

Рекомендательные системы

ШАД

Весна 2025

Лекция 2:

**Качество рекомендательных
систем**

Ресар первого семинара

- **Precision@10**
 - **Jaccard: 0.0288**
 - **NPMI: 0.0587**
 - **SVD: 0.0413**

Как оценить рексистему?

- **offline** vs **online** оценка
- **offline** -метрики
- **аспекты** качества
- **online** -метрики
- **ключевые проблемы**

Метрики: Offline vs Online

	Где берем Ground Truth	Связь с бизнесом
Offline	<p>из исторических данных</p> <ul style="list-style-type: none">- быстро: посчитать метрики можно сразу после обучения- неточно: разметка есть не для всех объектов	<ul style="list-style-type: none">- коррелирующая: рост метрик не всегда означает + для бизнеса- проверка “плохих” моделей не влияет на пользователей
Online	<p>из взаимодействия пользователей с новой моделью в проде (AB)</p> <ul style="list-style-type: none">- долго: нужно дождаться места в AB, раскатить, собрать логи- точно: все объекты размечены	<ul style="list-style-type: none">- прямая: метрики напрямую отражают бизнес-цели- “плохие” модели в проде приводят к потере денег и пользователей

Метрики: Offline vs Online

1. **Определяем метрики :**
 цель бизнеса →
 online →
 offline
2. **По offline: отбираем парочку моделей**
3. **Катим их в АВ**
4. **По online:**
 - **Все супер → раскатываем лучшую**
 - **Все не супер → goto п.1.**



Offline-метрики

$Q = \{q_i\}, i = \overline{1, N_Q}$ – множество запросов/пользователей

$D = \{d_i\}, i = \overline{1, N_D}$ – множество документов

$X = \{x_i\}$ – признаковые описания (q_i, d_i)

$Y = \{y_i\}$ – истинная разметка релевантности

$\hat{Y} = \{\hat{y}_i\}$ – предсказания модели

$rel_i = I(y_i = 1)$

Offline-метрики: что-то из классики

- Метрики регрессии.

Например,

$$RMSE_{Q_j} = \sqrt{\frac{1}{N_D} \sum_{i=1}^{N_D} (y_i - \hat{y}_i)^2}$$

- Не учитывают позиции

айтемов в выдаче y	5	rec1	\hat{y}	rec2
10	5	10	10	10
8	4	8	8	8
6	3	6	6	6
3	2	4	4	4
$RMSE_{Q_j}$	7.1		0	

- Метрики классификации

- Вообще тоже

- НО: есть AUC

$$AUC = P(y_+ > y_-)$$

- НО: пользователь видит только топ

- → AUC@k

Offline-метрики: MAP@k

$$Precision = \frac{1}{\hat{N}_{rel}} \sum_{v=1}^N rel_v$$



- ограничиваемся тем, что видел пользователь - **Precision@k**

$$Precision@k = \frac{1}{k} \sum_{v=1}^k rel_v$$

- усредняем по выдаче - **AveragePrecision@k**

$$AP@k = \frac{1}{N_{rel@k}} \sum_{j=1}^k P@j \cdot rel_j$$

- усредняем по всем запросам-
MeanAveragePrecision@k

k	Url	reli	P@k	P@k*reli	AP@k
1	 Карбонара	1	1	1	1
2	 Приготовить пиццу	0	0.5	0	0.5
3	 Болоньезе	1	0.67	0.67	0.835
4	ζυμαρικά με μανιτάρια	? → 1	0.75	0.75	0.807
5	 Паста с рыбой	1	0.8	0.8	0.805
6	 Макарошки	1	0.83	0.83	0.81

$$MAP@k = \frac{1}{N_Q} \sum_{i=1}^N (AP@k)_i$$

Offline-метрики: Recall@k, CG@k


$$Recall = \frac{1}{N_{rel}} \sum_{v=1}^N rel_v$$

- точное N_{rel} неизвестно
- ограничиваемся тем, что видел пользователь

$$Recall@k = \frac{1}{N_{rel}} \sum_{v=1}^k rel_v$$

- общая “полезность” рекомендаций: Cumulative Gain

$$CG@k = \sum_{v=1}^k y_i$$

k	Url	rel _i	R@k
1	 Карбонара	1	0.2
2	 Приготовить пиццу	0	0.2
3	 Болоньезе	1	0.4
4	ζυμαρικά με μανιτάρια	? → 1	0.6
5	 Паста с рыбой	1	0.8
6	 Макарошки	1	1

