



# Retrieval at Scale

Machine Learning Research @ Allegro

Aleksandra Osowska-Kurczab  
Jacek Szczerbiński

Research Engineer  
Junior Research Engineer

**allegro**

# We are the biggest e-commerce player of European origin

The platform of growth opportunities

**22M**

Customers

Employer of choice

**89%**

employees recommend working at Allegro

The **biggest** and still growing e-commerce team in CEE

**+7,200**

People

Customer satisfaction

**NPS 77.9**

Best in sector  
Apple < Netflix < **Allegro** < Tesla



**allegro**

**allegro pay.**

**one box**  
by allegro

**eBilet.pl**

**CENEOL**

# Agenda

1. Machine Learning at Allegro
2. Introduction to Retrieval
3. Retrieval at scale: production cases at Allegro
  - a. Recommendations
  - b. Visual Search
  - c. Learning to Rank
4. Summary

# allegro

---

Machine Learning at Allegro

## DZIAŁY KORZYSTCI

- Elektronika
- Moda
- Dom i Ogród
- Supermarket
- Dziecko
- Uroda
- Zdrowie
- Kultura i rozrywka
- Sport i turystyka
- Motoryzacja
- Nieruchomości
- Kolekcje i sztuka
- Firma i usługi
- eBilet.pl
- Allegro Lokalnie

## WSZYSTKIE KATEGORIE

Okazje dnia kończą się za:

11 : 26 : 42

SUPER CENA



- [xiaomi](#)  
Telefony i Akcesoria > Xiaomi
- [xiaomi](#)  
RTV i AGD > [Telewizory](#)
- [xiaomi mi band 6](#)
- [xiaomi mi band 5](#)
- [xiaomi redmi note 10 pro](#)
- [xiaomi redmi 9](#)
- [smartwatch xiaomi](#)
- [Smartfon Xiaomi Redmi Note 10 Pro 6 GB / 128 GB szary](#)
- [Smartfon Xiaomi Redmi 9A 2 GB / 32 GB szary](#)
- [Smartfon Xiaomi Redmi Note 11 6/128 GB szary](#)

 szukaj też w opisach ofert

Jak kupować na Allegro?



Allegro Smart!



HP Instant Ink



Wiosną wybór kwitnie



Wiosenny ogród



Pierwszy

Cześć Jacek!

Dokończ aktywację Allegro Pay. Kupuj, sprawdzaj i płac później lub w mniejszych splatach.

Dostawa z kurierem z Allegro Smart!

Dodaj kolejną osobę do Allegro Family

Okazja wybrana dla Ciebie



ZOBACZ WIĘCEJ OFERT

Gry video idealne na prezent



99,00 zł SMART!

SMERFY MISJA ZŁOŚĆ SMERFASTYCZNA PL NINTENDO  
zapłać później z   
36 osób kupiło

109,00 zł SMART!

JUST DANCE 2018 MOVE PS3 Nowa gra - PŁYTA - FOLIA  
zapłać później z 



Kategorie ▾

Okazje do -70%

Promocje z Monetami

Allegro Lokalnie

Premiery

Allegro Inspiruje

szukasz „xiaomi poco x3” (28 614 ofert)

★ OBSERWUJ WYSZUKIWANIE

## Podkategorie

## popularne

Elektronika	
Smartfony i telefony komórkowe	85
Xiaomi	85
Akcesoria GSM	28300
Etui i pokrowce	21177

## wszystkie

Telefony i Akcesoria	28472
RTV i AGD	89
Motoryzacja	27
Komputery	21

WIECEI

## Filtry z kategorii

Cena (zł)

Oryginalne akcesoria i części BMW  
AUTORYZOWANYCH DEALERÓW

SPRAWDZ &gt;

ROZŁĄCZ TE SAME OFERTY

sortowanie

trafność: największa

1

z 100

&gt;



Oferta sponsorowana ⓘ

Plecki POCO X3 bezbarwny

4,42 ★★★★★ 12 ocen produktu

Stan: Nowy

11,90 zł SMART



Oferta sponsorowana ⓘ

Smartfon Xiaomi Pocophone F3 8 GB / 256 GB biały

4,88 ★★★★★ 231 ocen produktu

Kolor: biały Przekątna ekranu: 6.67" Wbudowana pamięć: 256 GB Pamięć RAM: 8 GB

1 849,00 zł SMART

Kategorie



Okazje do -70%

Promocje z Monetami

Allegro Lokalnie

Premiery

Allegro Inspiruje

szukasz „xiaomi poco x3” (85 ofert)

OBSERWUJ WYSZUKIWANIE

Allegro - Elektronika - Telefony i Akcesoria - Smartfony i telefony komórkowe

## Podkategorie

cofnij do Telefony i Akcesoria

Xiaomi

85

## Filtry z kategorii

## Cena (zł)

 poniżej 225 zł 225 zł do 600 zł 600 zł do 1250 zł 1250 zł do 1750 zł powyżej 1750 zł

od

- do

Kup Galaxy Z Fold3 | Flip3 5G,  
a smartwatch Galaxy Watch4 (BT) oraz  
słuchawki Galaxy Buds2 otrzymasz w prezencie.

Promocja trwa do 04.04 do 01.05.2022 r. Rysik S Pen Edition jest sprzedawany oddzielnie. Szczegóły w regulaminie.**ROZŁĄCZ TE SAME OFERTY**

sortowanie

trafność: największa



1

z 1

Szukasz: xiaomi poco x3



Oferta sponsorowana

Smartfon Xiaomi PocoPhone F3 8 GB / 256 GB biały

4,88 ★★★★★ 231 ocen produktu

Kolor: biały Przekątna ekranu: 6,67" Wbudowana pamięć: 256 GB Pamięć RAM: 8 GB

**1 849,00 zł SMART!**92,45 zł x 20 rat **raty zero** szczegóły w ofercie

dostawa pojutrze



Kategorie



Okazje do -70%

Promocje z Monetami

Allegro Lokalnie

Premiery

Allegro Inspiruje

Allegro - Elektronika - Telefony i Akcesoria - Smartfony i telefony komórkowe - Xiaomi - POCO F3 - Smartfon Xiaomi Pocophone F3 8 GB / 256 GB biały

## Smartfon Xiaomi Pocophone F3 8 GB / 256 GB biały

4,88 ★★★★★ 231 ocen i 67 recenzji



od Oficjalny sklep Xiaomi



poleca 99,4%

### Smartfon Xiaomi POCO F3 5G 8/256GB Arctic White

**1849,00 zł SMART**92,45 zł x 20 rat **raty zero**

12 Monet

15 osób kupiło

Przewidywana dostawa: pojutrze u Ciebie

Dostawa za darmo

DOSTAWA

Zwrot za darmo  
14 dni na odstąpienie od umowy

ZWROTY

Wygodne płatności  
Allegro Pay, raty zero

PŁATNOŚCI

Kolor/wzór



Faktura:

Wystawiam fakturę VAT

## Dane podstawowe

Stan: **Nowy**  
Kod producenta: **POCO F3**

Typ: **Smartfon**

## Parametry fizyczne

Głębokość produktu: **7.8 mm**

## POPATRZ W OCZY BESTII

### Najwyraźniejszy jak dotąd wyświetlacz POCO

- Wysoka częstotliwość odświeżania 120Hz
- Niezależnie od tego, czy czytasz, czy grasz w gry, ekran jest płynniejszy niż kiedykolwiek.
- Obraz odświeża do 120 klatek na sekundę. Ciesz się niespotykaną gładkością!
- Odświeżanie dotyku 360Hz
- Częstotliwość próbkowania dotykowego została zwiększona do 360Hz oraz bardziej dopracowany algorytm dotykowy
- Telefon natychmiastowo reaguje na dotknięcia opuszków palców, dzięki czemu gra i zwykła praca jest wyjątkowo płynna.

## Wyświetlacz AMOLED E4

- Dzieki nowej generacji wysokiej klasy materiału ekranu, zużycie energii jest zmniejszone o 15%, dzięki czemu bateria działa dłużej.



**1 849,00 zł**

**SMART!**

Smartfon Xiaomi  
POCO F3 5G  
8/256GB Night Black

**1 499,00 zł**

**SMART!**

Smartfon Xiaomi  
POCO X4 Pro 5G  
6/128GB Laser Blue

**1 799,00 zł**

**SMART!**

Smartfon Xiaomi  
Redmi Note 11 Pro  
5G 8/128GB Grey

**1 193,49 zł**

**SMART!**

Smartfon Xiaomi  
POCO M4 Pro 5G  
6/128GB POCO

**1 599,00 zł**

**SMART!**

Smartfon Xiaomi  
Redmi Note 11 Pro  
6/128GB Grey

**1 699,00 zł**

**SMART!**

Smartfon Xiaomi 11  
Lite 5G NE 8/128GB  
Peach Pink

**2 599,00 zł**

**SMART!**

Smartfon Xiaomi 11T  
Pro 8/256GB  
Moonlight White

**1 099,00**

**SMART!**

Smartfon Xi.  
Redmi Note  
4/128GB Bl



Rozmawiasz z:  
[anna\\_zzz](#)

9:46 Twoja przesyłka „Kocyk bawełniany HIT ma...” + 9 innych jest już w drodze. Przewidywany czas dostawy: czwartek, 3 września.

30 sierpnia 2020

Dzień dobry, czy moja paczka dotrze przed weekendem?

09:13



Cześć, tu chatbot Allina.  
Nie wysyłam Twojej wiadomości, bo mogę Ci od razu powiedzieć, że przesyłka „Kocyk bawełniany HIT ma...” + 9 innych jest już w drodze i powinna dotrzeć w **czwartek, 3 września**.

Czy nadal chcesz wysłać swoją wiadomość do sprzedającego?

[TAK, CHCĘ](#) [NIE CHCĘ](#)

napisz wiadomość...  

Nasza sztuczna inteligencja Allina wylapuje słowa kluczowe i może na ich podstawie odpowidać na pytania. Jeśli sobie tego nie życzysz, możesz [wyłączyć Allinę](#).

mimi3r1@163.com  
Do: allegro.kontakt 

Client:79381628, kamaagnes

We have been unable to contact the buyer during the discussion, please help, thank you.

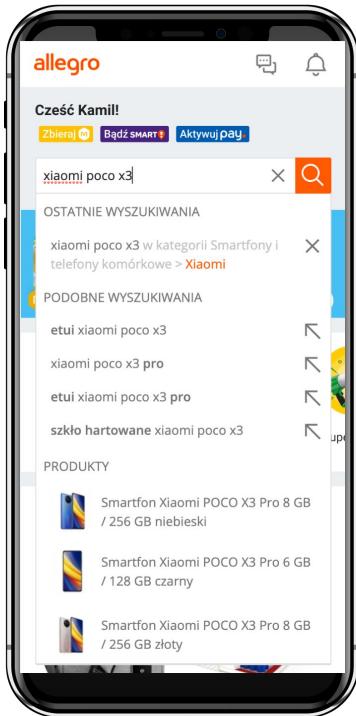
[Odpowiedz](#) [Odpowiedz wszystkim](#) [Prekaż](#) [Skomentuj](#)

1. Wybierz właściwy szablon

INTENCJA	NAZWA SZABLONU	Pan
Proszę o interwencję w dyskusji	Zakończenie 	<a href="#">Pan</a>
Proszę o interwencję w dyskusji	Wypowiedź w dyskusji 	<a href="#">Pan</a>
Proszę o interwencję w dyskusji	Wypowiedź i termin 	<a href="#">Pan</a>

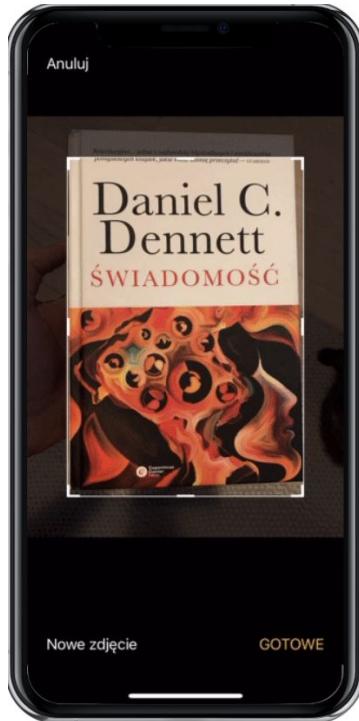
2. Jeżeli nie możesz skorzystać z szablonu, wskaż powód

[Błędne intencje](#)  [Właściwe intencje, lecz brak szablonu](#)  [Niewłaściwa płeć](#)

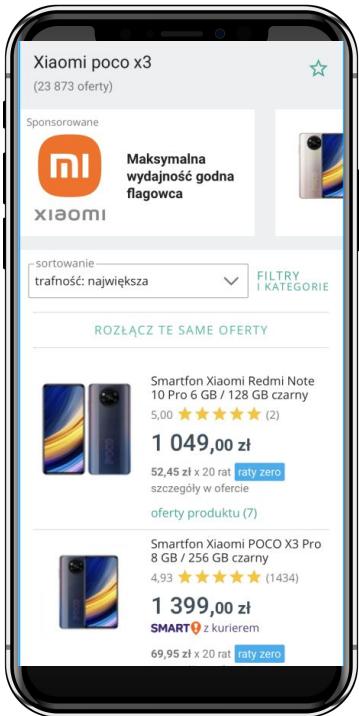


## Find the product

Recommendations  
Ads  
Personalisation  
Machine translation



## Visual Search



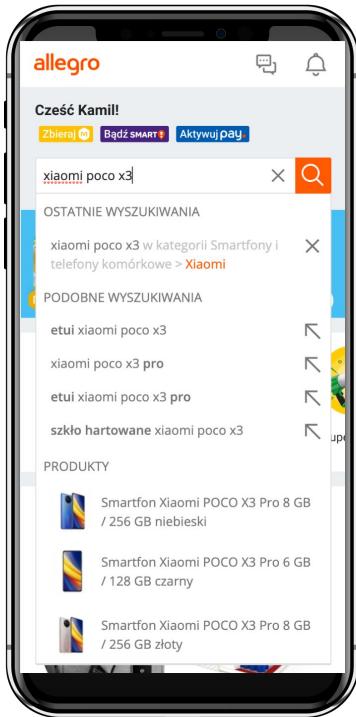
## Choose the best offer

Ranking  
Product matching  
Recommendations  
Category classification  
Ads



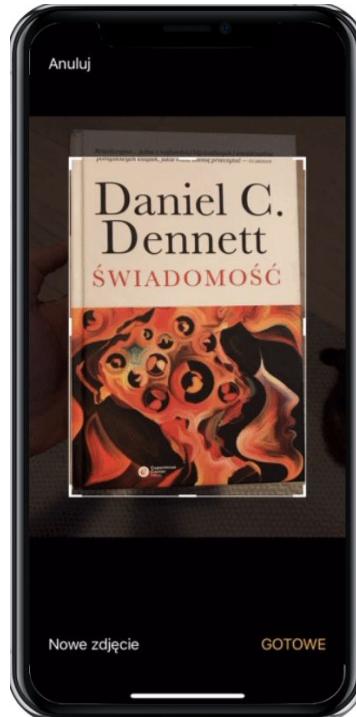
## Buy

Attribute extraction  
Description cleaning  
Recommendations

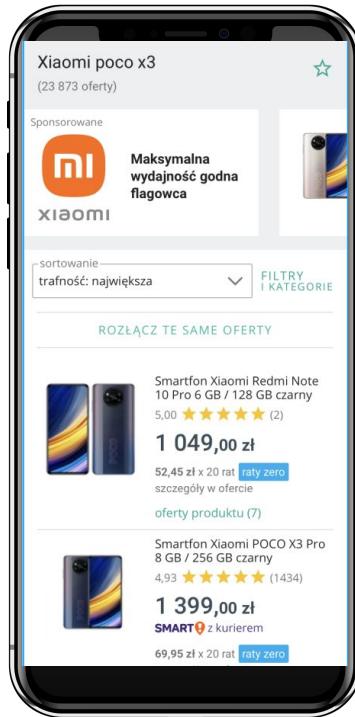


## Find the product

**Recommendations**  
Ads  
Personalisation  
Machine translation



## Visual Search



## Choose the best offer

**Ranking**  
Product matching  
Recommendations  
Category classification  
Ads



## Buy

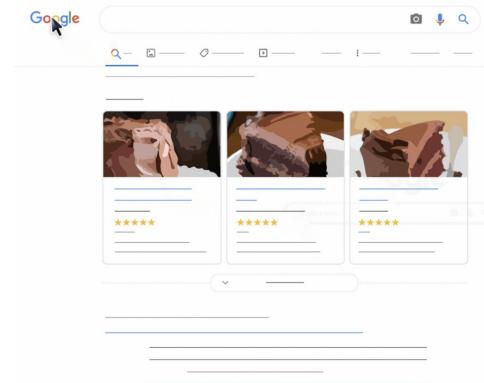
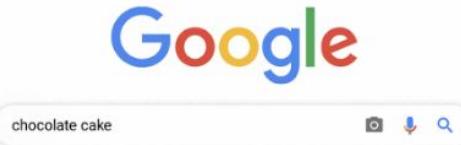
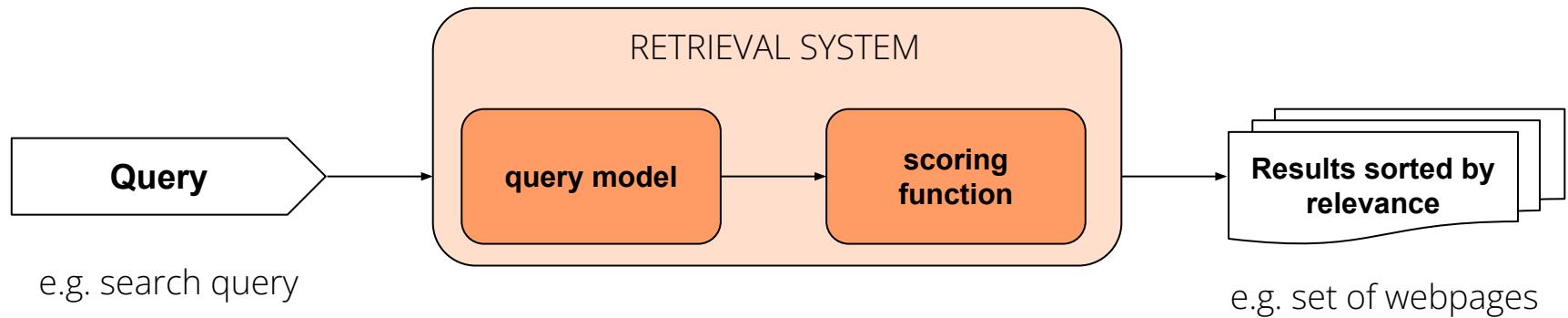
Attribute extraction  
Description cleaning  
Recommendations

