релевантности</li></ul>
минусы	<ul style="list-style-type: none"><li>сложность вычисления</li><li>зависит от k</li></ul>	<ul style="list-style-type: none"><li>сложность вычисления</li><li>зависит от k</li></ul>

# Метрики, QueryAuc и Hitrate@K

## QueryAUC

### Classic type

$$\frac{\sum_q \sum_{i,j \in q} \sum I(a_i, a_j) \cdot w_i \cdot w_j}{\sum_q \sum_{i,j \in q} \sum w_i \cdot w_j}$$

The sum is calculated on all pairs of objects  $(i, j)$  such that:

- $t_i = 0$
- $t_j = 1$

$$\bullet I(x, y) = \begin{cases} 0, & x < y \\ 0.5, & x = y \\ 1, & x > y \end{cases}$$

<https://catboost.ai/docs/en/concepts/loss-functions-ranking#QueryAUC>



<https://www.evidentlyai.com/ranking-metrics/evaluating-recommender-systems#hit-rate>

# Вопросы

# Полезные материалы

## КНИГИ

- [Practical Recommender Systems](#)
- [Recommender Systems Handbook](#)
- [Personalized Machine Learning](#)
- [Building Recommendation Systems in Python and JAX](#)

# Полезные материалы

## Тг каналы

- [https://t.me/ods recommender systems](https://t.me/ods_recommender_systems)
- [https://t.me/Recsys IR Travel](https://t.me/Recsys_IR_Travel)
- <https://t.me/info retriever>
- <https://t.me/WazowskiRecommends>
- [https://t.me/knowledge accumulator](https://t.me/knowledge_accumulator)

# Полезные материалы

## Конференции

- ACM Recommender Systems
- ECIR (European Conference on Information Retrieval)
- SIGIR (Conference on Research and Development in Information Retrieval)
- WSDM (ACM International Conference on Web Search and Data Mining)
- WWW (The Web Conference)