Offline-метрики: NDCG@k

- **Normalized Discounted Cumulative Gain**

$$DCG@k = \sum_{i=1}^k \frac{y_i}{\log_2(i+1)}$$

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

i	log(i+1)			IDCG-DCG
		Ideal	rec	
1	1	10	8	(10 + 5.06) - (8 + 6.33) = 0.73
2	1.58	8	10	
3	2	8	6	(4 + 2.59) - (3 + 3.45) = 0.14
4	2.32	6	8	
DCG		21.65	20.78	
NDCG		1	0.96	

Offline-метрики: pFound@k

- **вероятность психануть и перестать искать**

$$pBreak = 0.15$$

- **вероятность, что в i -м документе что-то полезное**

$$y_i$$

- **вероятность долистать до i позиции**

$$pLook_1 = 1$$

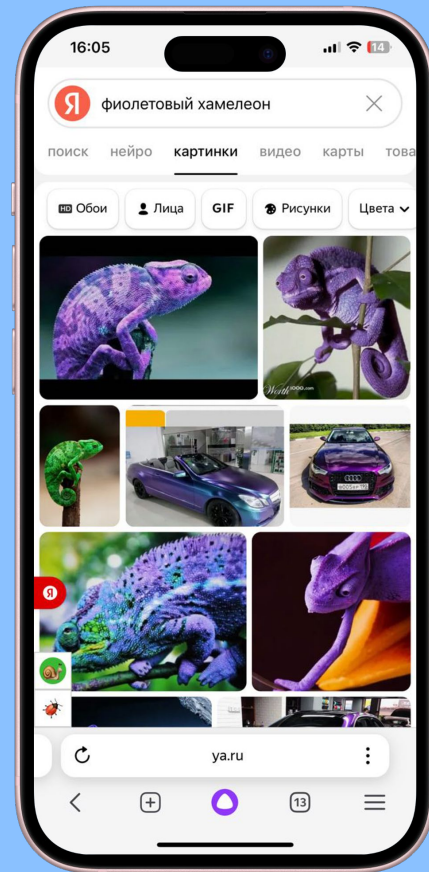
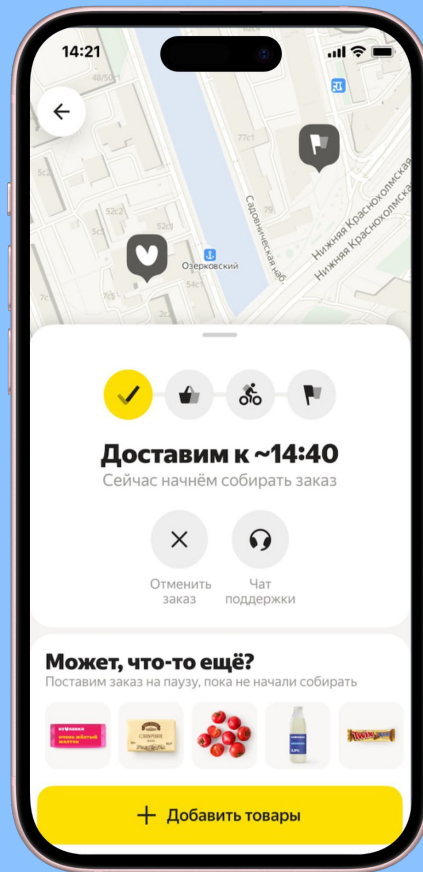
$$pLook_i = pLook_{i-1} (1 - y_{i-1}) (1 - pBreak)$$