# allegro

---

Introduction to retrieval tasks

## Retrieval in general



# Retrieval applications

- Search (document retrieval)
- Recommendations
- QA retrieval

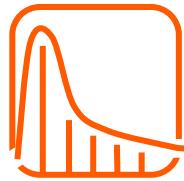


# Challenges in retrieval



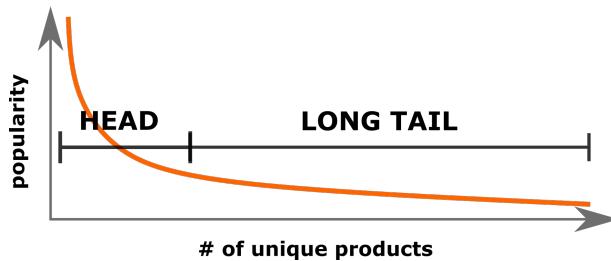
## Definition of relevance

*What is a good recommendation?*



## Skewed datasets

*Long tail & cold start*



## Fairness & interpretability

*Diverse & unbiased*

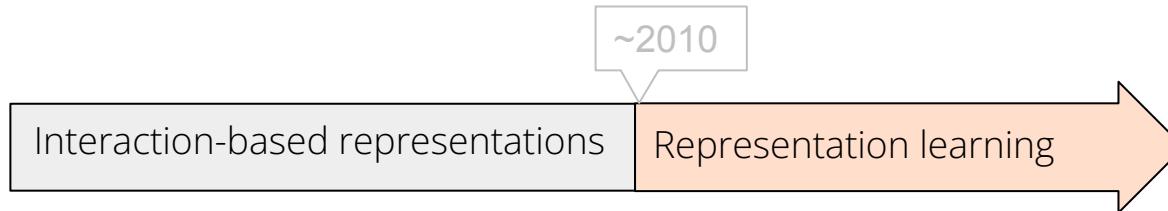


## Scale

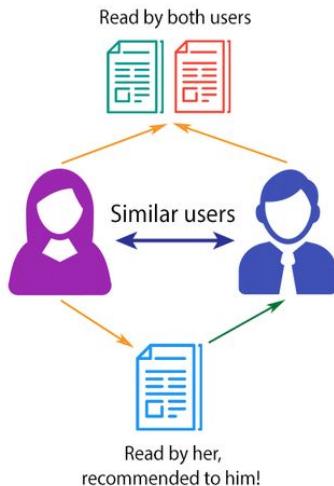
*Millions of entities and queries*



# Machine Learning meets retrieval



## COLLABORATIVE FILTERING



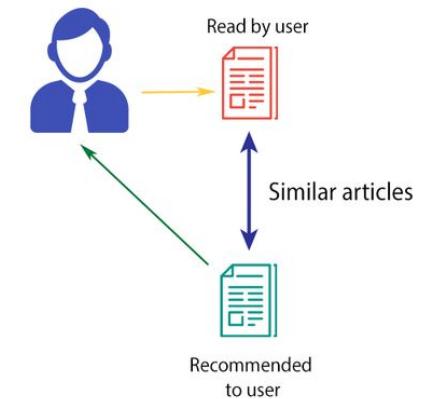
## Probabilistic retrieval

- collaborative filtering
- matrix factorization

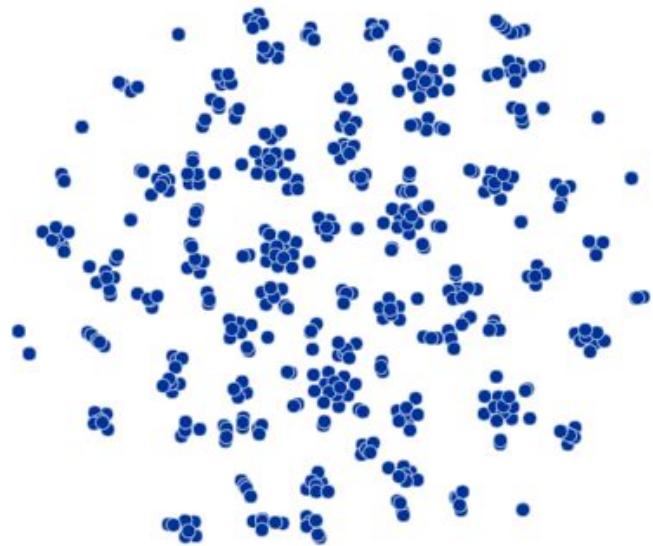
## Embedding-based retrieval

- DNNs, CNNs
- siamese NNs
- metric learning

## CONTENT-BASED FILTERING



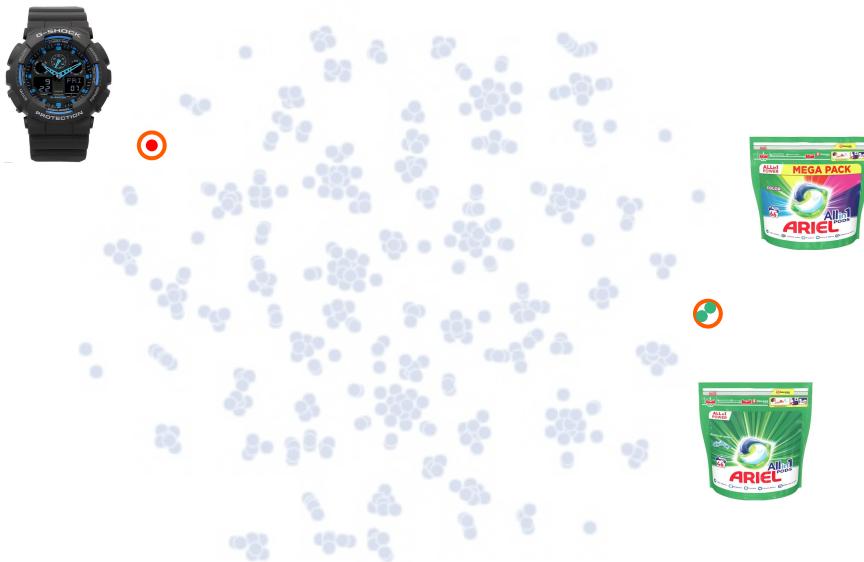
## Key concepts in large scale retrieval



latent space

each item is represented as an **embedding**  
(learned feature vector)

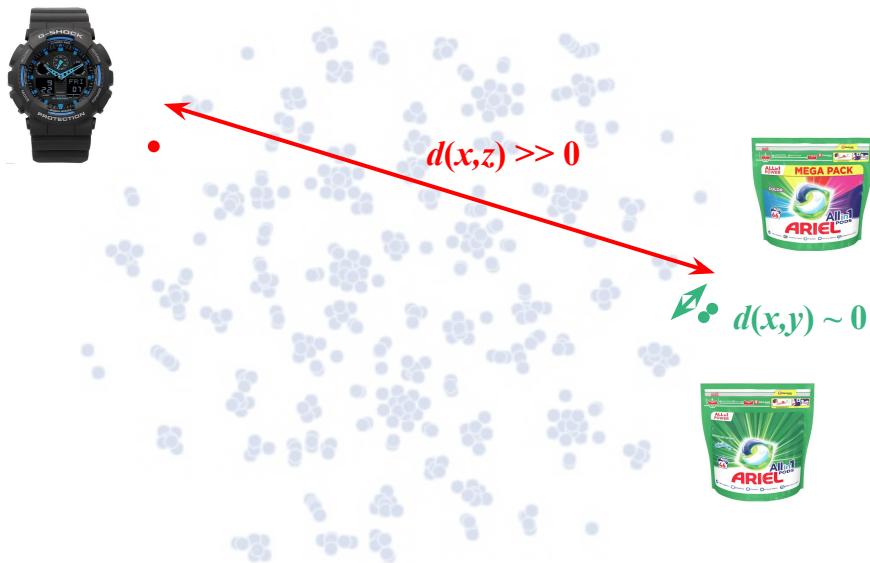
## Key concepts in large scale retrieval



each item is represented as an **embedding**  
(learned feature vector)

- similar items ( $x,y$ ) are close in the embedding space
- dissimilar items ( $x,z$ ) are far away in the embedding space

## Key concepts in large scale retrieval

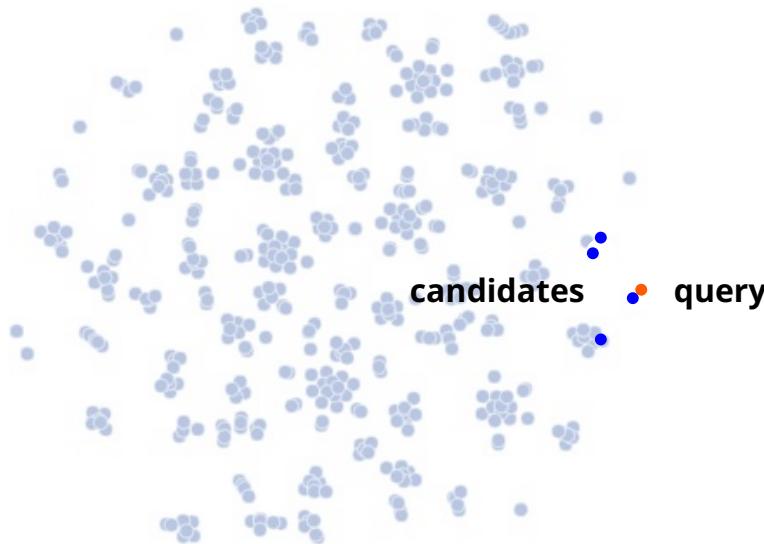


each item is represented as an **embedding**  
(learned feature vector)

- similar items (x,y) are close in the embedding space
- dissimilar items (x,z) are far away in the embedding space

similarity can be computed with a **distance metric**  
(e.g. dot product, L2) or own **scoring function**

# Key concepts in large scale retrieval



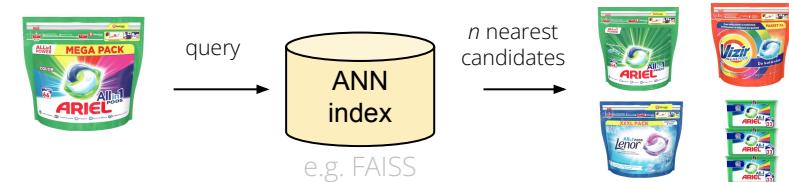
Nearest Neighbor Search

each item is represented as an **embedding**  
(learned feature vector)

- similar items ( $x,y$ ) are close in the embedding space
- dissimilar items ( $x,z$ ) are far away in the embedding space

similarity can be computed with a **distance metric**  
(e.g. dot product, L2) or own **scoring function**

To retrieve similar items upon query we can use an **index** based on chosen distance metric



# allegro

---

Large-scale retrieval at Allegro

## Retrieval at scale

**22M**

customers  
each month



**20M**

searches  
every day



**250M**

offers



**135k**

retailers



**14k**

categories



**2M**

new & changed  
offers every day



**120 ms**

search latency



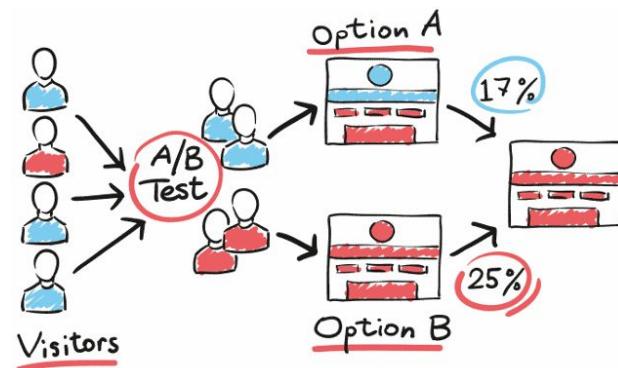
**99.9%**

SLA



# Data-driven approach to development

**Offline** evaluation → A/B Testing → **Online** deployment

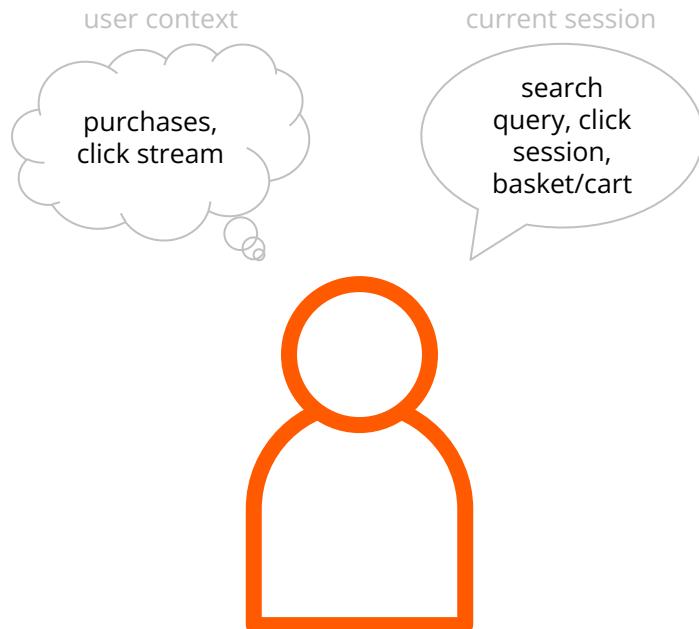


# Recommendations

---

Supporting the user in catalog  
exploration

## Goal of recommender system



**allegro**

wants to help the user in exploration of its  
enormous offer catalogue

# Where can you find recommendations?

- Offer page
- Main page
- Listing (search page)
- Basket (shopping cart)
- Thank you page
- Archived/Ended offers



**250M**  
clicks in  
reco carousels  
every month

**Filter cartridge Dafi Unimax 4 pcs**  
4.94 ★★★★★ 1374 ratings and 108 reviews



**FILTERS CARTRIDGE water DAFI UNIMAX 4pcs BRITA MAXTRA \***  
**33,99 zł SMART with courier**

**coupons for 5 PLN** 6 people purchased

**Free Smart! delivery with purchases from 40 zł from multiple sellers**

**Delivery from 6.70 zł** Estimated delivery: Thu, 10 Feb. – Tue, 15 Feb. at your place

**Free return** 14 days to withdraw from the agreement

**Quantity** 1 + out of 99596 units

**ADD TO CART** **BUY NOW**

**Allegro guarantees safe shopping** You will receive the purchased item or we will give you a refund. Check details.

**Clip a coupon for 5 PLN** A coupon is available for selected offers for the order value from 100 PLN. See the requirements in the Terms & Conditions

**CLIP COUPON**

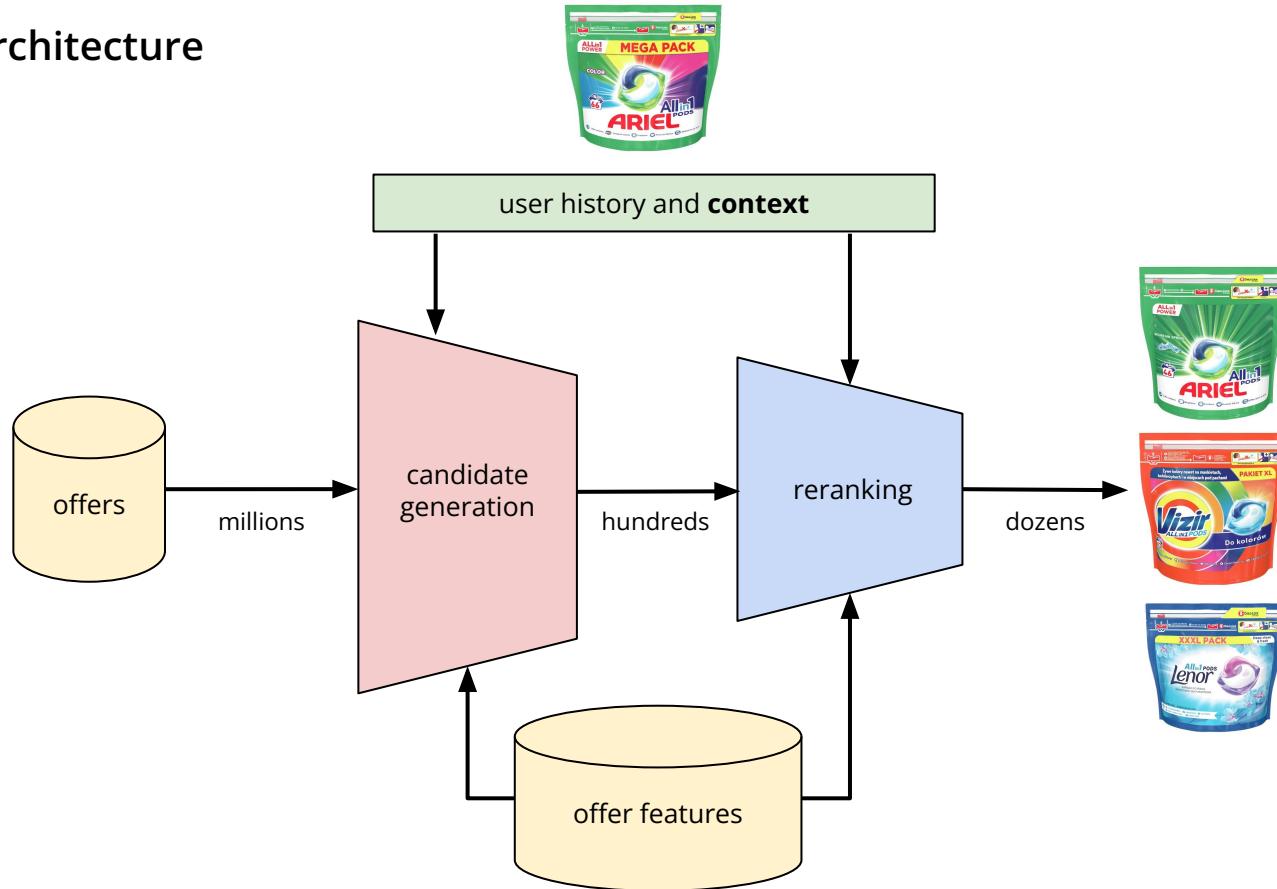
**Seller's other offers**

<b>35,90 zł</b> <b>SMART with courier</b> FILTERING JUG DAFI LUNA + 2x FILTER + BOTTLE 42,60 zł with delivery 48 people purchased	<b>43,90 zł</b> <b>SMART with courier</b> Dafi Unimax magnesium filter cartridge 4 pcs MG + 50,60 zł with delivery 111 people purchased	<b>33,90 zł</b> <b>SMART with courier</b> 4x refill for filter jug DAFI UNIMAX BRITA 40,60 zł with delivery 116 people purchased	<b>31,69 zł</b> <b>SMART with courier</b> 4x FILTER WESSPER for BRITA MAXTRA DAFI UNIMAX Aqua 38,39 zł with delivery 3 people purchased	<b>33,49 zł</b> <b>SMART with courier</b> FILTERING JUG DAFI ASTRA + 2x FILTER + BOTTLE 40,19 zł with delivery 291 people purchased	<b>58,90 zł</b> <b>SMART with courier</b> BRITA DAFI A + 6x HARD W FILTER 65,60 zł with delivery 24 people purchased
---	---	--	---	---	--

