User-centric Design and Evaluation of Exploratory Search and Recommender Systems

Dorota Głowacka

Exploratory Search and Personalisation

Research interests: interactive information retrieval, exploratory search, recommender systems, beyond-accuracy evaluation, virtual reality

Curious about my thoughts on **knitting** and **AI**? Check out the next issue of **Laine** magazine!

- What We Evaluate When We Evaluate Recommender Systems: Understanding Recommender Systems' Performance using Item Response Theory, **RecSys 2023**
- The Dark Matter of Serendipity in Recommender Systems, **CHIIR 2024**
- Sample, Nudge and Rank: Exploiting Interpretable GAN Controls for Exploratory Search, **IUI 2024**
- Behind the Scenes: Adapting Cinematography and Editing Concepts to Navigation in Virtual Reality, **CHI 2024**

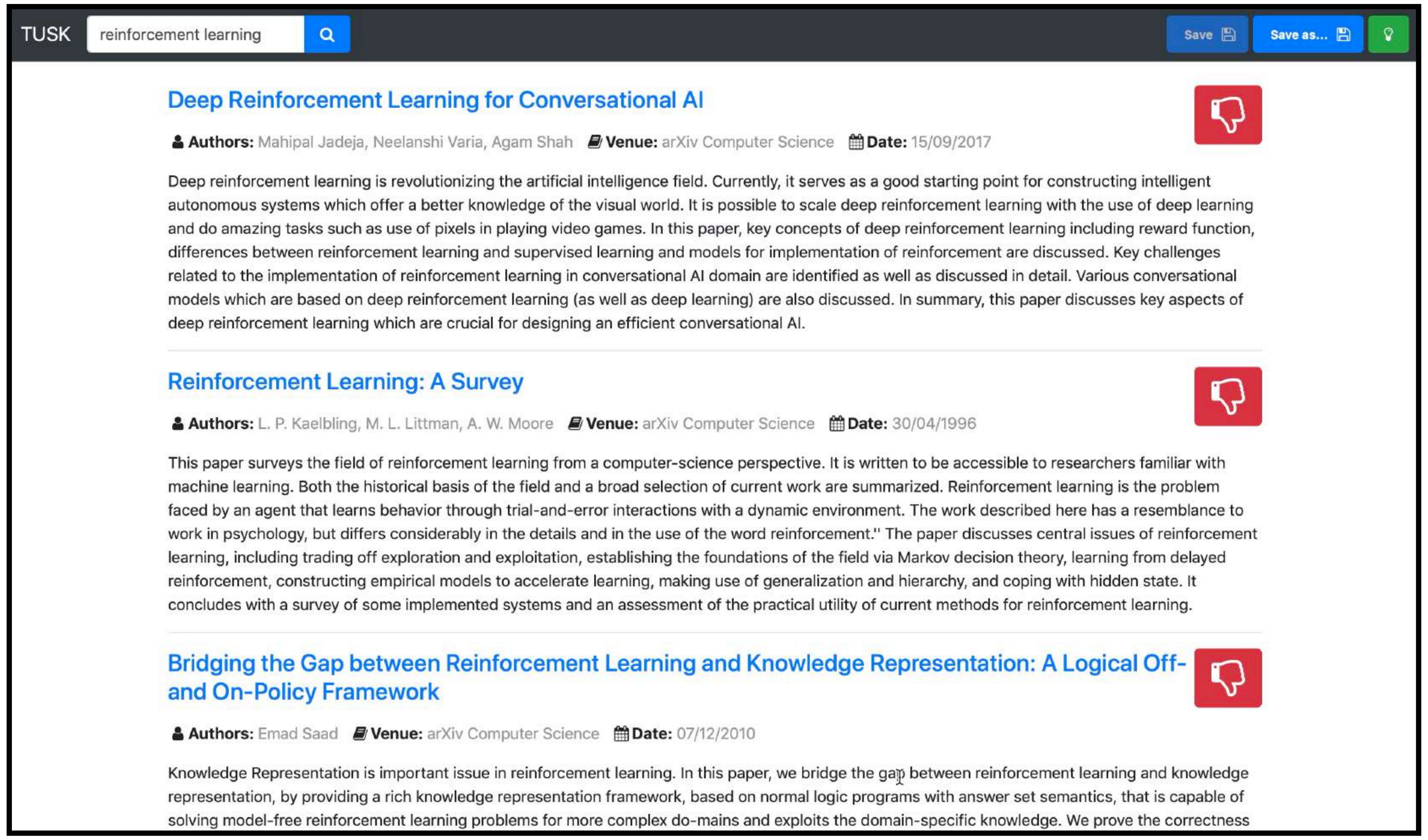
https://glowacka.org/

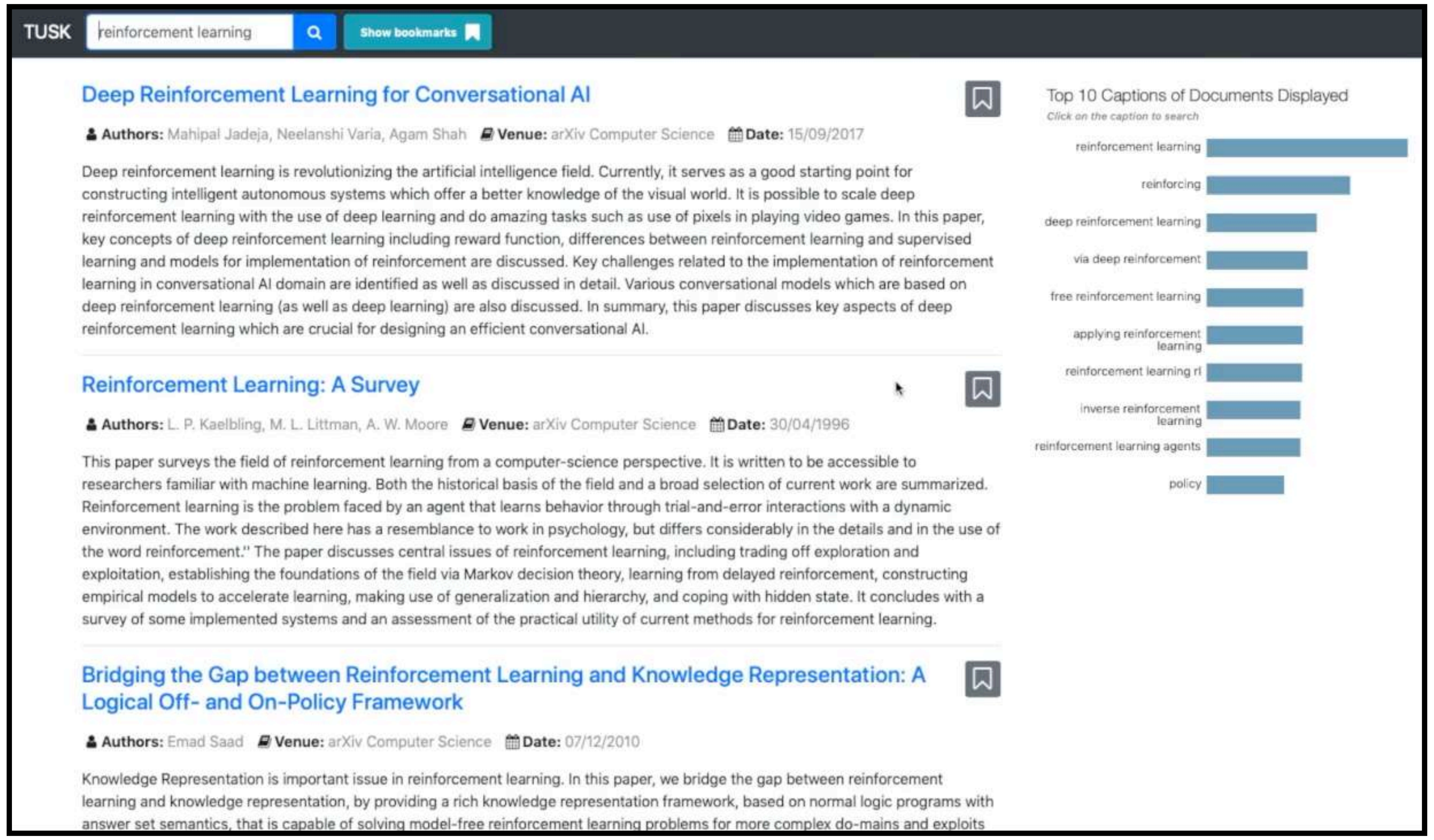
User-centric Design and Evaluation of Exploratory Search and Recommender Systems

Exploratory search: how do we support knowledge acquisition?