- **вероятность найти ответ за первые k элементов**

$$pFound@k = \sum_{i=1}^k pLook_i \cdot y_i$$

Offline-метрики: HR@k

- **HitRate@k** -
хотя бы один
позитив в топ-k
 $HR@k = 1(\text{rel}_i \leq k)$



Offline-метрики: MRR

- **Mean Reciprocal Rank (MRR) – первый позитив**

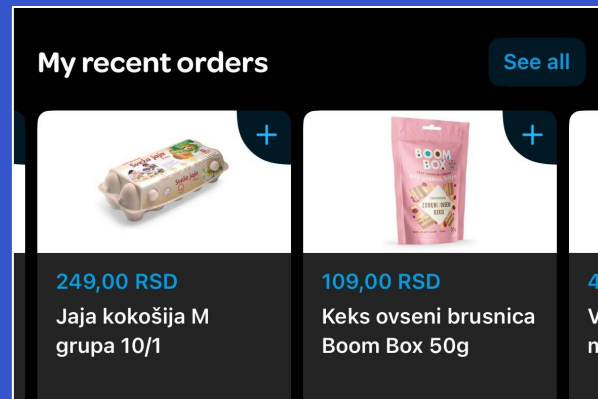
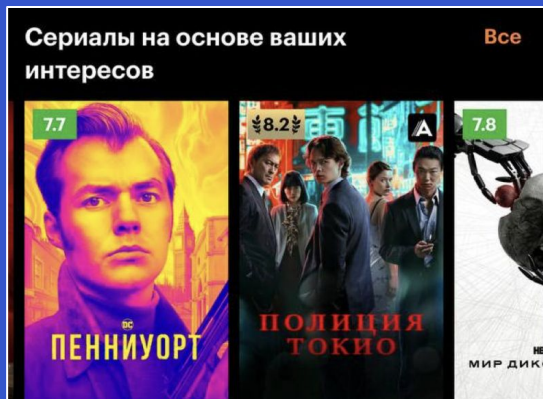
$$MRR = \frac{1}{firstRelevantPos}$$



Аспекты качества

Аспекты качества

- **Relevance & Conversion**
 - рекомендации должны быть **релевантными**
 - конвертироваться в **целевое действие**
 - **Precision@k, CTR**



Аспекты качества

- **Diversity**

- **рекомендации должны быть достаточно разнообразными**
- **category coverage**

$$\text{CategoryCoverage@}k = \frac{|\text{UniqueCategories@}k|}{|\text{Categories@}k|}$$

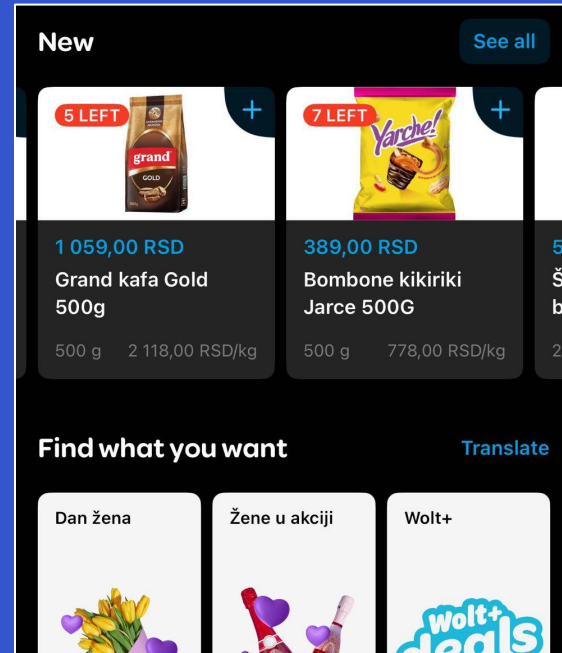
- **intra-list similarity**

$$ILS@k = \frac{2}{k(k-1)} \sum_{i=1}^k \sum_{j=i+1}^k \text{sim}(i, j)$$

Аспекты качества

- **Novelty**
 - способность рекомендательной системы предлагать **новые** или **малоизвестные** айтемы

$$Novelty@k = \frac{1}{k} \sum_{i=1}^k (1 - Popularity(i))$$



Аспекты качества

Discover y

User Discovery — насколько часто пользователи получают новые релевантные айтемы.

$$UserDiscovery = \frac{|I_{rec} \cap I_{new}|}{|I_{rec}|}$$

Item Discovery — насколько часто товары находят новых заинтересованных пользователей.