**ADD TO CART** **ADD TO CART** **ADD TO CART** **ADD TO CART** **ADD TO CART**

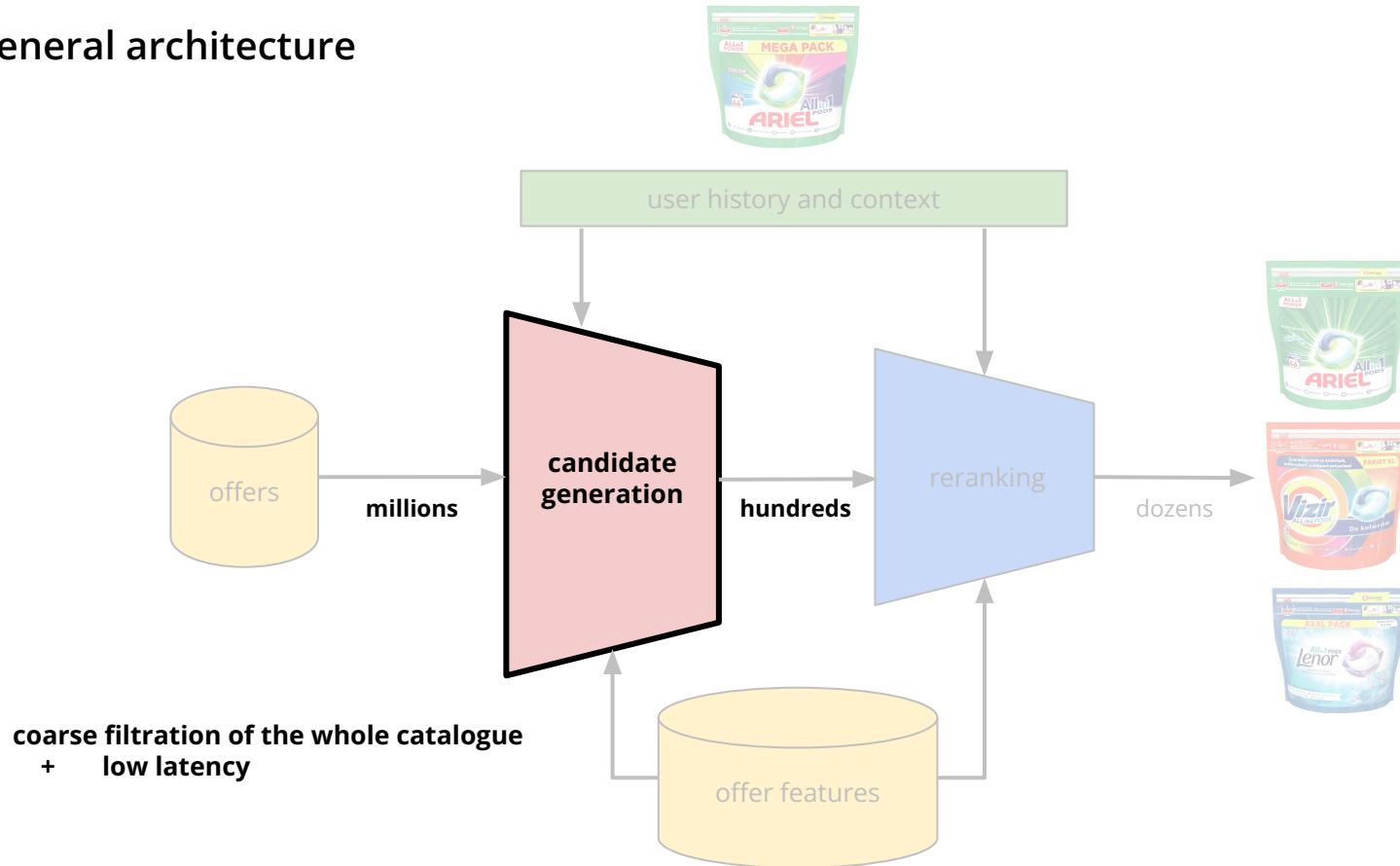
**See**  
Other items from this seller in the category Cartridges  
[All seller's items](#)  
[Seller's store](#)  
[Ask the seller](#)

## General architecture



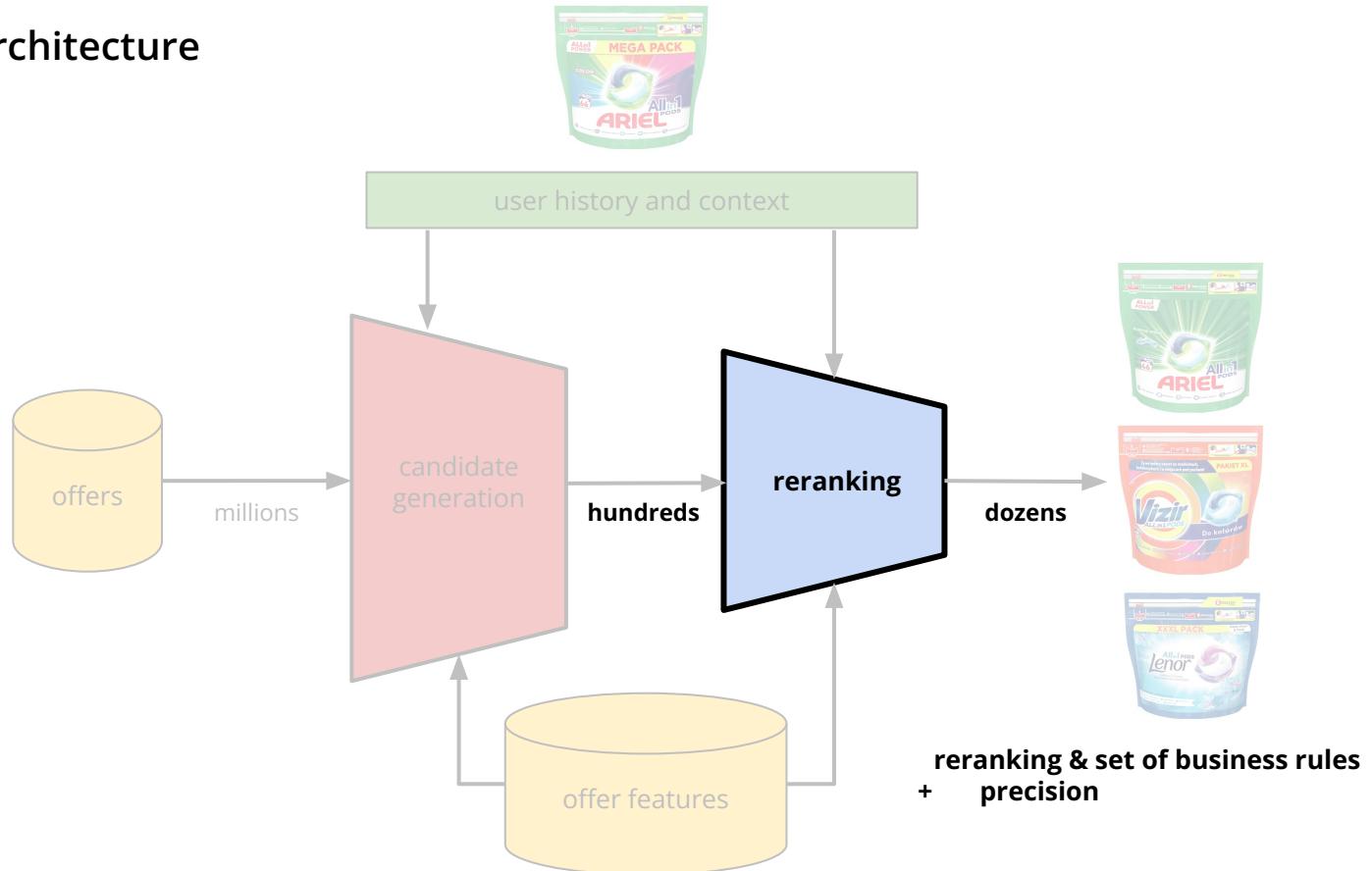
*Recommending What Video to Watch Next: A Multitask Ranking System, Zhao et al. 2019 (Google)*

## General architecture



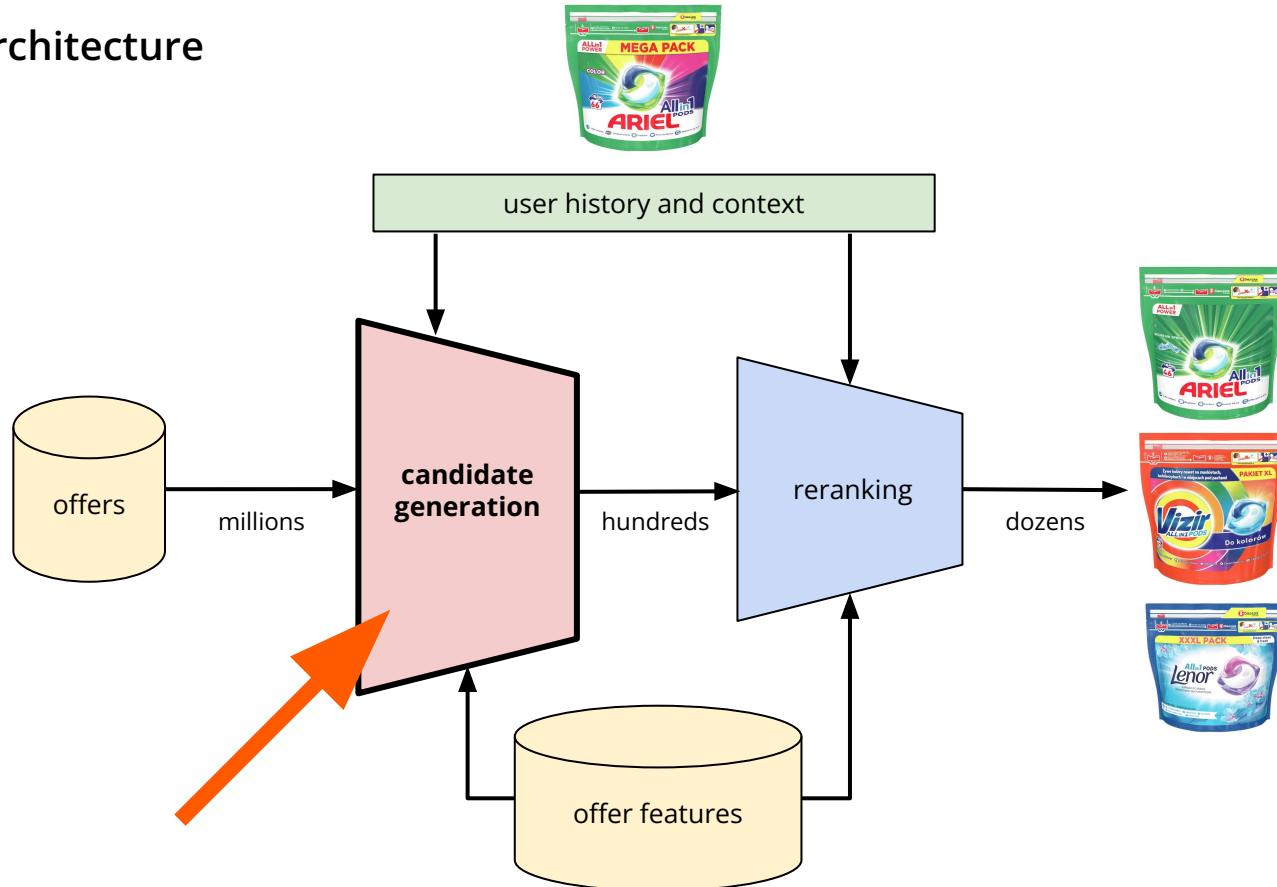
*Recommending What Video to Watch Next: A Multitask Ranking System, Zhao et al. 2019 (Google)*

## General architecture



*Recommending What Video to Watch Next: A Multitask Ranking System, Zhao et al. 2019 (Google)*

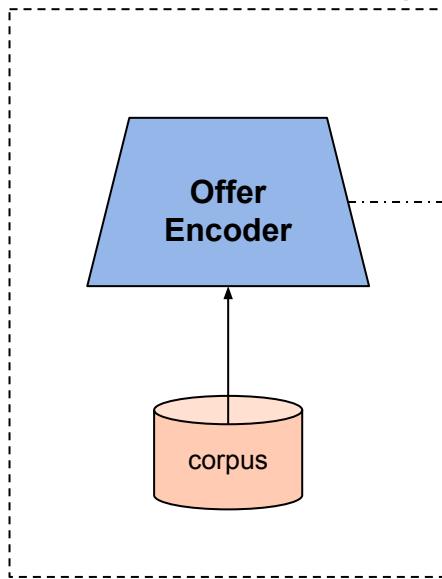
## General architecture



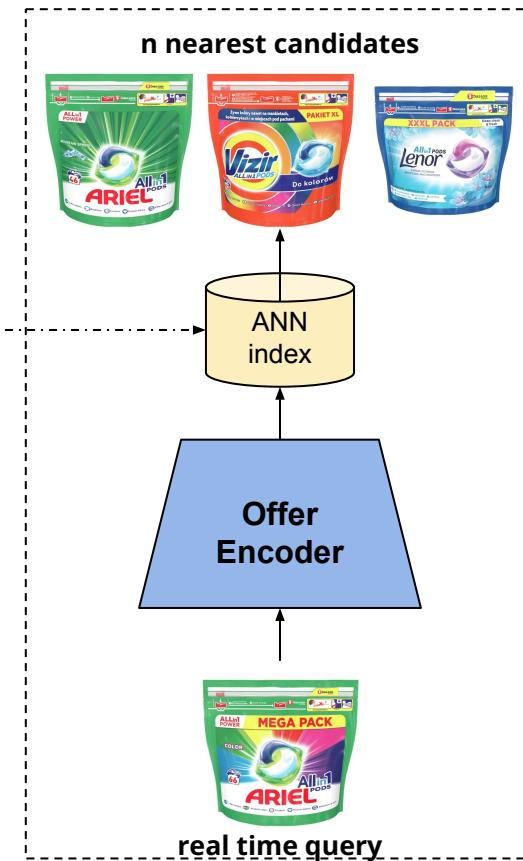
*Recommending What Video to Watch Next: A Multitask Ranking System, Zhao et al. 2019 (Google)*

# Candidate generation service

**Offline:** encoder training

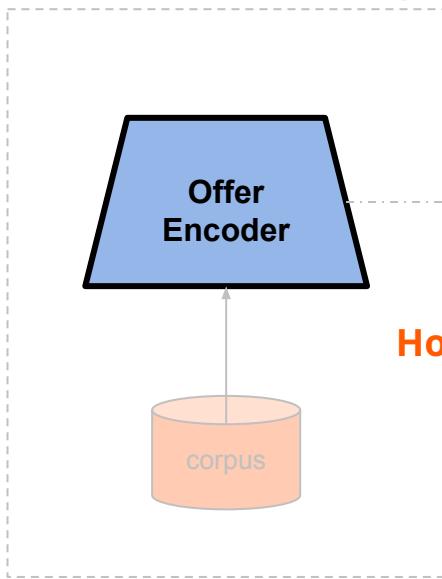


**Online:** candidate inference



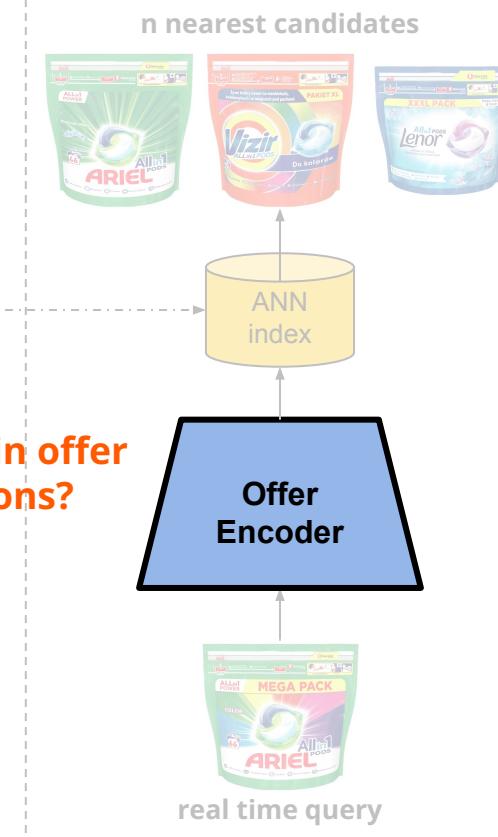
# Candidate generation service

**Offline:** encoder training

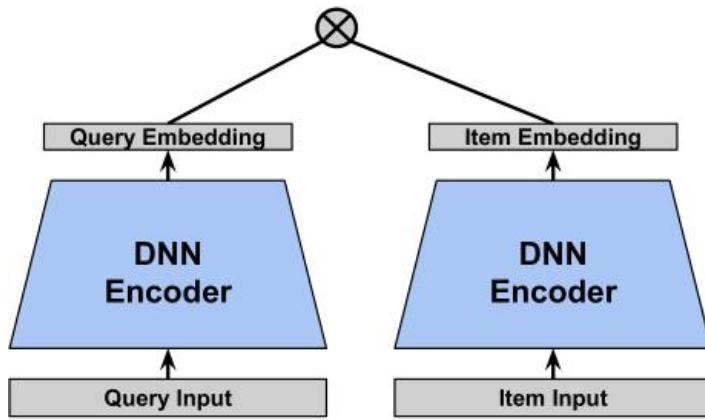


How can we train offer representations?