- Users performing exploratory search can be:
  - unfamiliar with their search domain
  - unsure how to achieve their goals
  - unsure what their goals are

- Methods to support users trying to acquire knowledge:
  - **System** learns from **user** (reinforcement learning - top)
  - **User** learns from **system** (result summarisation - bottom)





Query Suggestions as Summarization in Exploratory Search

Alan Medlar, Jing Li and Dorota Głowacka
University of Helsinki, Finland

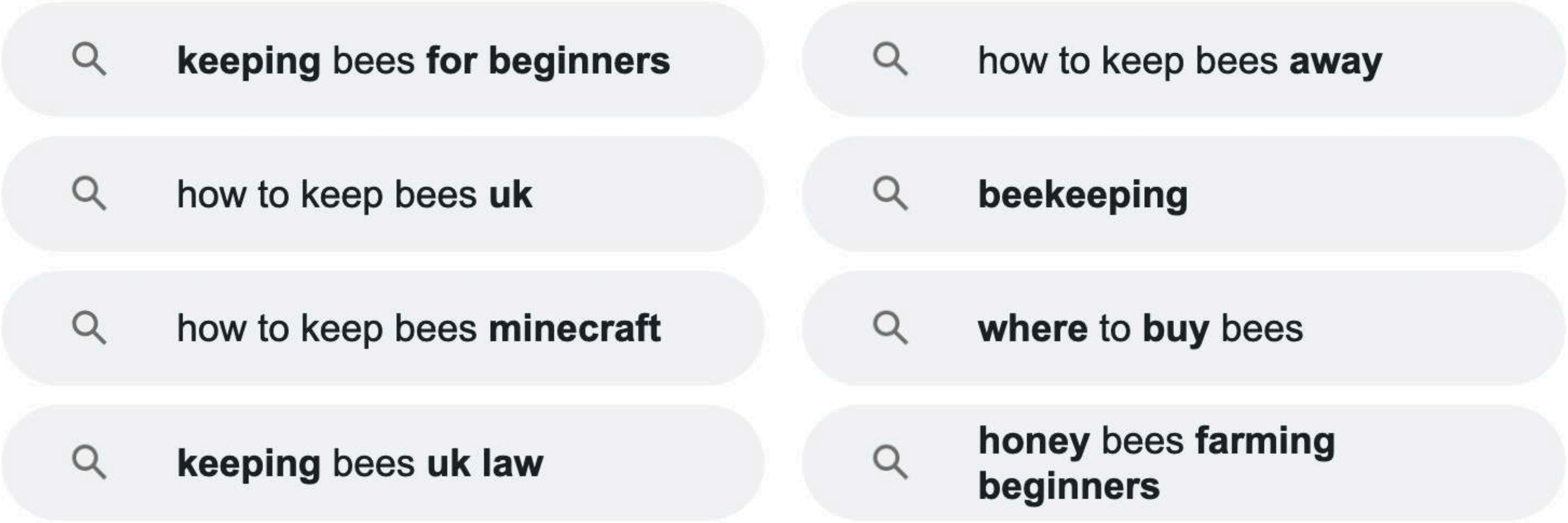
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Can query suggestions be used to support exploratory search?

- Exploratory search involves uncertainty w.r.t. search domain + information seeking goals
- Prior work focused on **search domain uncertainty**:
  - purchasing VOIP telephone
  - finding topically relevant newspapers articles
- Does it generalize to **scientific literature search**?
  - Cognitively demanding
  - Users highly uncertain about document relevance
  - Users scroll through far more search results

Query suggestions

- Query suggestions are queries displayed alongside search results:
  - follow-on queries
  - query reformulations
  - generated using query logs, pseudo-relevance feedback, concept hierarchies, etc.