$$ItemDiscovery = \frac{|U_{rec} \cap U_{new}|}{|U_{rec}|}$$

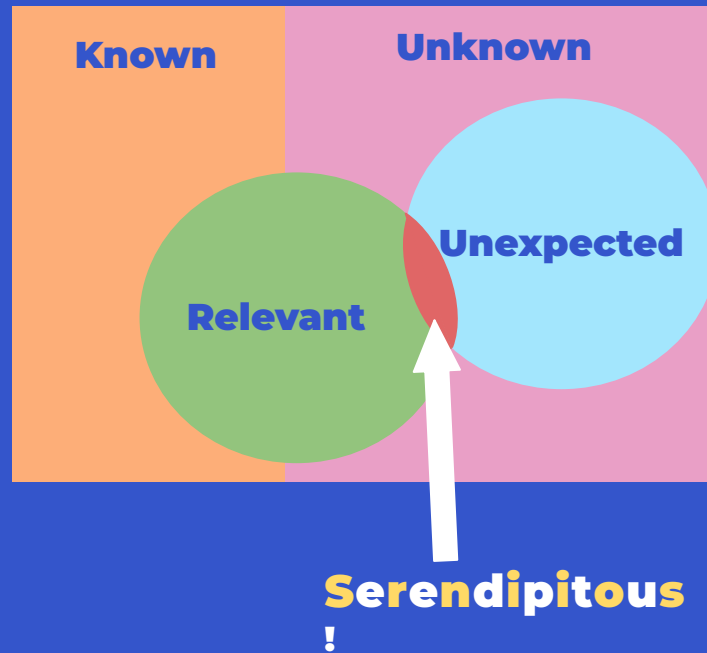


Аспекты качества

- **Serendipity**
 - способность системы рекомендовать **неожиданные**, но релевантные товары, которые пользователь **не искал** активно и **не нашел бы** самостоятельно
 - **WOW-эффект**

$$S@k = \frac{1}{k} \sum_{i=1}^k y_i (1 - P_{expected}(i))$$

$$S@k = \frac{1}{k} \sum_{i=1}^k y_i \left(1 - \max_{j \in R_{prim}} \text{sim}(i, j) \right)$$



Аспекты качества

- Retention

- рекомендации должны заставлять пользователей **возвращаться** в сервис: совершать **повторные** покупки / **пересматривать** или находить **новый контент** / ...
- **Retention Rate**

$$RR_{DayN} = \frac{|UsersDayN| - |NewUsers0N|}{|UsersDay0|}$$

$$ChurnRate_{DayN} = 1 - RR_{DayN}$$

- **Repeated Purchases Rate**

$$RPR = \frac{|RepeatedPurchases|}{|AllPurchases|}$$



Online-метрики

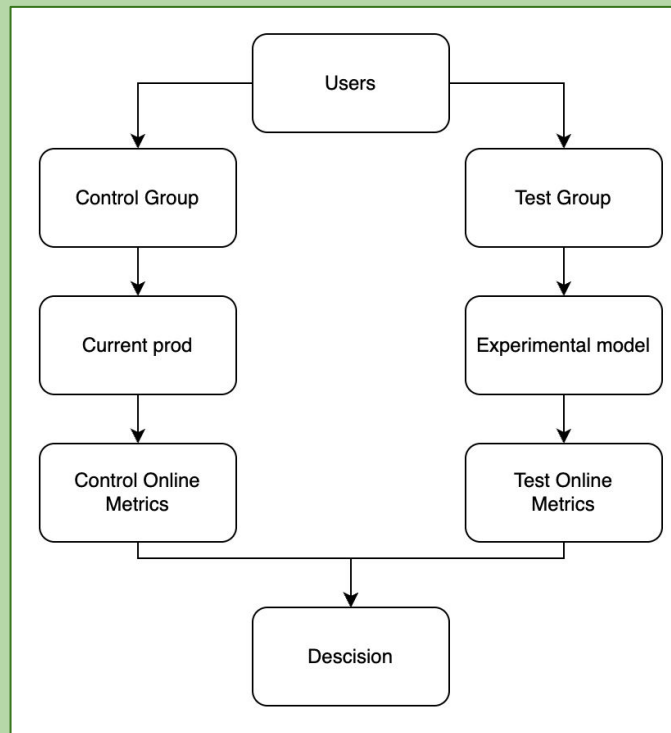
Online -метрики: АВ-тестирование

зачем:

- для определения наиболее эффективной версии продукта

как:





- сформулировать **гипотезы**
 - выбрать целевые **метрики**
 - и подумать про **метрики-антагонисты**
 - определить **сетап** эксперимента: длительность, размеры выборок, MDE, ...
 - **запустить** , подождать
- посчитать **метрики**, **p-value**, принять решение