**Online:** candidate inference



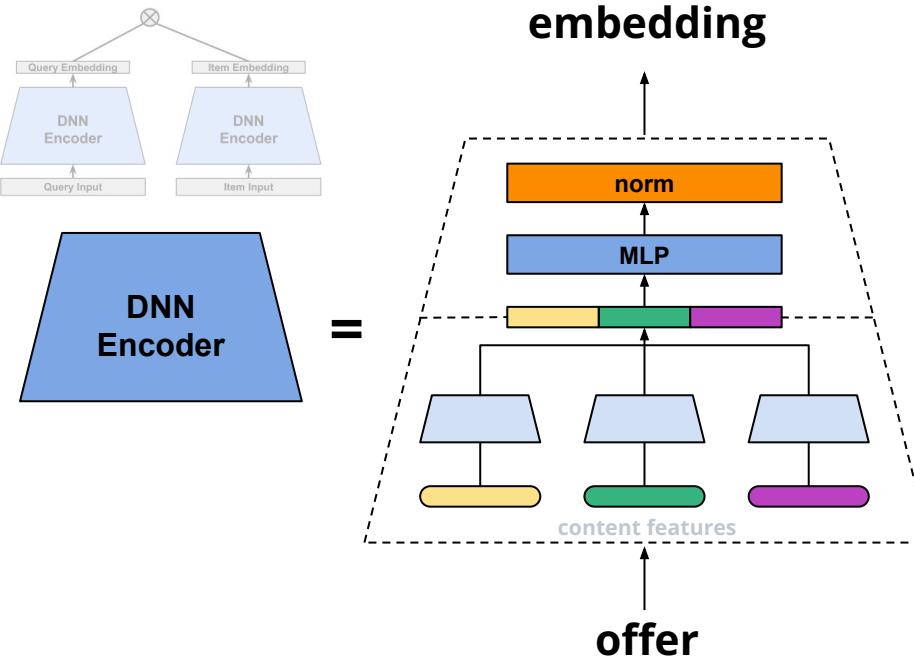
## Candidate generation with DNNs



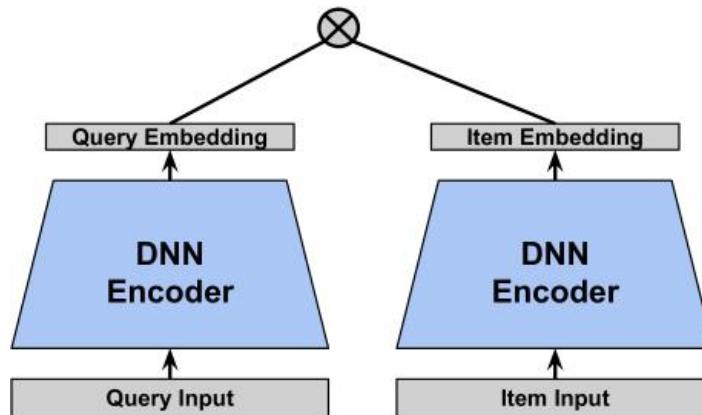
- **Training objective:** similar offers are close in embedding space
- **Hybrid system:**
  - **content-based** - offers are described purely by their content features (e.g. offer title, price)
  - **collaborative filtering** - trained on pairs of items clicked together

*Sampling-Bias-Corrected Neural Modeling for Large Corpus Item Recommendations, Yi et al. 2019 (Google)  
Personalized Embedding-based e-Commerce Recommendations at eBay, Wang et al. 2021 (eBay)*

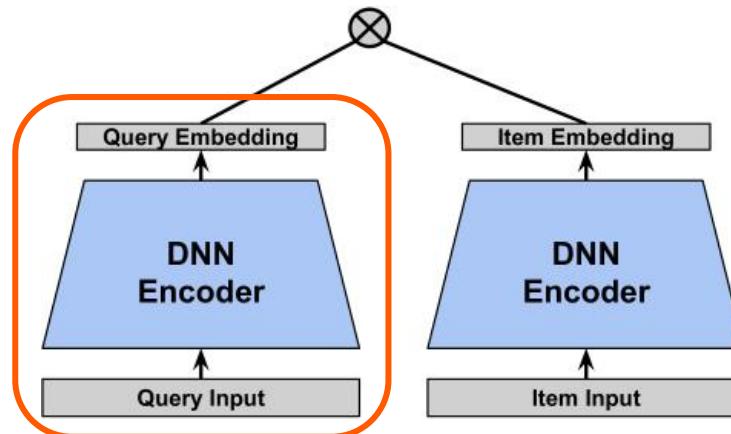
## DNN encoder architecture



## DNN encoder architecture



## DNN encoder architecture



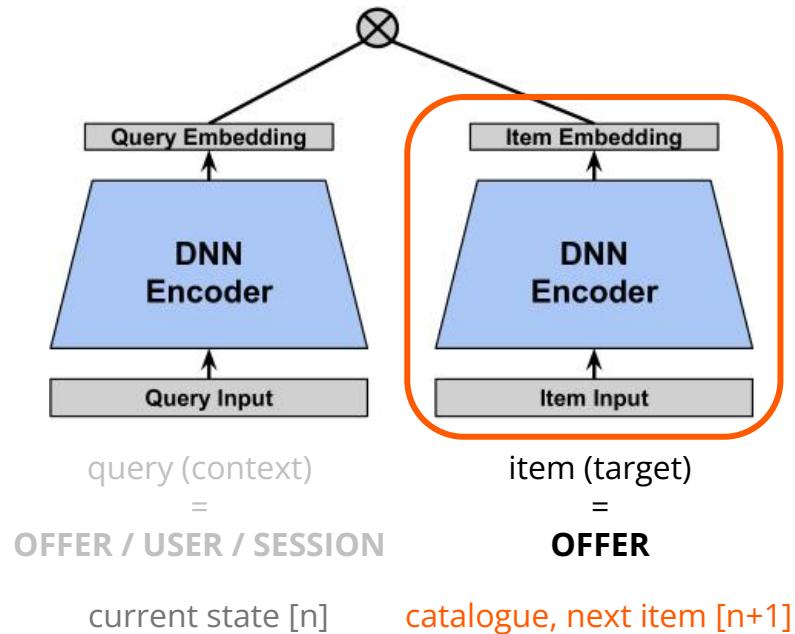
query (context)

=

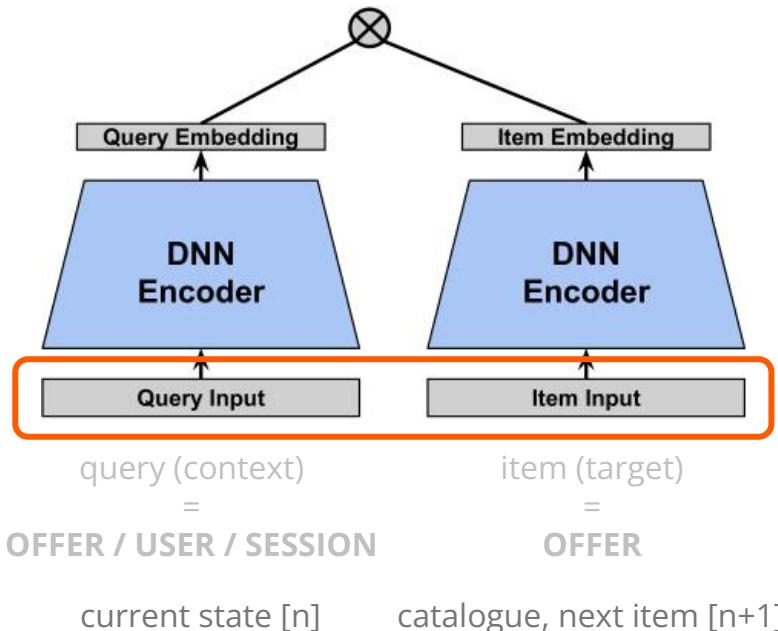
**OFFER / USER / SESSION**

current state [n]

## DNN encoder architecture

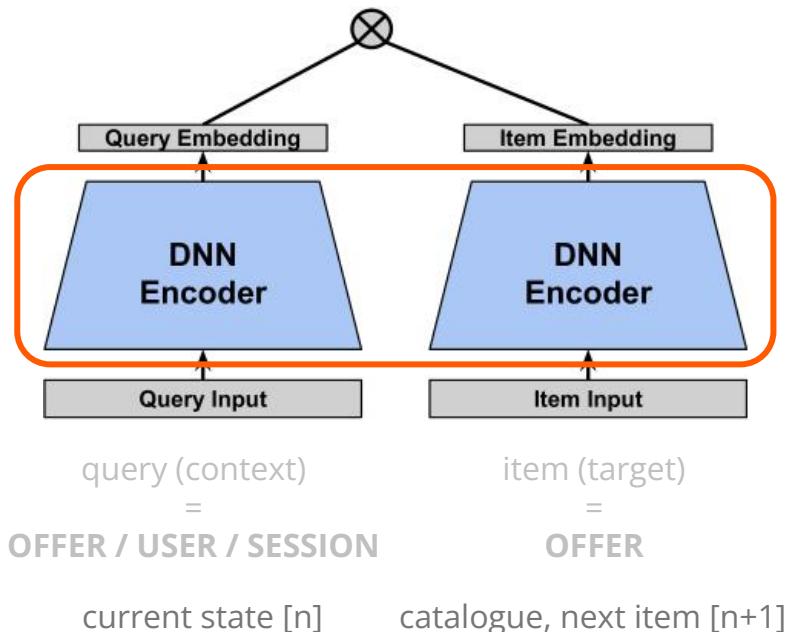


## DNN encoder architecture



trained on pairs of co-clicked offers  
(query, item) from user sessions

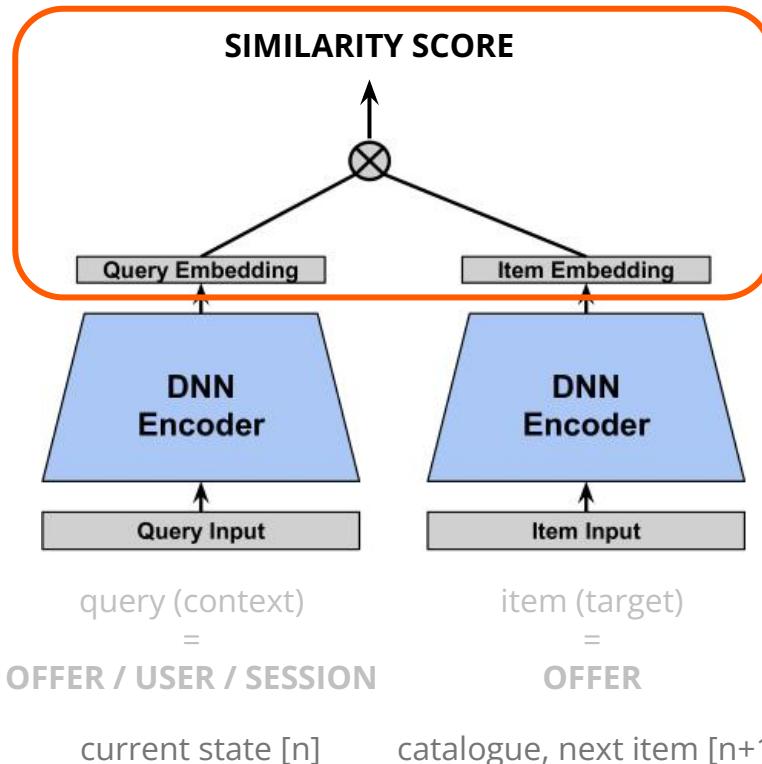
## DNN encoder architecture



shared weights (~Siamese NNs)

trained on pairs of co-clicked offers  
(*query, item*) from user sessions

## DNN encoder architecture

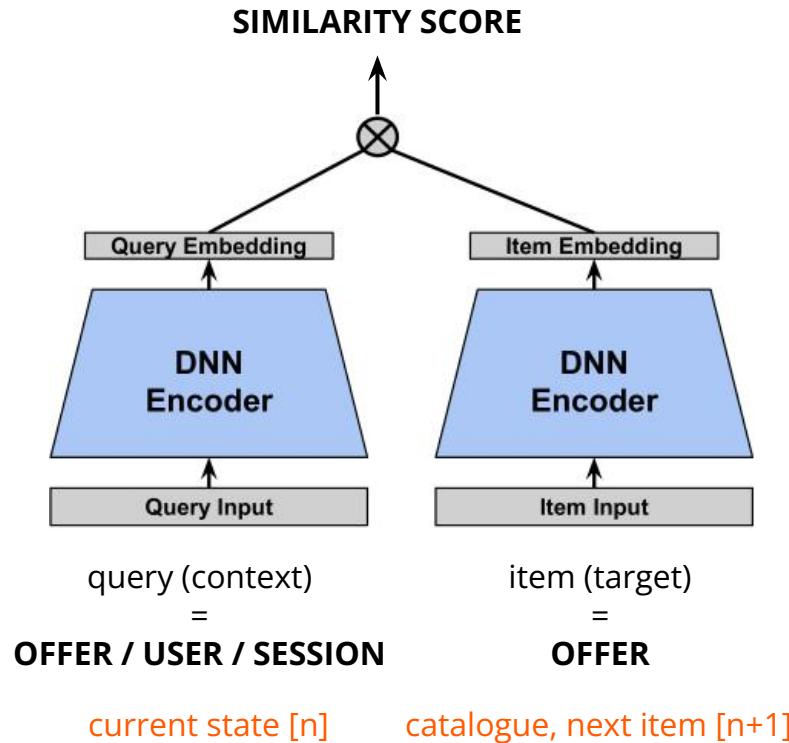


dot product / L2

shared weights (~Siamese NNs)

trained on pairs of co-clicked offers  
(*query, item*) from user sessions

## DNN encoder architecture



dot product / L2

shared weights (~siamese nns)

trained on pairs of co-clicked offers  
(*query, item*) from user sessions

## DNN training

Recommendations as a classification problem with the softmax probability  
**task**: predict=classify exact item

$$P(s_i|s_q) = \frac{e^{\gamma(v_i, v_q)}}{\sum_{j \in V} e^{\gamma(v_j, v_q)}}$$

## DNN training

Recommendations as a classification problem with the softmax probability

task: predict=classify exact item

$$P(s_i | s_q) = \frac{e^{\gamma(v_i, v_q)}}{\sum_{j \in V} e^{\gamma(v_j, v_q)}}$$

candidate item      query item      embeddings

retrieval set  
~millions      scoring function

The diagram illustrates the softmax probability formula for recommendation. It shows the inputs: a candidate item ( $s_i$ ) and a query item ( $s_q$ ). These inputs are processed through embeddings ( $v_i$  and  $v_q$ ) and a scoring function ( $\gamma$ ). The scoring function takes the dot product of the embeddings of the candidate and query items. The final probability is the softmax of this score, normalized by the scores of all items in the retrieval set (~millions).

## DNN training

Recommendations as a classification problem with the softmax probability  
task: predict=classify exact item

$$P(s_i|s_q) = \frac{e^{\gamma(v_i, v_q)}}{\sum_{j \in V} e^{\gamma(v_j, v_q)}}$$

Optimization criterion: Negative Log-likelihood of observed user clicks

$$\textit{loss} = - \sum_{j=1}^V y_j \log(\hat{y}_j)$$

*Personalized Embedding-based e-Commerce Recommendations at eBay*, Wang et al. 2021 (eBay)

## DNN training

Recommendations as a classification problem with the softmax probability  
**task**: predict=classify exact item

$$P(s_i|s_q) = \frac{e^{\gamma(v_i, v_q)}}{\sum_{j \in V} e^{\gamma(v_j, v_q)}}$$


**infeasible!**

Optimization criterion: **Negative Log-likelihood** of observed user clicks

V~millions

$$\text{loss} = - \sum_{j=1}^V y_j \log(\hat{y}_j)$$

*Personalized Embedding-based e-Commerce Recommendations at eBay*, Wang et al. 2021 (eBay)

## DNN training

Recommendations as a classification problem with the softmax probability  
**task**: predict=classify exact item

$$P(s_i|s_q) = \frac{e^{\gamma(v_i, v_q)}}{\sum_{j \in B} e^{\gamma(v_j, v_q)}}$$

**negative sampling**  
 instead

B~thousands

Optimization criterion: **Negative Log-likelihood** of observed user clicks

$$loss = - \sum_{j=1}^V y_j \log(\hat{y}_j)$$

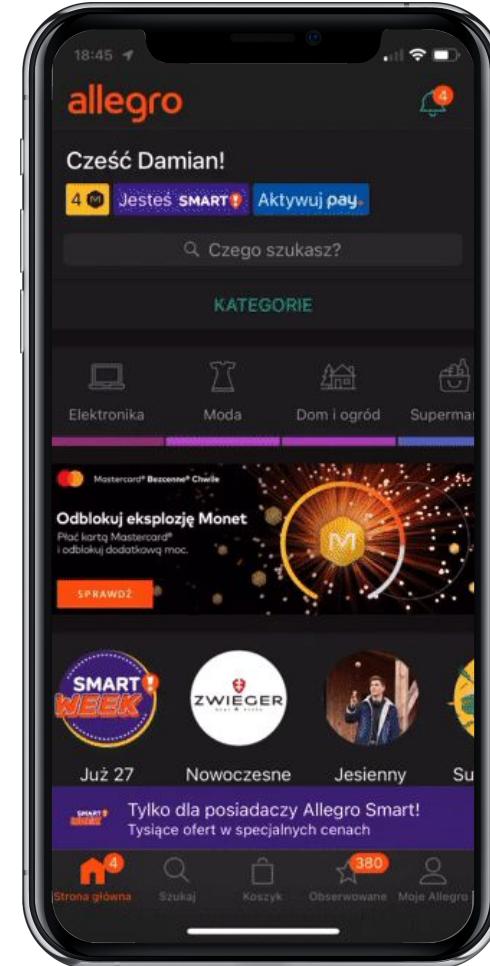
*Personalized Embedding-based e-Commerce Recommendations at eBay*, Wang et al. 2021 (eBay)

# Visual Search

---

Simplifying the search for  
visually-explainable objects

The screenshot shows a product page for a light-colored, knee-length robe with a belt. The main image features two women modeling the item. Below the main image are smaller images of the robe from different angles. The product title is "SUKIENKA ROZKŁOSZOWANA KOPERTOWY DEKOŁOT". The price is listed as 58,99 zł SMART. A "DODAJ DO KOSZYKA" (Add to Cart) button is visible.