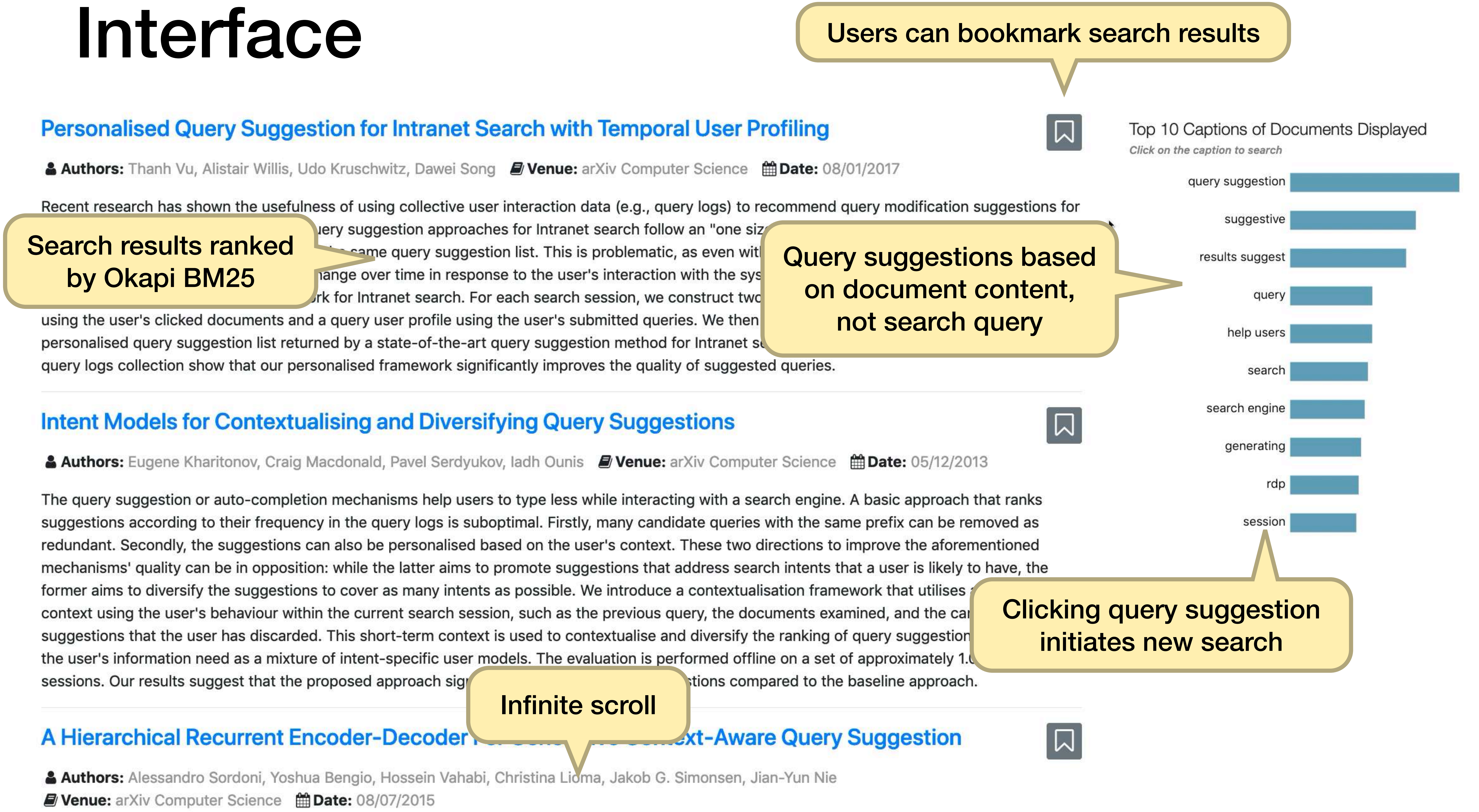
- Modern approaches based on word embeddings:

  - + identify semantically similar queries to the search query
  - **- users scroll through significantly more results during ES**

Our approach

- Query suggestions based on SERP embeddings (identify semantically similar queries to search results = **alternative queries**)
  - + independent of search query
  - + summarizes the contents of currently visible search results
  - + answers the question: **"*what am I looking at right now?*"**
- Search interface based on **infinite scroll**
  - + query suggestions change dynamically
  - + users can see when results are not relevant anymore