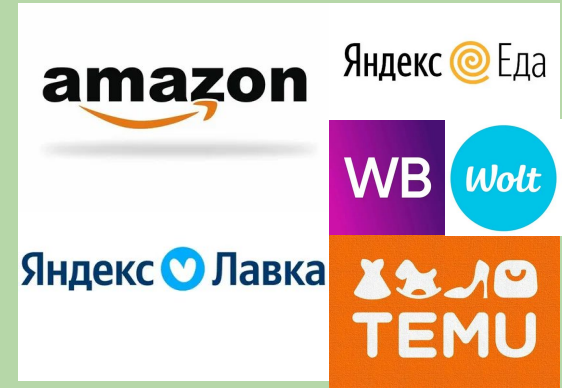


Online -метрики: база

- **релевантны ли рекомендации в моменте?**
 - **Explicit Feedback Count (Likes, Cart-updates, ...)**
 - **Click-Through Rate (CTR)**
 - **Time to First Interaction (TTFI)**
 - **Bounce Rate**
- **не скучно ли?**
 - **Diversity Score**
 - **Discovery Score**
- **а еще придете?**
 - **Retention Rate**
 - **Daily/Weekly/... Active Users (DAU/MAU/...)**
 - **Life-time value (LTV)**

Online -метрики: E-com

-  Conversion to Purchase
- 
 - Average Revenue per User (ARPU)
 - Average Order Value (AOV)
 - Gross Merchandise Volume (GMV)
-  Time to Purchase
-  Cart Abandonment Rate



How to **накосячить** :

- **растим** : Conversion to Purchase
- **получаем** : в топе товары по скидке / дешевые товары / товары из истории
- **роняем** : все денежные метрики
- **растим** : AOV
- **получаем** : в топе дорогие товары / товары, которые пользователь итак берет всегда
- **роняем** : DAU

Online -метрики: Новостные платформы и соцсети

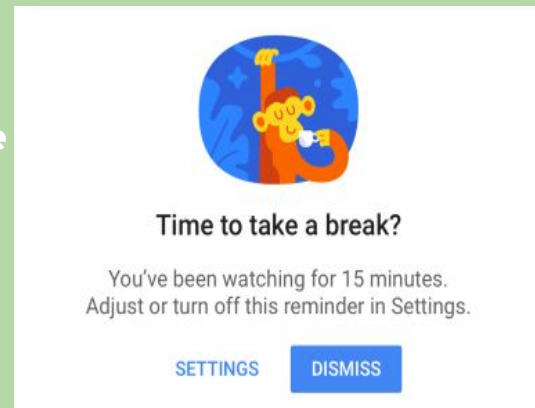
-  Dwell Time
-  Scroll Depth
-  Read Depth
-  Popularity Score
-  Novelty Score
-  Sentiment Balance

ЭТО **СКРЫВАЛИ** ОТ НАС ВСЮ
ЛЕКЦИЮ!!! **CTR РАСТЕТ НА
ГЛАЗАХ!** ПРОСТО
ПОРЕКОМЕНДУЙТЕ...

- **самый кликбейтный
контент**
- **и уроните все что можно -
Dwell Time, Read Depth,
Retention, ...**

Online -метрики: Стриминговые сервисы

- 
 - Session Length/Total Watch/Listening Time
 - Average Watch/Listening Time
 - Streaming Time
-  Completion Rate
-  Skip Rate



How to накосячить:

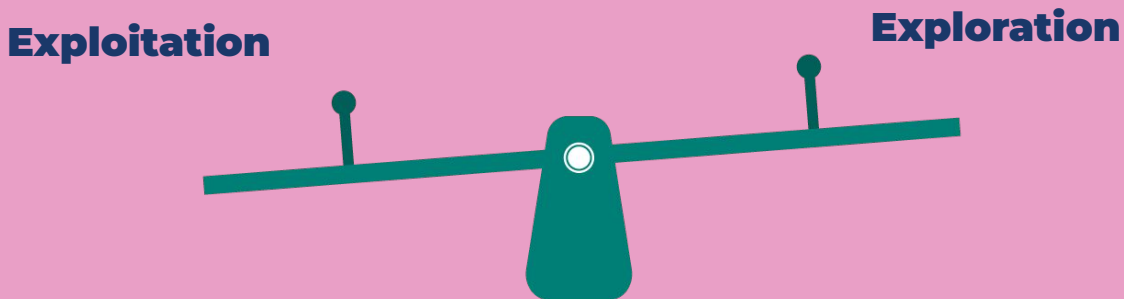
- **растим** : Completion Rate
- **получаем** : в топе короткие видео
- **роняем** : Session Length

How to раскосячить:

- показать “заботу о пользователе”:
Autoplay VS “Time to break”