**Over 120M offers indexed (~50% of Allegro)**

## Task: find visually similar items

query  
(user photo)

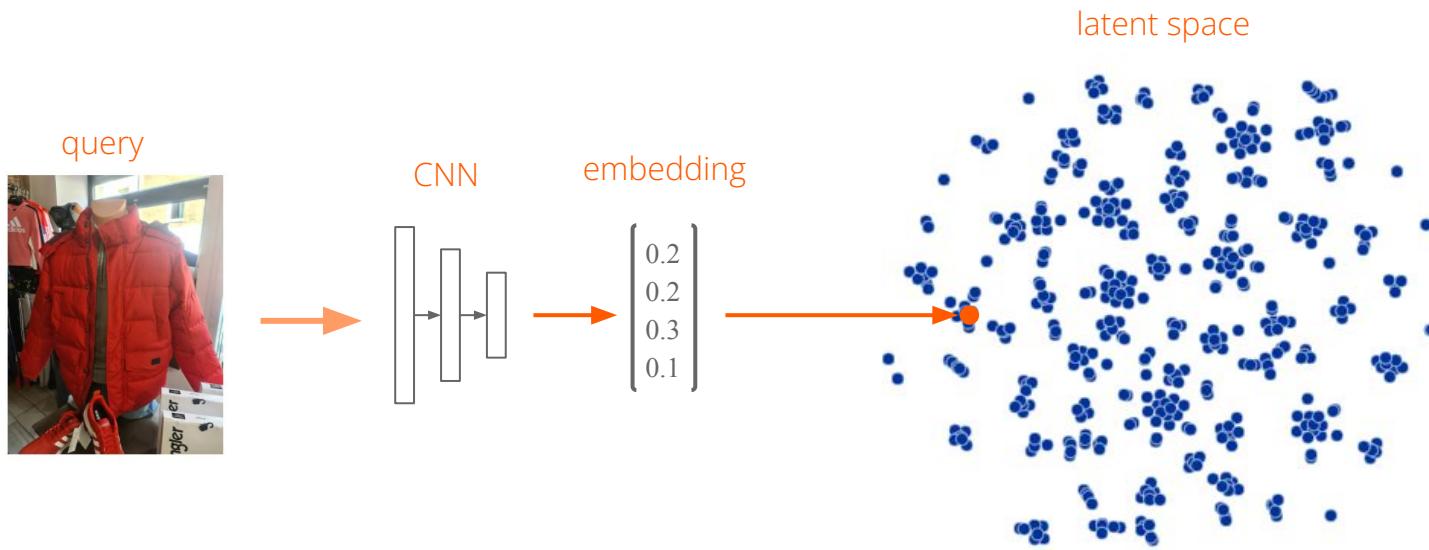


5 most similar images  
(Allegro offers)

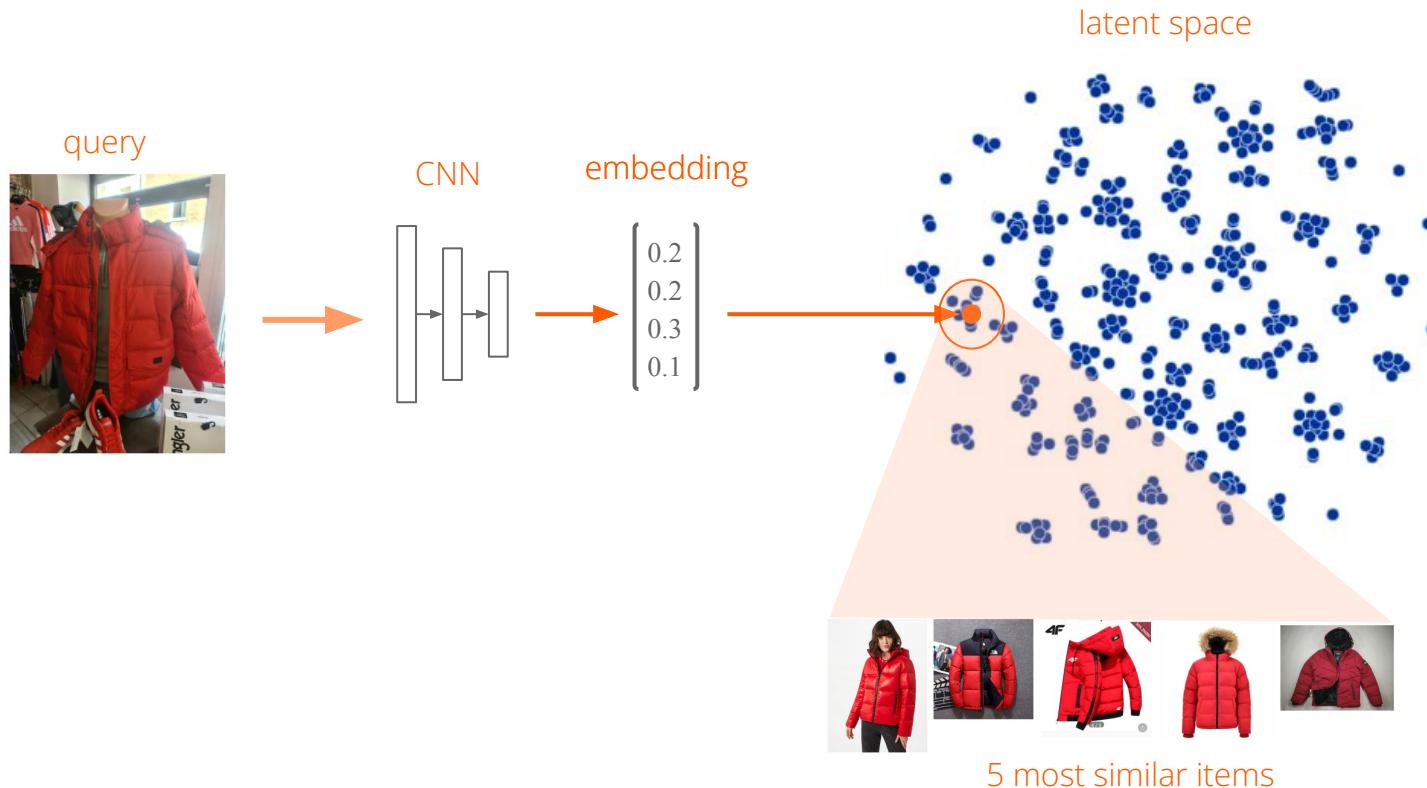


metric: accuracy@5

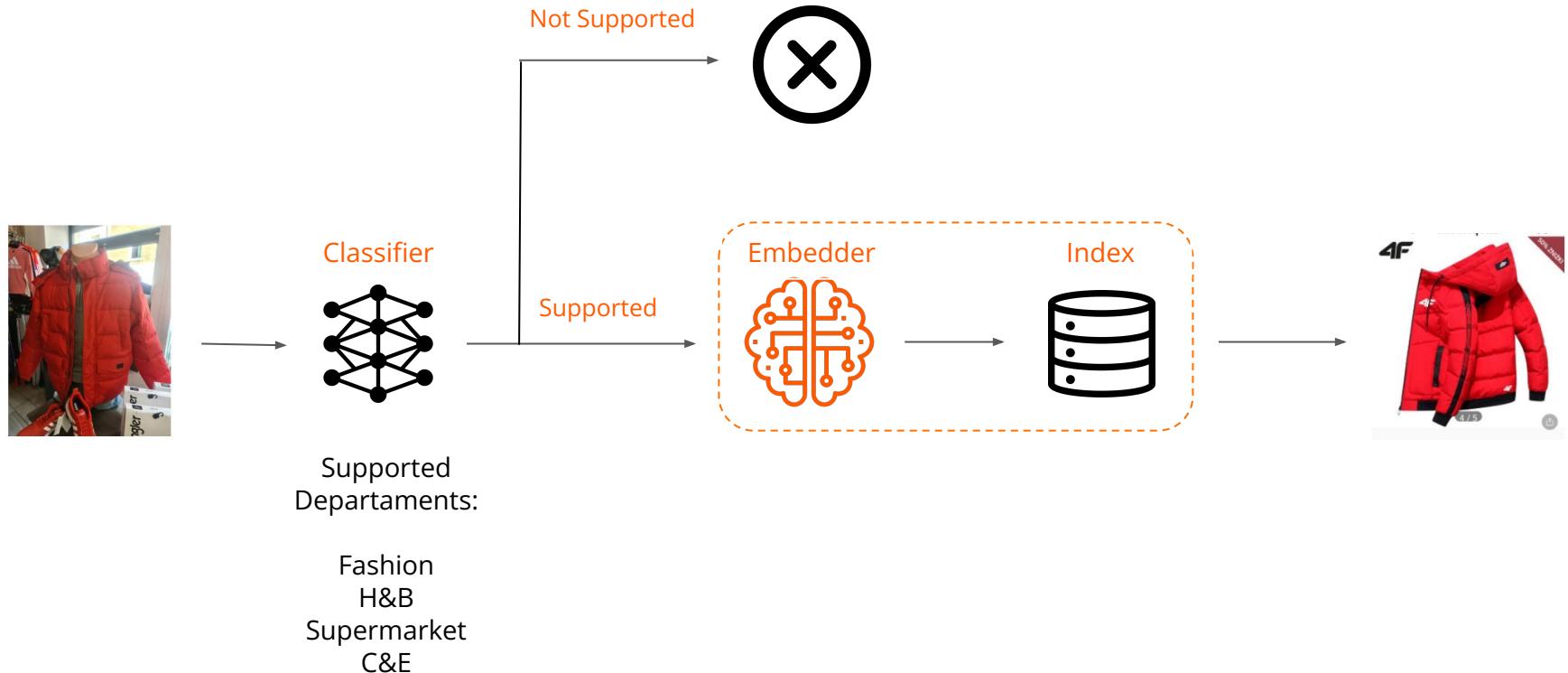
## Representing similarity



## Representing similarity



# Deployment

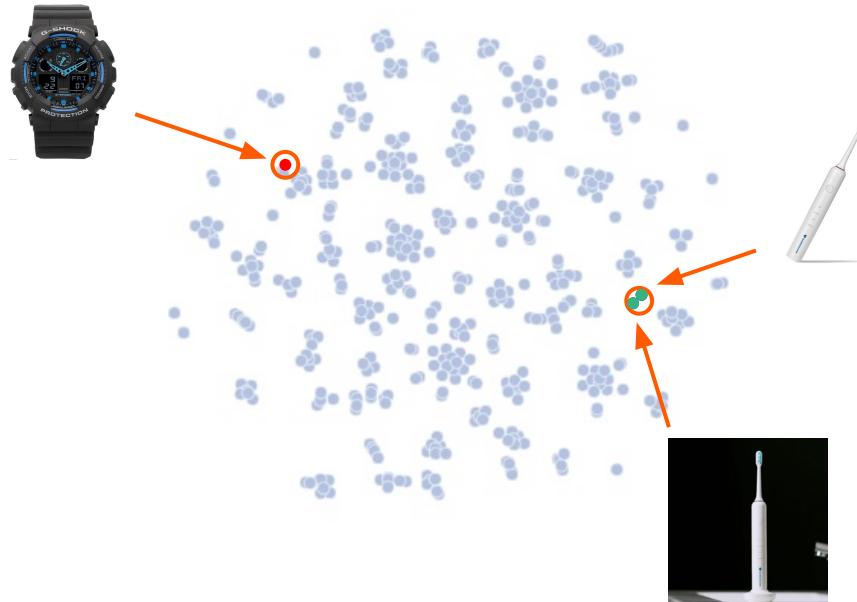


# accuracy@5

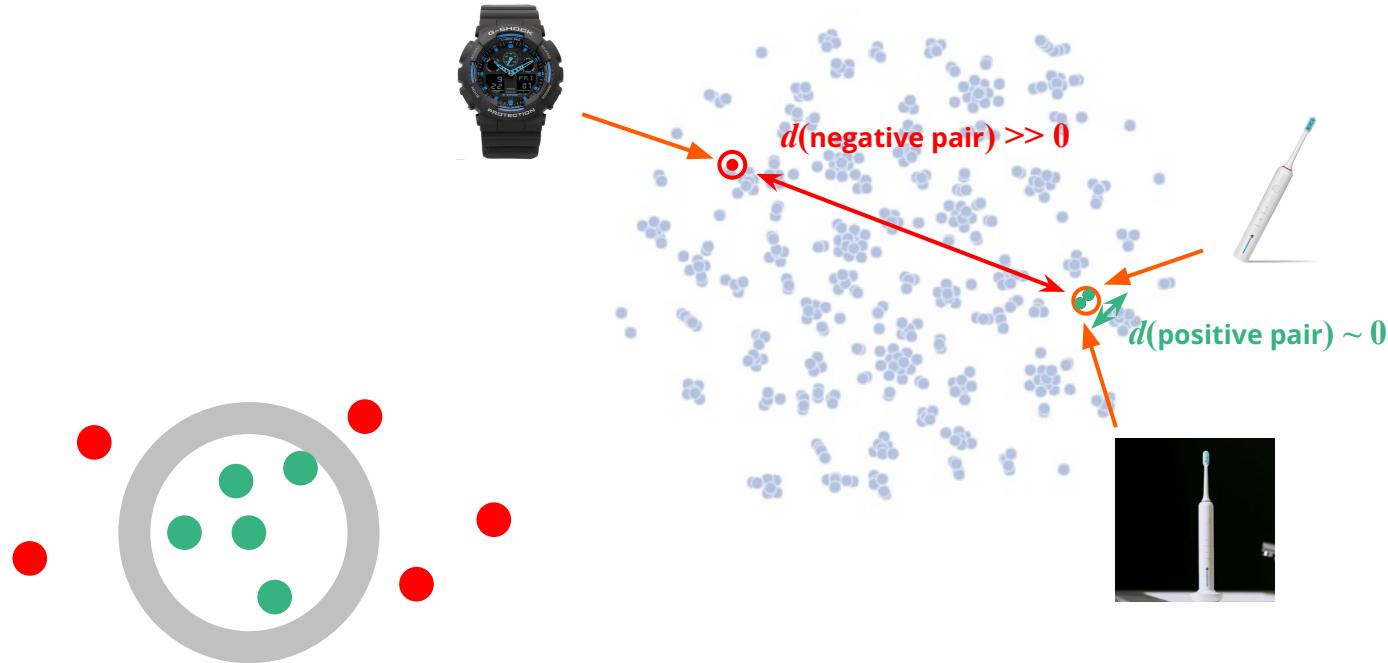
ImageNet 35%



## Metric learning

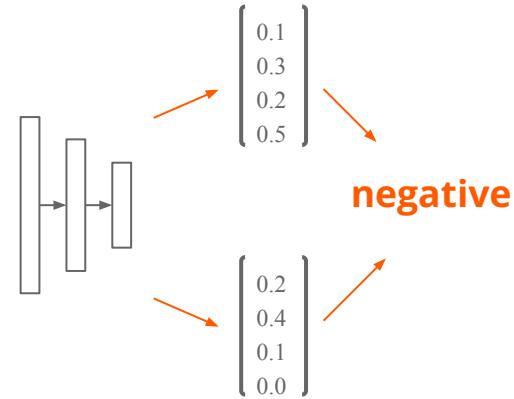
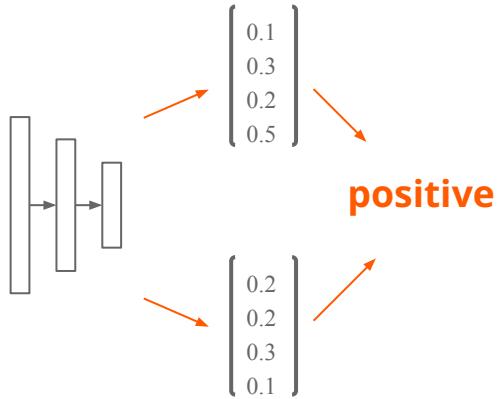


## Metric learning

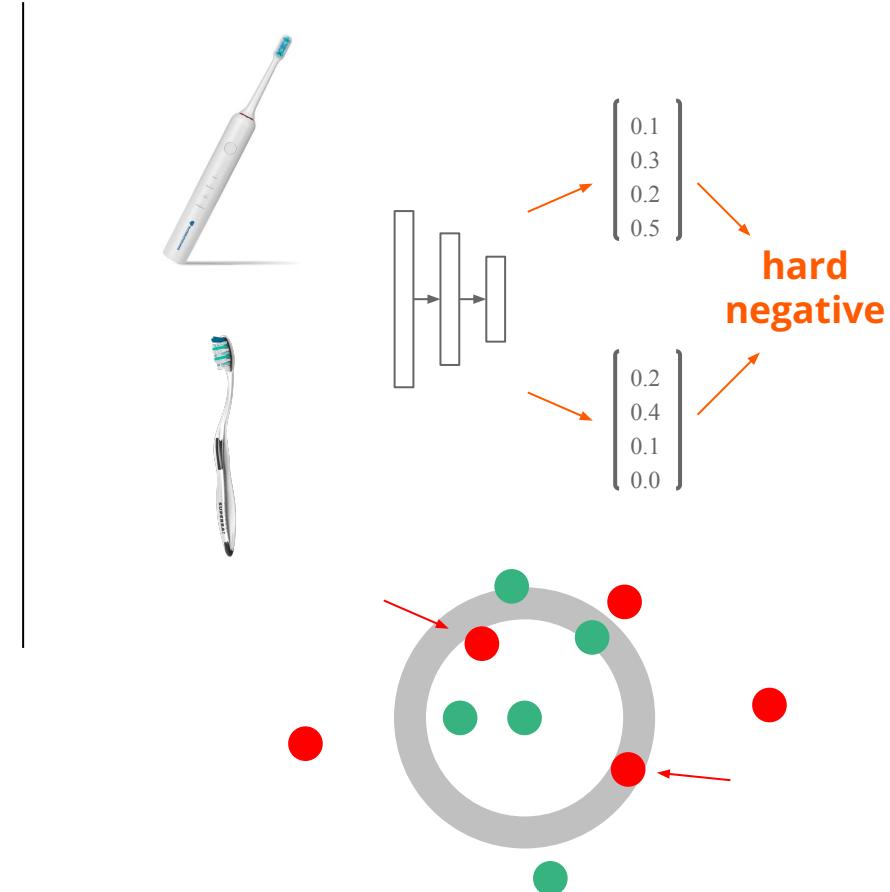
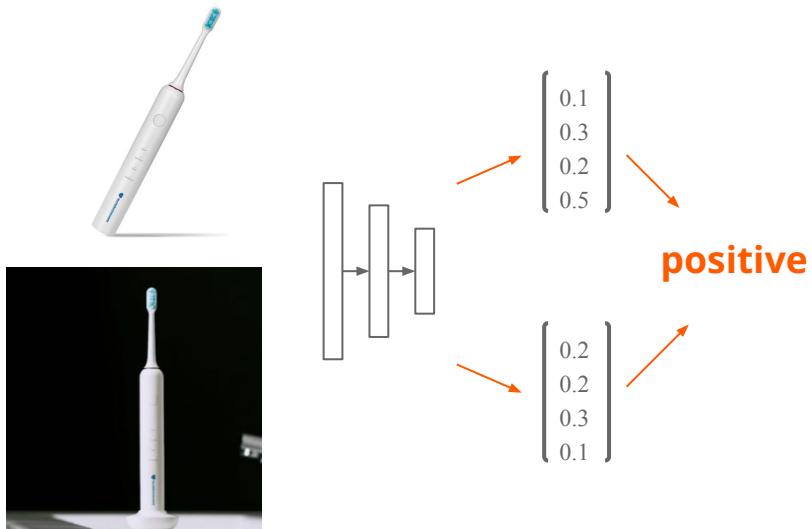


$$\text{Triplet loss: } L = \sum_{\text{triplets}} [ d_{\text{positive}} - d_{\text{negative}} + \text{margin} ]_+$$

## Image pairs from Allegro offers



## Image pairs from Allegro offers



## accuracy@5



# Generating positive pairs

Useful images

Shop



Street



Confusing images

Multiple



Detail



Trash

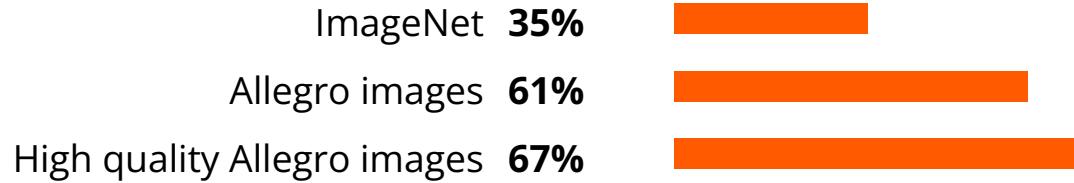


Other



Requirement: image filtering and deduplication

## accuracy@5



## Domain shift: street to shop

User images: street



Allegro images: shop



# Synthetic pairs

Let's simulate street images!

Seamless clone



Gaussian blurring clone



# Cut and paste?

It's not that easy!

## Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection

Debidatta Dwibedi      Ishan Misra      Martial Hebert  
The Robotics Institute, Carnegie Mellon University  
`debidatta@cmu.edu, {imisra, hebert}@cs.cmu.edu`



Figure 6: Different blending modes used while generating datasets. These modes help the model in ignoring artifacts arising from pasting objects on background scenes. More

# Synthetic pairs

Let's simulate street images!

Seamless clone



Gaussian blurring clone



## Natural background vs. uniform texture

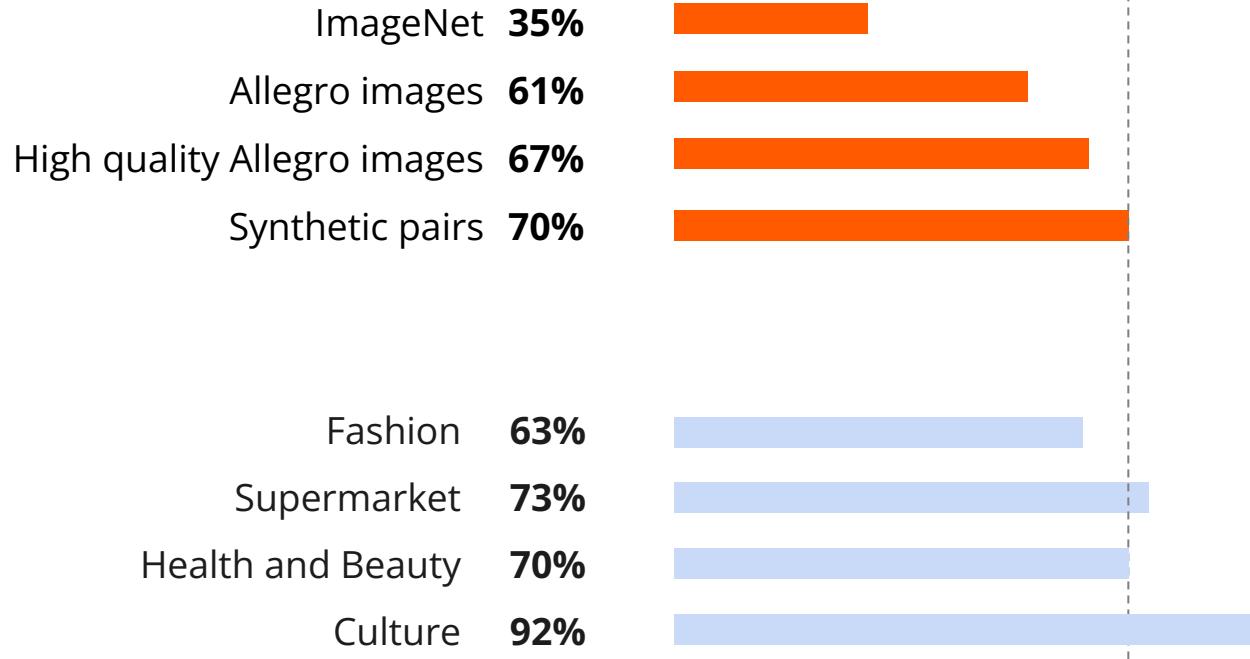
Natural images



Uniform texture



## accuracy@5



# Learning to Rank

---

Real-time search service

Magic



1.



2.



3.



4.



5.



6.



7.



Re-Ranking

4.



5.



3.



1.



7.



6.



2.

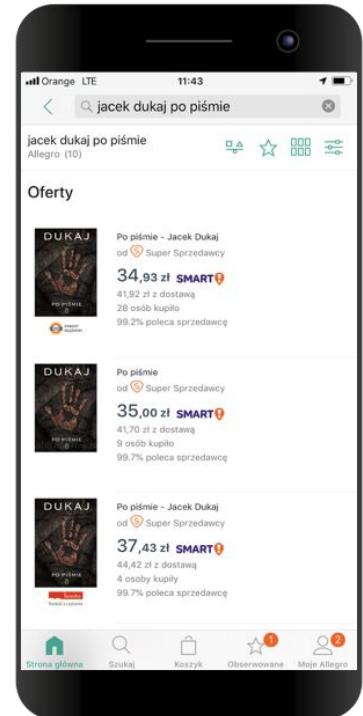


# Search architecture

Query  
anything you like



etui motorola moto e6s      szukaj wielu      Etui i pokrowce      SZUKAJ



# Search architecture

Query  
anything you like

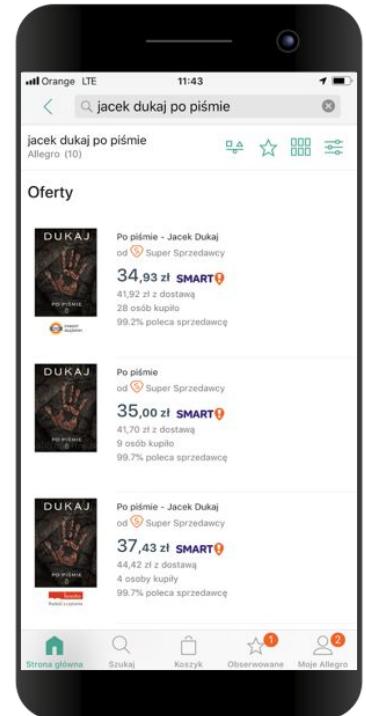


etui motorola moto e6s      szukaj wielu      Etui i pokrowce      SZUKAJ

250M offers

text-based  
search  
(inverted index) + offline  
scoring  
(XGBoost)

Matching offers



# Search architecture

Query  
anything you like



etui motorola moto e6s      szukaj wielu      Etui i pokrowce      SZUKAJ

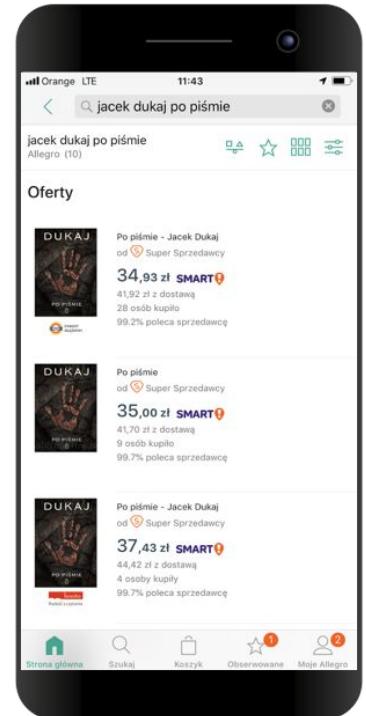
250M offers

text-based  
search  
(inverted index) +  
offline  
scoring  
(XGBoost)

Matching offers

reranking  
(DNN)

Listing with 300 top offers re-ranked



# Search architecture

Query  
anything you like



szukaj wielu

SZUKAJ

250M offers

text-based  
search  
(inverted index) +  
offline  
scoring  
(XGBoost)

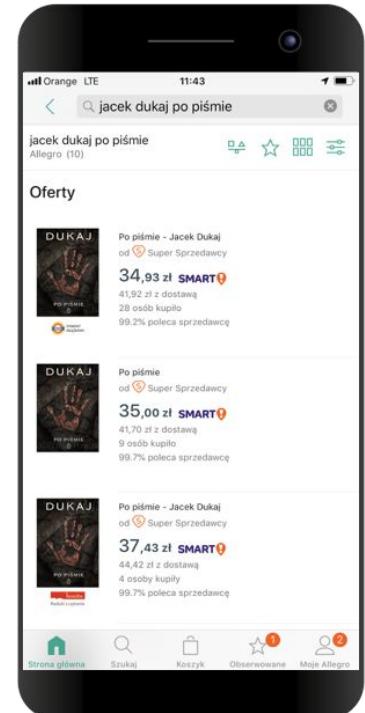
Matching offers

reranking  
(DNN)  
+ 10% nDCG  
+ 4.5% search GMV

Listing with 300 top offers re-ranked



clicks & purchases



# Learning to Rank

Maximizing utility by optimizing nDCG

## Objective:

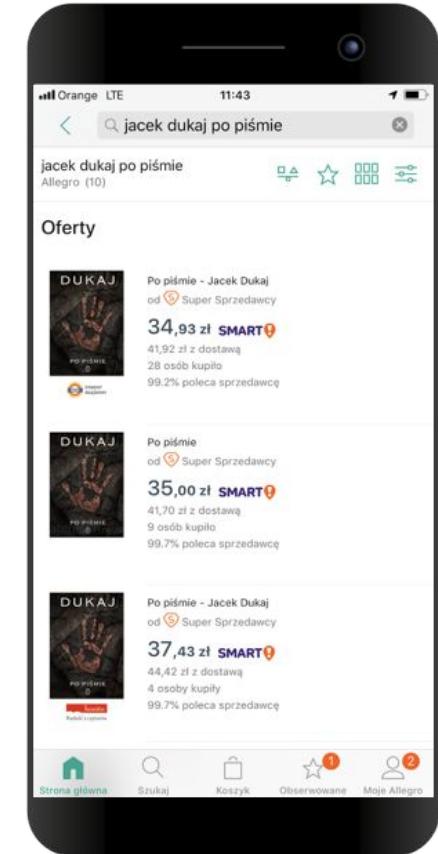
Maximize the **utility** of the selected subset (300 offers)  
by presenting the most relevant items first

## How to measure utility?

Standard metric: Normalized Discounted Cumulative Gain

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$$

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$



# Learning to Rank

Maximizing utility by optimizing nDCG

## Objective:

Maximize the **utility** of the selected subset (300 offers)  
by presenting the most relevant items first

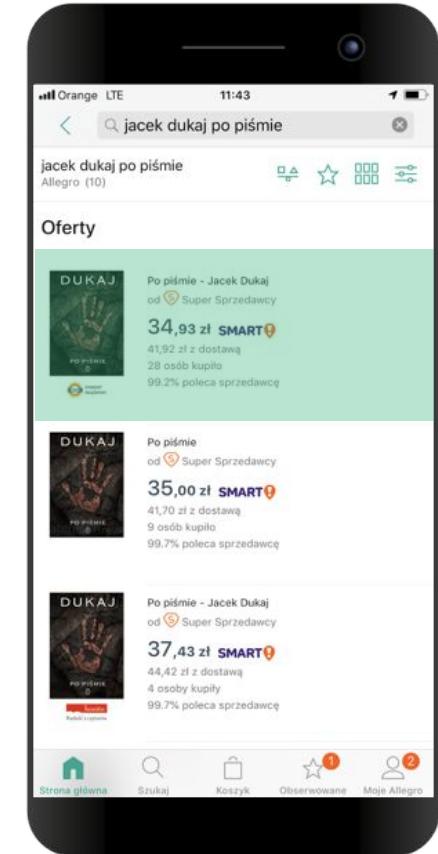
$$nDCG = 1$$

## How to measure utility?

Standard metric: Normalized Discounted Cumulative Gain

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$$

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$



# Learning to Rank

Maximizing utility by optimizing nDCG

## Objective:

Maximize the **utility** of the selected subset (300 offers)  
by presenting the most relevant items first

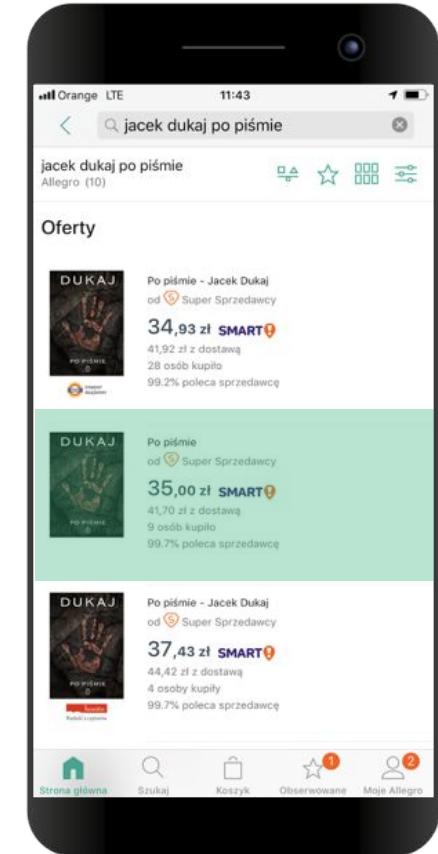
## How to measure utility?

Standard metric: Normalized Discounted Cumulative Gain

$$\text{DCG}_p = \sum_{i=1}^p \frac{\text{rel}_i}{\log_2(i+1)}$$

$$\text{nDCG}_p = \frac{\text{DCG}_p}{\text{IDCG}_p}$$

$$nDCG \simeq 0.63$$



# Learning to Rank

Maximizing utility by optimizing nDCG

## Objective:

Maximize the **utility** of the selected subset (300 offers)  
by presenting the most relevant items first

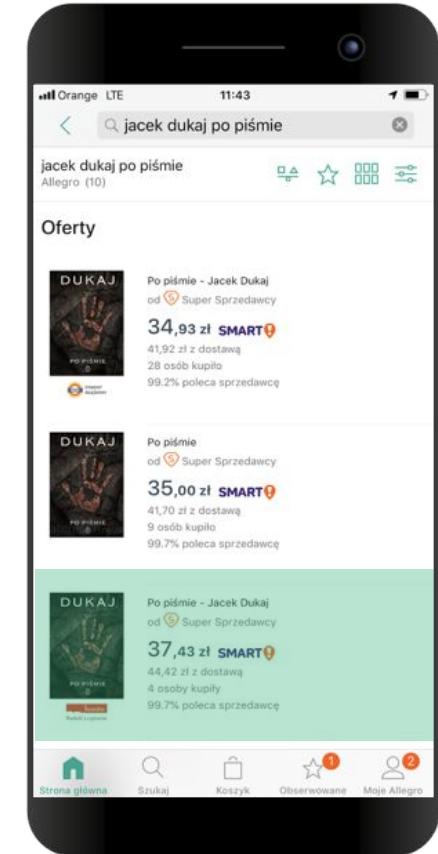
## How to measure utility?

Standard metric: Normalized Discounted Cumulative Gain

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$$

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

$$nDCG = 0.5$$



# Learning to Rank

Maximizing utility by optimizing nDCG

## Objective:

Maximize the **utility** of the selected subset (300 offers)  
by presenting the most relevant items first

## How to measure utility?

Standard metric: Normalized Discounted Cumulative Gain

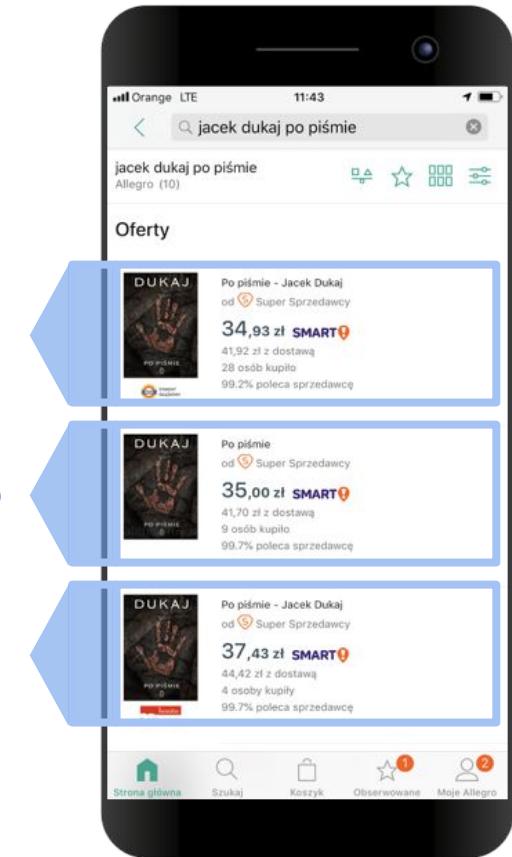
$$\text{DCG}_p = \sum_{i=1}^p \frac{\text{rel}_i}{\log_2(i+1)}$$

$$\text{nDCG}_p = \frac{\text{DCG}_p}{\text{IDCG}_p}$$

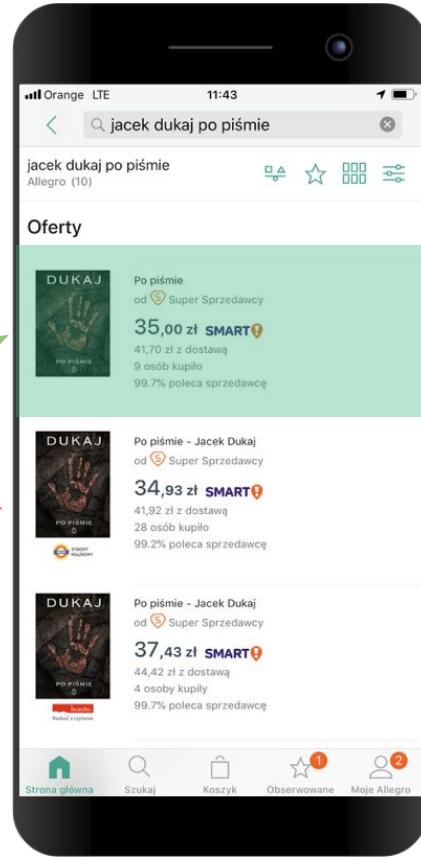
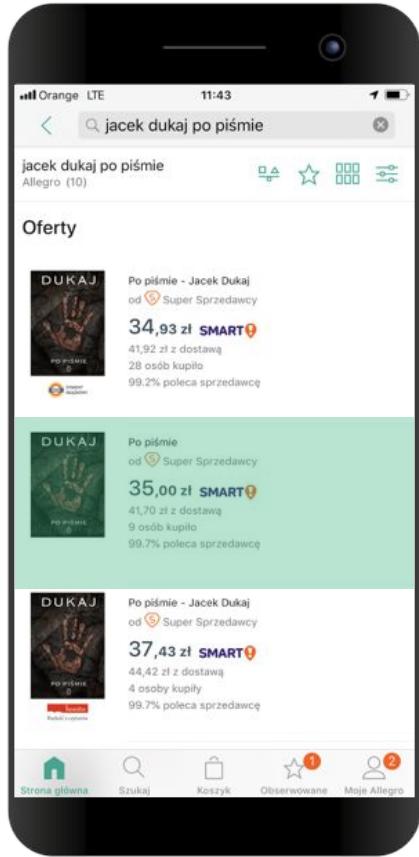
(0.19, 0.42, 0.75, ...)