Ключевые проблемы

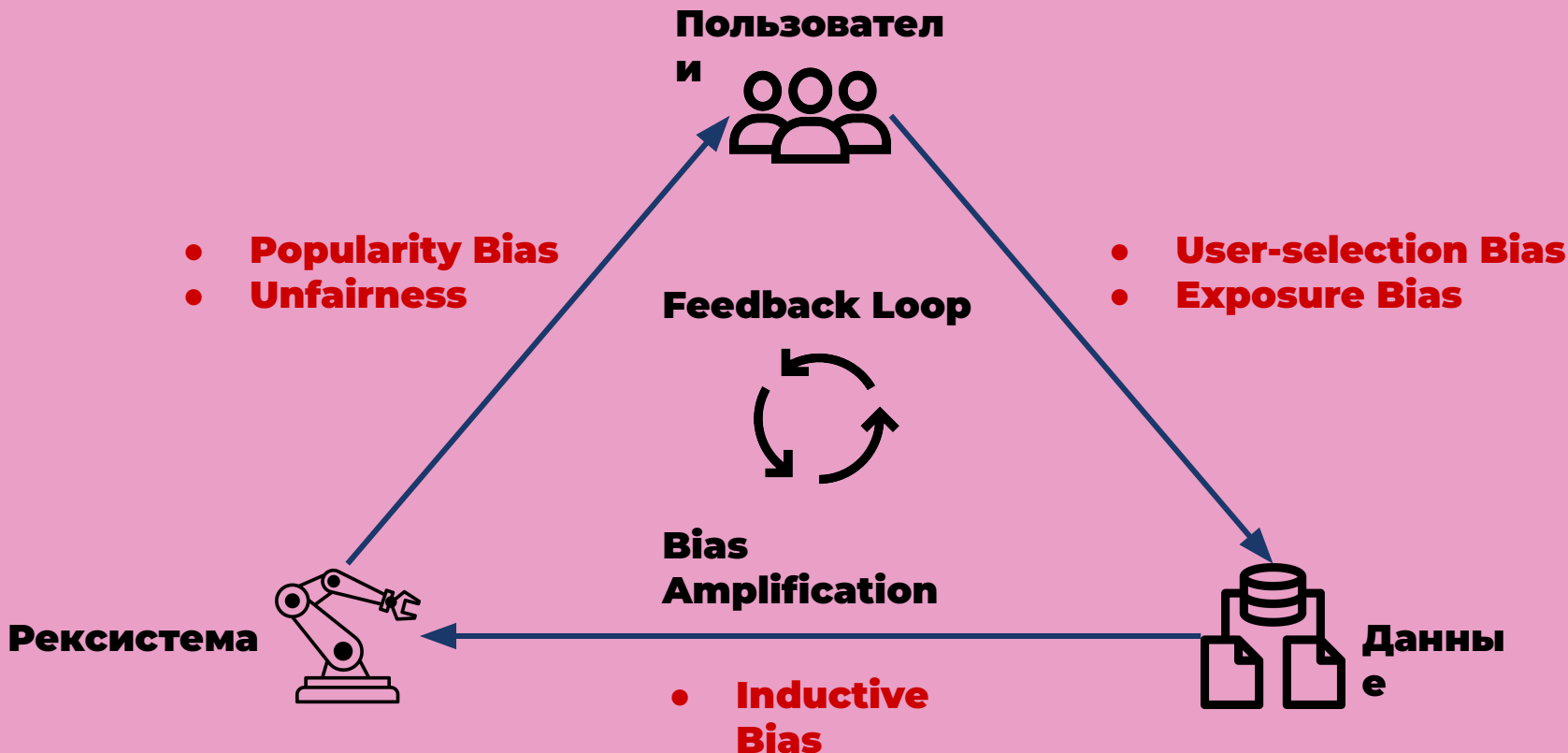
Ключевые проблемы: Exploration-Exploitation Trade-Off



- **известные релевантные** айтемы -> высокое **вовлечение в моменте**
- **пузыри интересов,** скучно

- **слишком много нерелевантных айтемов**
- **исследование** новых интересов -> выше **long-term влияние**

Ключевые проблемы: Feedback Loop



Ключевые проблемы: Холодный старт

- **Новые пользователи**
 - Рекомендовать популярное
 - Гибридные модели (демография/внешние данные)
 - Онбординг
- **Новые айтемы**
 - Бустинг новинок
 - Гибридные модели (контент)
 - RL
- **Новые платформы**
 - Предобученные модели
 - Внешние данные



Спасибо!

**Алена Зайцева,
Группа рекомендаций Лавки,
Белград 2025**