(0.22, -0.35, 0.21, ...)

(0.12, 0.45, 0.11, ...)



$$nDCG \simeq 0.63$$



$$nDCG = 1$$

We cannot optimize NDCG directly using gradient methods, its derivative is zero or not defined.

## NeuralNDCG: Direct Optimisation of a Ranking Metric via Differentiable Relaxation of Sorting

Przemysław Pobrotyn\* and Radosław Białobrzeski\*

ML Research at Allegro.pl  
[mlr@allegro.pl](mailto:mlr@allegro.pl)

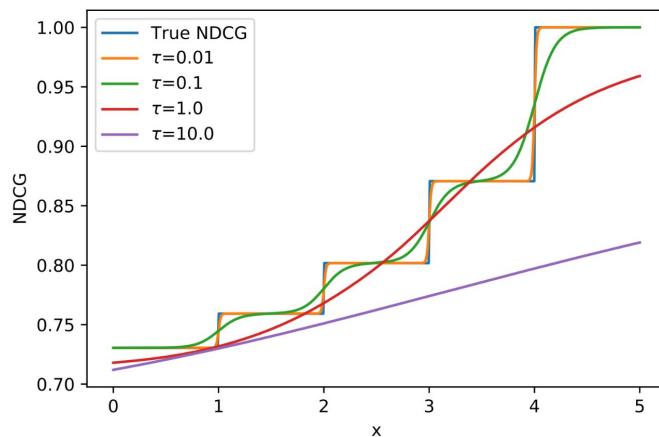


Fig. 2: Given ground truth  $y = [1, 2, 3, 4, 5]$  and a list of scores  $s = [1, 2, 3, 4, x]$ , we vary the value of the score  $x$  and plot resulting NDCG induced by the scores along with Deterministic NeuralNDCG for different temperatures  $\tau$ .

## Choose the right loss function

Metric-driven loss function: NDCGLoss2++

$$\text{NDCG} = \sum_i G_i / D_i$$

relevance gain / discount

$$G_i = \frac{2^{y_i} - 1}{\text{IDCG}} \quad 0 \leq y_i \leq 1 \quad D_i = \log_2(1 + i)$$

When two offers  $i$  and  $j$  are swapped,  
NDCG changes:

$$\Delta\text{NDCG}(i, j) = |G_i - G_j| \left| \frac{1}{D_i} - \frac{1}{D_j} \right|$$

$\Delta\text{NDCG}(i, j) > 0$  when  $i$  and  $j$  are in correct order

## Choose the right loss function

Metric-driven loss function: NDCGLoss2++

$$\text{NDCG} = \sum_i G_i / D_i$$

relevance gain / discount

$$G_i = \frac{2^{y_i} - 1}{\text{IDCG}} \quad 0 \leq y_i \leq 1 \quad D_i = \log_2(1 + i)$$

When two offers  $i$  and  $j$  are swapped,  
NDCG changes:

$$\Delta\text{NDCG}(i, j) = |G_i - G_j| \left| \frac{1}{D_i} - \frac{1}{D_j} \right|$$

$\Delta\text{NDCG}(i, j) > 0$  when  $i$  and  $j$  are in correct order

Let's construct a pairwise loss weighed by  $\Delta\text{NDCG}$

$$l(s, y) = \sum_{\substack{y_i > y_j \\ \text{scores vs. clicks}}} \text{sum over pairs} \quad \text{score difference} \quad -(s_i - s_j)$$

## Choose the right loss function

Metric-driven loss function: NDCGLoss2++

$$\text{NDCG} = \sum_i G_i / D_i$$

relevance gain / discount

$$G_i = \frac{2^{y_i} - 1}{\text{IDCG}} \quad 0 \leq y_i \leq 1 \quad D_i = \log_2(1 + i)$$

When two offers  $i$  and  $j$  are swapped,  
NDCG changes:

$$\Delta\text{NDCG}(i, j) = |G_i - G_j| \left| \frac{1}{D_i} - \frac{1}{D_j} \right|$$

$\Delta\text{NDCG}(i, j) > 0$  when  $i$  and  $j$  are in correct order

Let's construct a pairwise loss weighed by  $\Delta\text{NDCG}$

$$l(s, y) = \sum_{\substack{y_i > y_j \\ \text{sum over pairs}}} \text{scores vs. clicks}$$

$$\text{score difference} \\ [1 / (1 + e^{-(s_i - s_j)})]$$

logistic function  $> 1/2$  when  $s_i > s_j$

## Choose the right loss function

Metric-driven loss function: NDCGLoss2++

$$\text{NDCG} = \sum_i G_i / D_i$$

relevance gain / discount

$$G_i = \frac{2^{y_i} - 1}{\text{IDCG}} \quad 0 \leq y_i \leq 1 \quad D_i = \log_2(1 + i)$$

When two offers  $i$  and  $j$  are swapped,  
NDCG changes:

$$\Delta\text{NDCG}(i, j) = |G_i - G_j| \left| \frac{1}{D_i} - \frac{1}{D_j} \right|$$

$\Delta\text{NDCG}(i, j) > 0$  when  $i$  and  $j$  are in correct order

Let's construct a pairwise loss weighed by  $\Delta\text{NDCG}$

$$l(s, y) = - \sum_{\substack{y_i > y_j \\ \text{scores vs. clicks}}} \log_2 [1 / (1 + e^{-(s_i - s_j)})]$$

sum over pairs

score difference

logistic function  $> 1/2$  when  $s_i > s_j$

$$0 < -\log_2 [...] < 1 \text{ when } s_i > s_j$$

$$1 < -\log_2 [...] \text{ when } s_i < s_j$$

## Choose the right loss function

Metric-driven loss function: NDCGLoss2++

$$\text{NDCG} = \sum_i G_i / D_i$$

relevance gain / discount

$$G_i = \frac{2^{y_i} - 1}{\text{IDCG}} \quad 0 \leq y_i \leq 1 \quad D_i = \log_2(1 + i)$$

When two offers  $i$  and  $j$  are swapped,  
NDCG changes:

$$\Delta\text{NDCG}(i, j) = |G_i - G_j| \left| \frac{1}{D_i} - \frac{1}{D_j} \right|$$

$\Delta\text{NDCG}(i, j) > 0$  when  $i$  and  $j$  are in correct order

Let's construct a pairwise loss weighed by  $\Delta\text{NDCG}$

$$l(s, y) = - \sum_{\substack{y_i > y_j \\ \text{scores vs. clicks}}} \Delta\text{NDCG}(i, j) \log_2 [1 / (1 + e^{-(s_i - s_j)})]$$

sum over pairs

logistic function  $> 1/2$  when  $s_i > s_j$

$$0 < -\log_2 [...] < 1 \text{ when } s_i > s_j$$

$$1 < -\log_2 [...] \text{ when } s_i < s_j$$



# Lidia Wojciechowska

Research Engineer

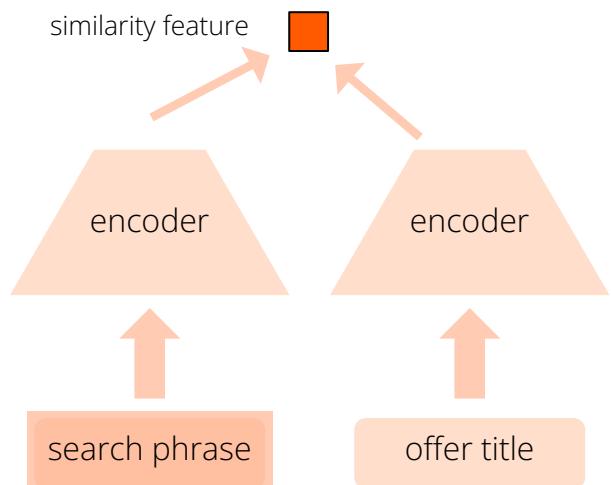


czwartek, 9 czerwca 2022

KNSI Golem

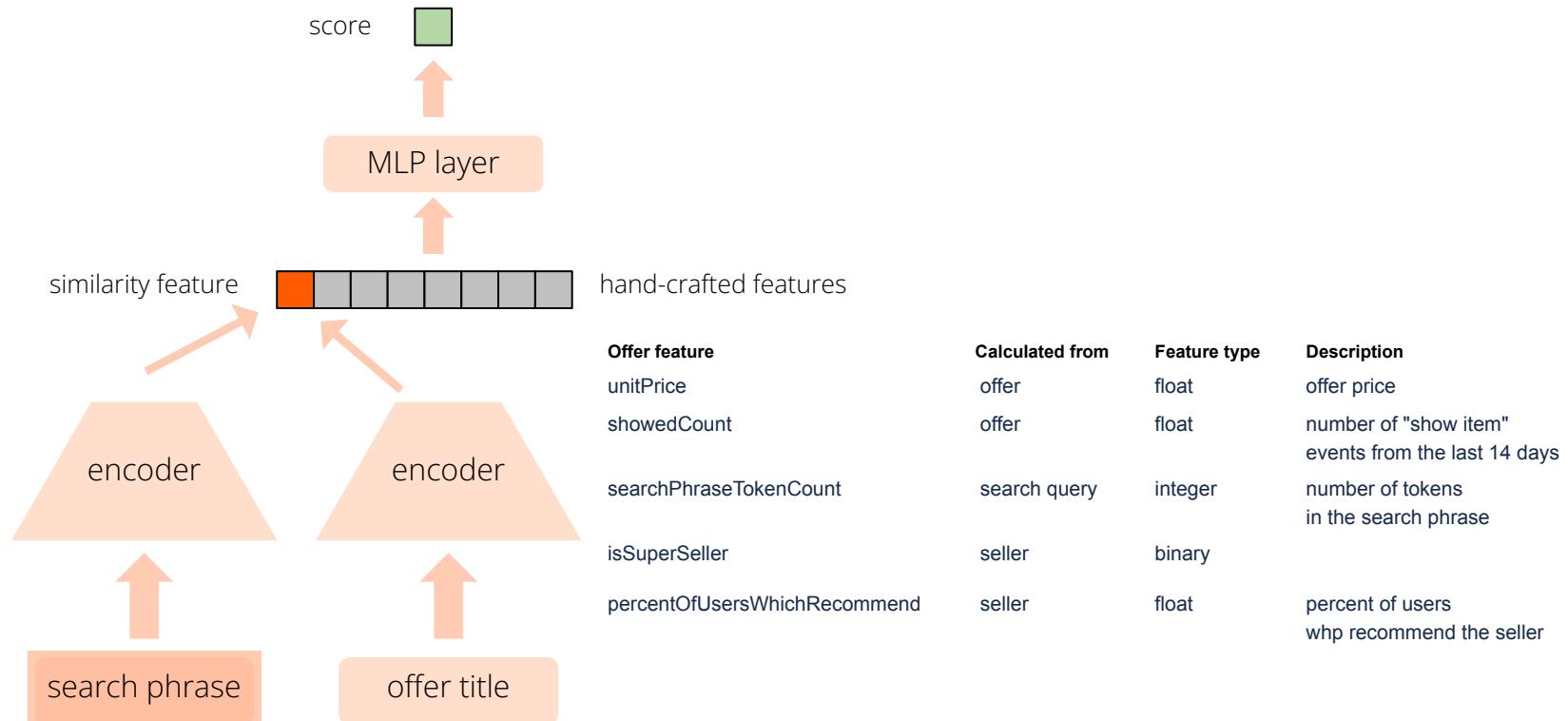
# Numeric features vs. deep textual features

We want both at the same time!



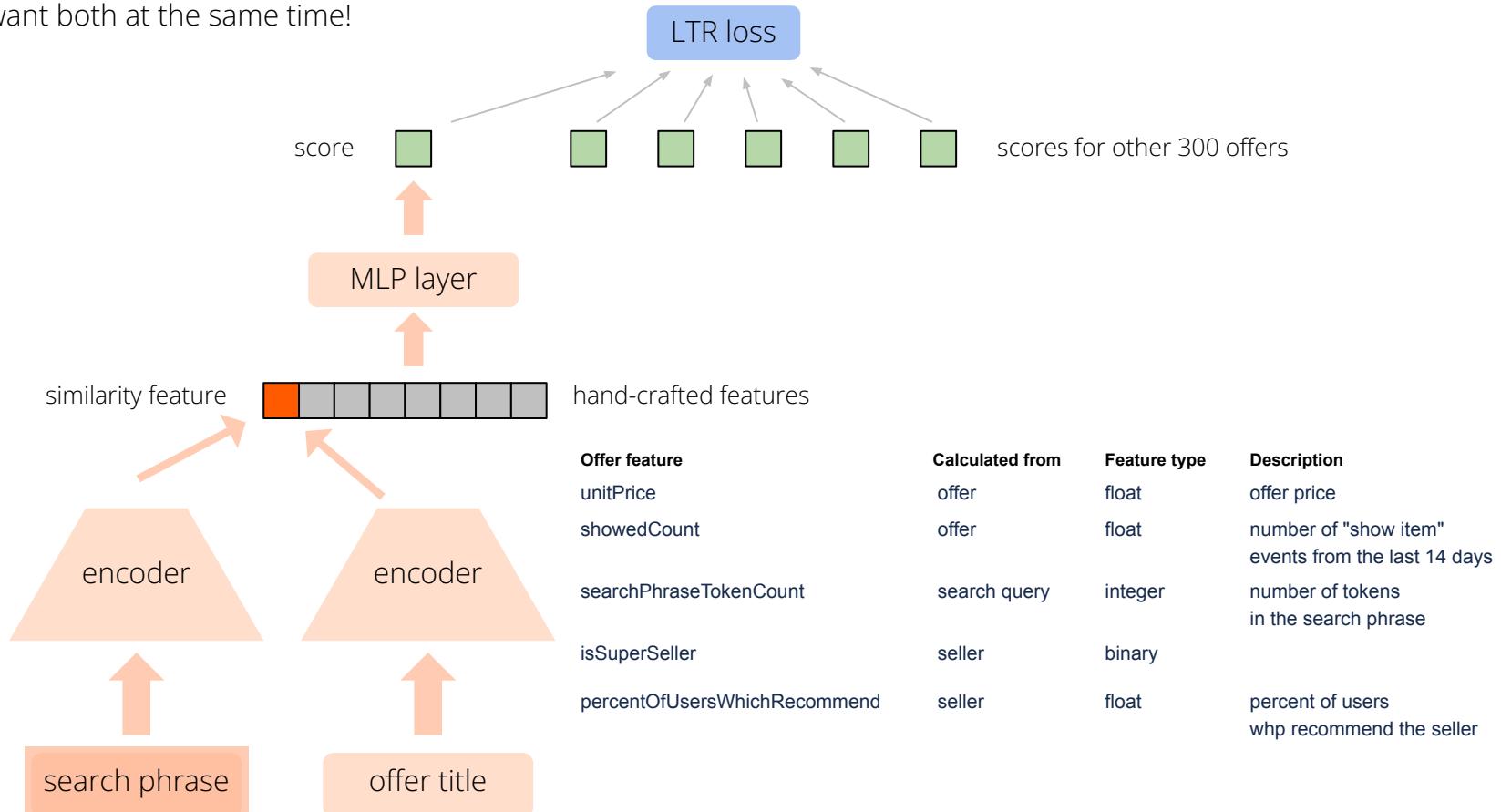
# Numeric features vs. deep textual features

We want both at the same time!



# Numeric features vs. deep textual features

We want both at the same time!



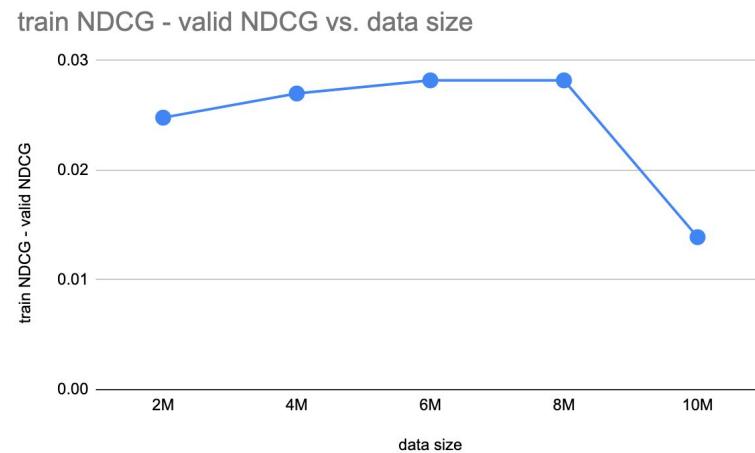
## Take advantage of scale

You need to use at least 10M examples to avoid overfitting

The reranking model is re-trained every week on 50M listings.

Infrastructure: GPUs @ Google Cloud

**more data > heavier model**



# allegro

---

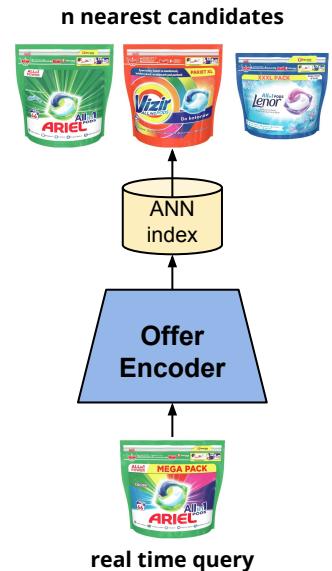
Summary

# The Essence of Retrieval

Define Similarity



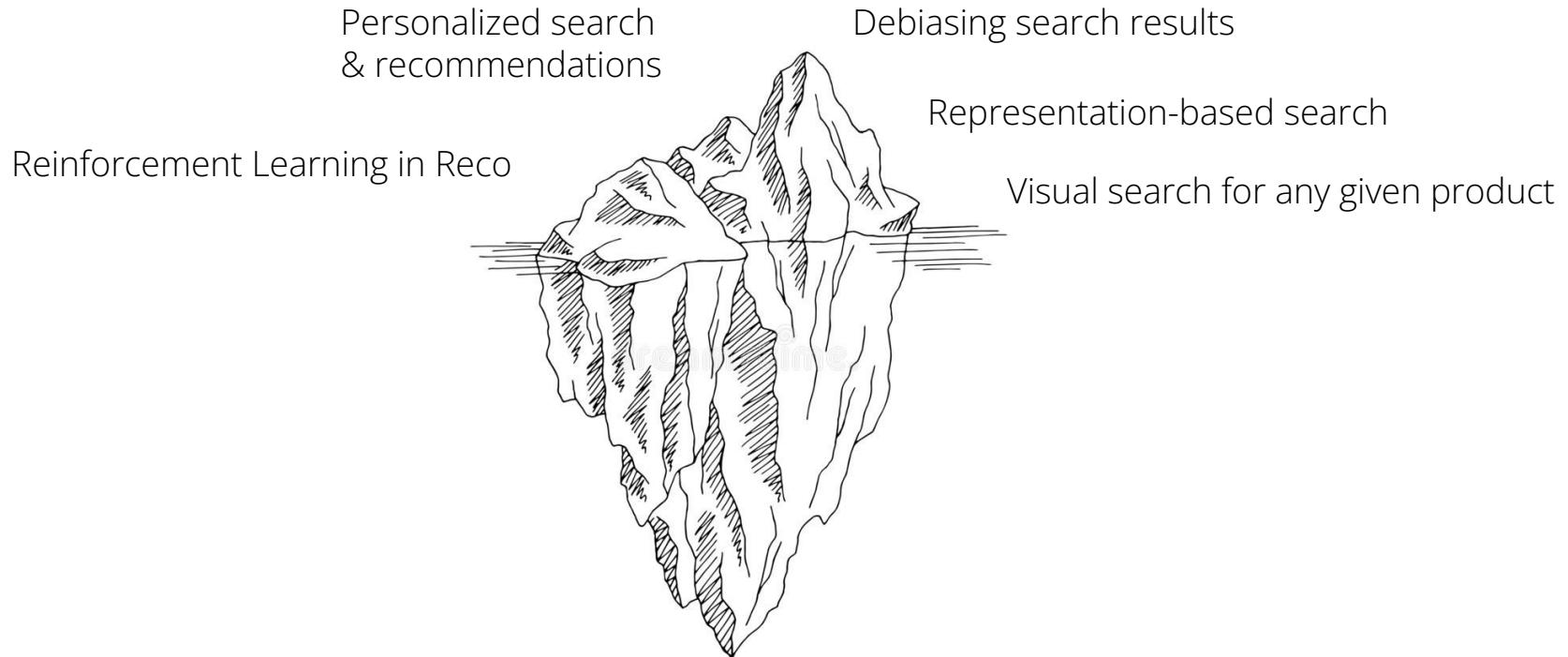
Generate candidates



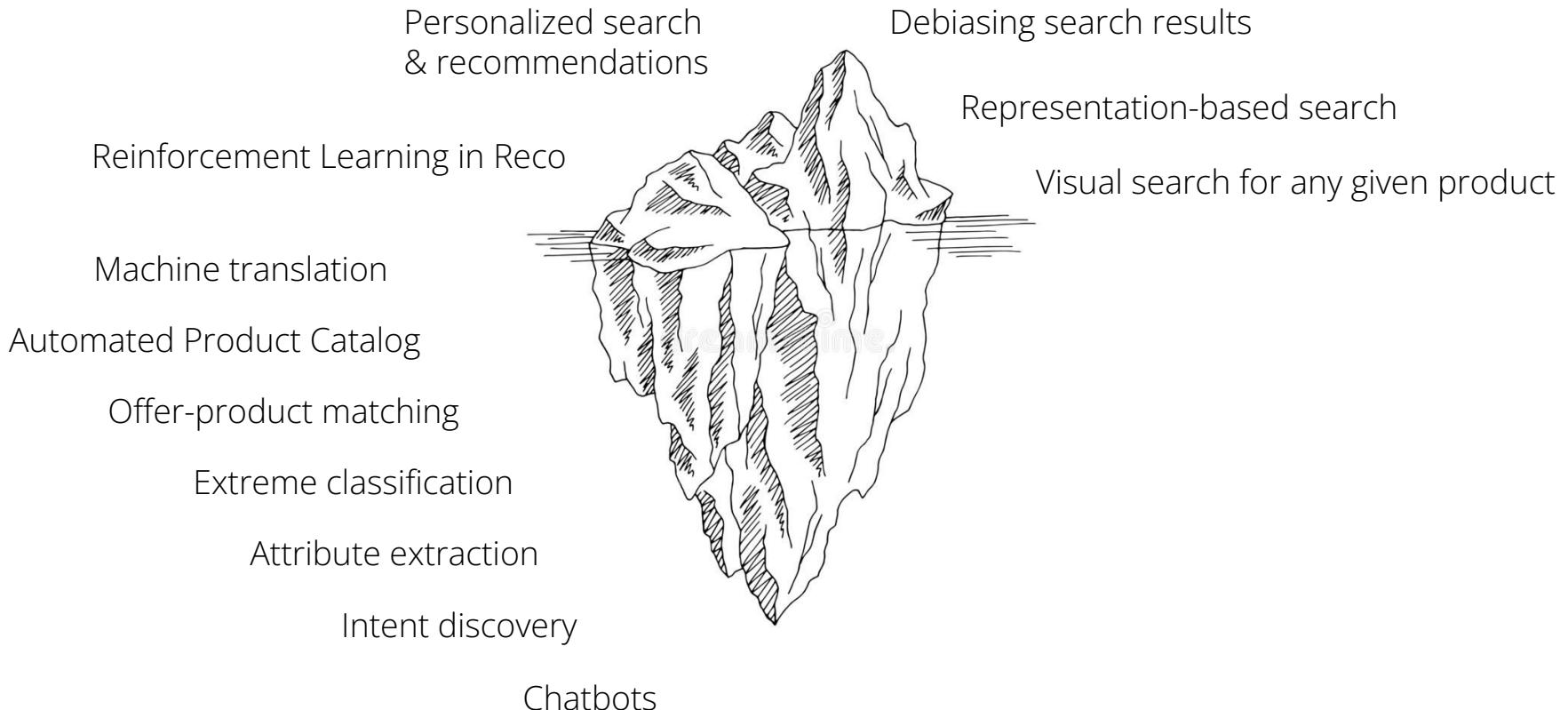
Rank the results

1. Usmarzne etui na karty Banki samopryklejne od Super Sklepsera do karty kredytowej. Zm. Nowy 27,99 zł smart@kaufland 32,00 zł do dostawy
2. Bla WODOODPORNE POKROWIEC NA TELEFON PLATE BAGNO 29,99 zł smart@kaufland 32,00 zł do dostawy
3. Biegone warranty ETUI DO MOTOROLA MOTO E5S Perfek Magnet + Szkło 15,99 zł smart@kaufland 17,00 zł do dostawy
4. Biegone warranty ETUI DO MOTOROLA MOTO E5 SMART MAGNET + SZKŁO Etui na telefon Wysok model Zm. Nowy 12,99 zł smart@kaufland 14,99 zł do dostawy
5. Biegone warranty ETUI DO MOTOROLA MOTO E5 SMART MAGNET + SZKŁO Etui na telefon Wysok model Zm. Nowy 12,99 zł smart@kaufland 14,99 zł do dostawy
6. Biegone warranty ETUI DO SAMSUNG GALAXY A50S Perfek Magnet + SZKŁO Etui na telefon Wysok model Zm. Nowy 24,99 zł smart@kaufland 27,99 zł do dostawy
7. Biegone warranty ETUI DO SAMSUNG GALAXY A50 Perfek Magnet + SZKŁO Etui na telefon Wysok model Zm. Nowy 24,99 zł smart@kaufland 27,99 zł do dostawy

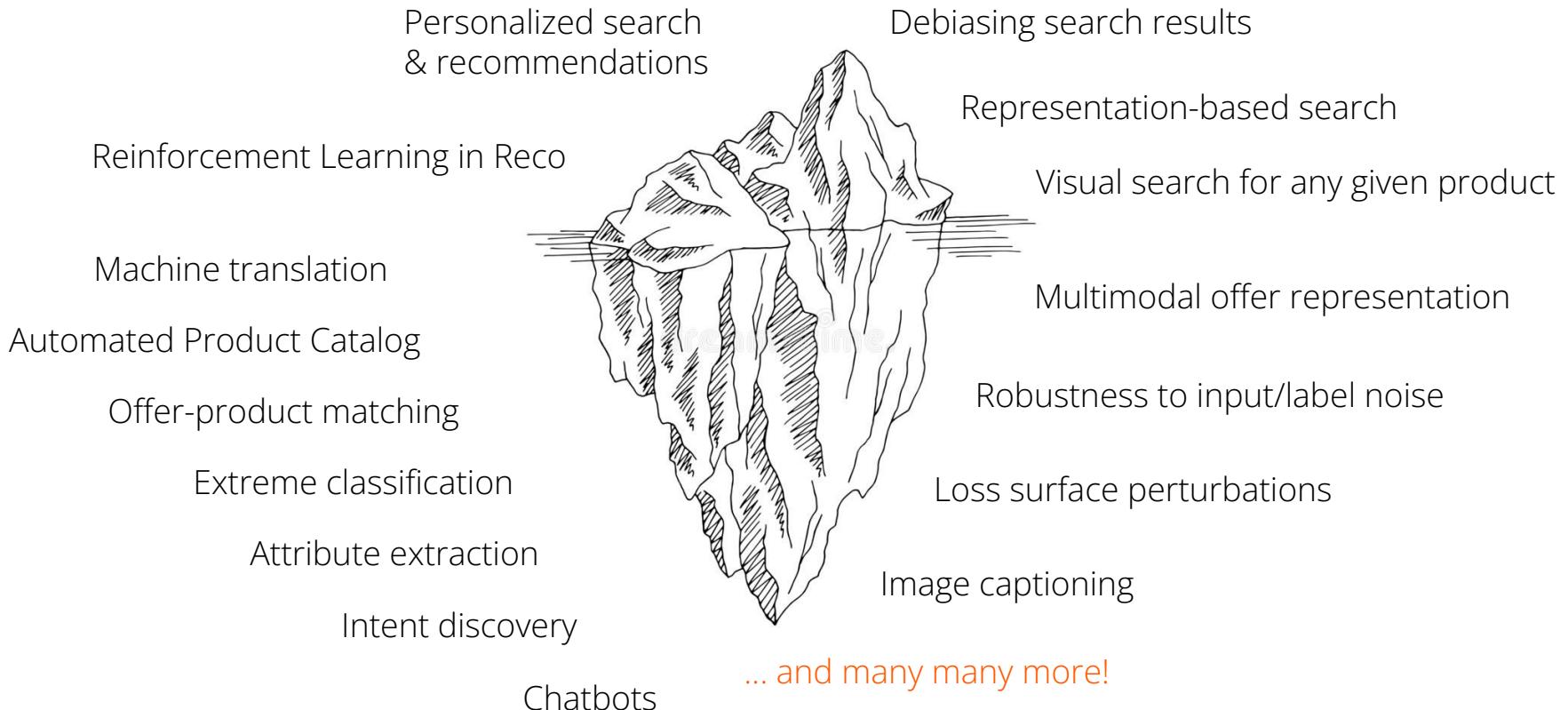
## Tip of an iceberg



## Tip of an iceberg



# Tip of an iceberg



# Machine Learning Research Team: 36 Engineers

Poznań / Warszawa / Kraków / Toruń / Prague

[ml.allegro.tech](http://ml.allegro.tech)



## We're making Allegro

...and everyone likes it!



## ML at scale

Data + infrastructure



## You'll never deploy alone

MLOps, Data Annotators,  
Developers, Analysts



\* BĘDE PROGRAMOWAĆ W SYLWE Patrzalka.pl

# Machine Learning Research Team: 36 Engineers

Poznań / Warszawa / Kraków / Toruń / Prague



**We're making Allegro**  
...and everyone likes it!



## Business problems motivate ML research

Share<sup>1</sup> more than 2x the size of the 2nd largest non-grocery retailer in Poland



## ML at scale

Data + infrastructure



## 20% research time

Research projects, weekly seminars



## You'll never deploy alone

MLOps, Data Annotators,  
Developers, Analysts



## Conferences & publications

We attend top venues (NeurIPS, ICML, etc.)  
and publish at ML conferences

### NeuralNDCG: Direct Optimisation of a Ranking Metric via Differentiable Relaxation of Sorting

Przemysław Pobrotyn<sup>\*</sup>  
ML Research at Allegro.pl  
przemyslaw.pobrotyn@allegro.pl

Radosław Białobrzeski<sup>†</sup>  
ML Research at Allegro.pl  
radoslaw.bialobrzeski@allegro.pl

#### ABSTRACT

Learning-to-Rank (LTR) algorithms & Information Retrieval metrics like NDCG are the most popular metrics for ranking. But derivatives are either undefined or zero-them unusable for gradient-based optimizers of learning approaches. Standard LTR loss functions are only lower-bounds, causing a mismatch between the evaluation criteria. In this paper, by proposing NeuralNDCG, we extend NDCG to DCG. Since DCG relies on the gradient, we obtain NeuralNDCG by reusing the neural network architecture. Then we obtain a new ranking loss function, curving approximation to the evalution gap. Finally, we evaluate our approach and introduce two variants of the proposed empirical evaluation shows that our proposed approach based on differentiable relaxation is competitive with the state-of-the-art.

CCS CONCEPTS  
• Information systems → Learning

KEYWORDS  
Learning-to-Rank  
Information Retrieval  
ACM Reference Format  
Przemysław Pobrotyn, Radosław Białobrzeski, Ireneusz Gawlik. 2023. NeuralNDCG: Direct Optimisation of a Ranking Metric via Differentiable Relaxation of Sorting. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023), held online, October 2–6, 2023, in Paris, France. Association for Computational Linguistics, 1–11. DOI: <https://doi.org/10.11695/EMNLP2023-1010>

### HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish

Robert Mroczkowski<sup>1</sup>, Piotr Rybak<sup>1</sup>, Alina Wróblewska<sup>2</sup>, Ireneusz Gawlik<sup>1,3</sup>

<sup>1</sup>ML Research at Allegro.pl

<sup>2</sup>Institute of Computer Science, Polish Academy of Sciences

<sup>3</sup>Department of Computer Science, AGH University of Science and Technology  
(firstname.lastname)@allegro.pl, alina@ipipan.waw.pl

#### Abstract

While most of the research related to analyzing and improving BERT-based models focuses on En-

### KLEJ: Comprehensive Benchmark for Polish Language Understanding

Piotr Rybak<sup>1</sup>, Robert Mroczkowski<sup>1</sup>, Janusz Tracz<sup>2</sup>, and Ireneusz Gawlik<sup>1,2</sup>

<sup>1</sup>ML Research at Allegro.pl,

ul. Grunwaldzka 182, 60-166 Poznań, Poland

(firstname.lastname)@allegro.pl

<sup>2</sup>AGH University of Science and Technology

Faculty of Computer Science, Electronics and Telecommunications,

Department of Computer Science

al. A. Mickiewicza 30, 30-059 Kraków, Poland

#### Abstract

In recent years, a series of Transformer-based models unlocked major improvements in general natural language understanding (NLU) tasks. However, the lack of benchmarks did not allow for a fair comparison of the proposed methods. However, such benchmarks are crucial for the development of languages. To alleviate this issue, we introduce a comprehensive multi-task benchmark for the Polish language understanding, accompanied by an extensive analysis. It consists of a diverse set of tasks, adapted from the well-known GLUE for named entity recognition, question-answering, textual entailment, and others. We also introduce a new sentiment analysis task for the e-commerce domain, named Allegro Reviews (AR). The proposed benchmark scheme promotes models that generalize to different NLU tasks; the benchmark in-

These models significantly improved the state-of-

the-art on numerous language understanding tasks.

Since then, the number of benchmarks increased and the new Transformer-based (Vaswani et al., 2017) models are being published every month to claim the latest state-of-the-art performance.

Such a pace of development would not be possible without standardized and publicly available NLU evaluation benchmarks. Among the most popular ones is the recently introduced GLUE (Wang et al., 2019a) consisting of a collection of tasks such as question answering, sentiment analysis, and text classification with test sets containing diverse set of domains. Some tasks come with numerous training examples, while others have limited training data. On top of that, for some tasks, the training set represents a different domain than the test set. This promotes models that learn general language representations and are effective at

Follow us!



[ml.allegro.tech](https://ml.allegro.tech)

[blog.allegro.tech](https://blog.allegro.tech)

[podcast.allegro.tech](https://podcast.allegro.tech)



# Thank you for your attention

---

[aleksandra.kurczab@allegro.pl](mailto:aleksandra.kurczab@allegro.pl)  
[jacek.szczerbinski@allegro.pl](mailto:jacek.szczerbinski@allegro.pl)

allegro

We are looking for  
**Research  
Engineers**



- Interesting and challenging projects (NLP, Vision, Ranking, RL)
- 20% of time for pure research
- Weekly seminars
- We attend top ML conferences (ACL, ICLR, NeurIPS, etc.)

[ml.allegro.tech](http://ml.allegro.tech)