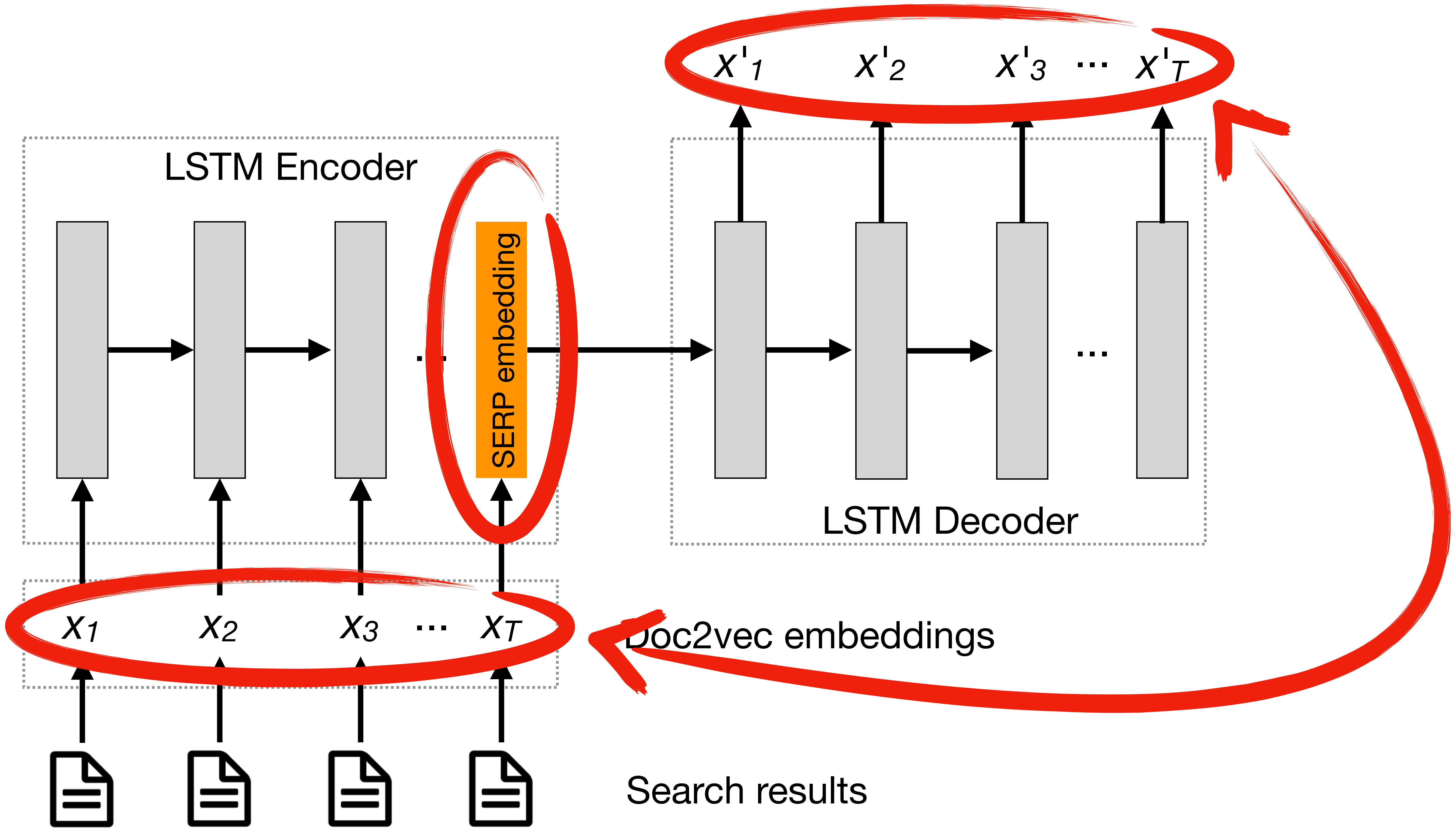
Interface



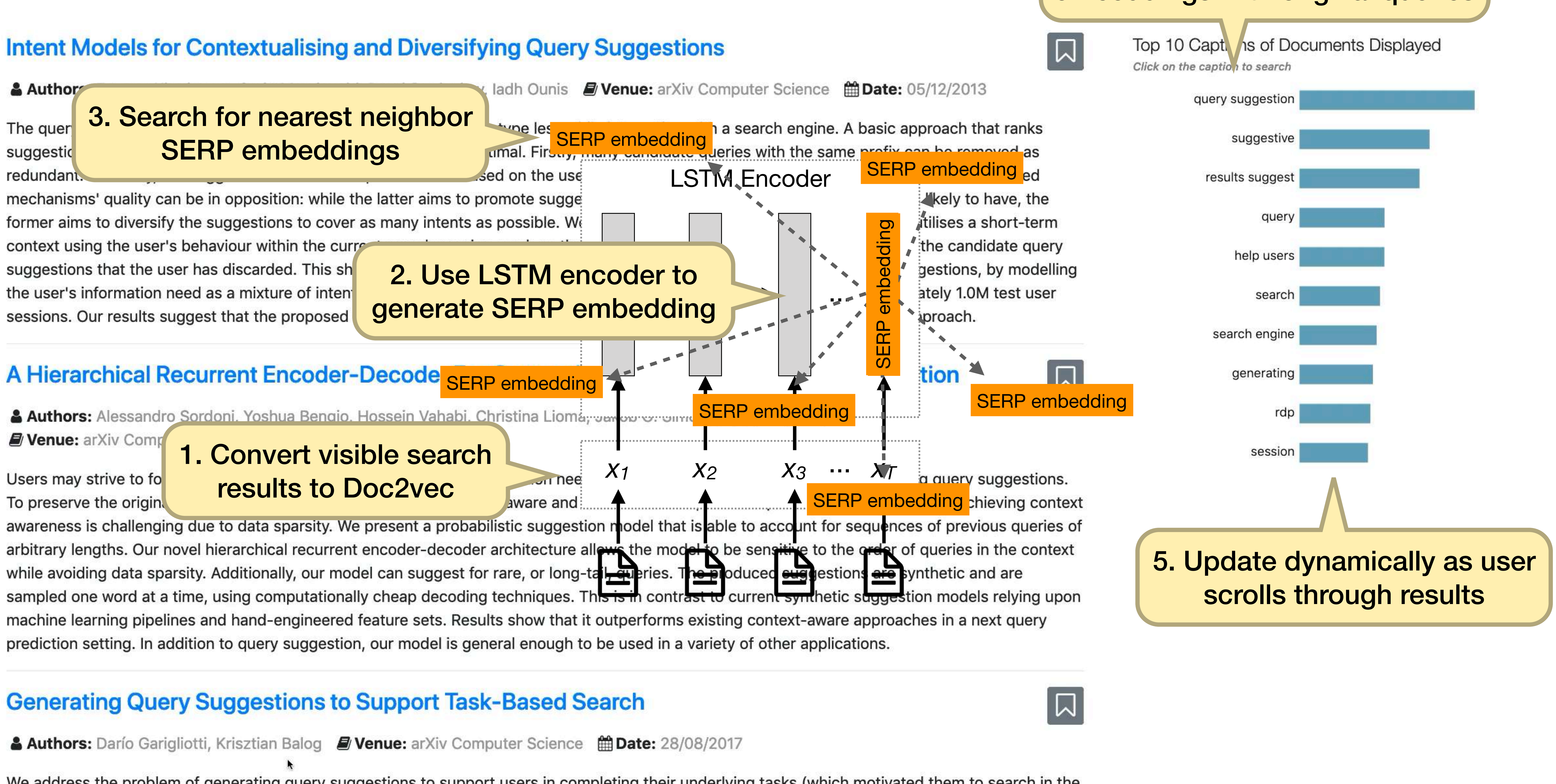
Users can bookmark search results
Personalised Query Suggestion for Intranet Search with Temporal User Profiling
Authors: Thanh Vu, Alistair Willis, Udo Kruschwitz, Dawei Song
Venue: arXiv Computer Science
Date: 08/01/2017
Recent research has shown the usefulness of using collective user interaction data (e.g., query logs) to recommend query modification suggestions for
uery suggestion approaches for Intranet search follow an "one siz
same query suggestion list. This is problematic, as even wit
ange over time in response to the user's interaction with the sys
rk for Intranet search. For each search session, we construct two
using the user's clicked documents and a query user profile using the user's submitted queries. We then
personalised query suggestion list returned by a state-of-the-art query suggestion method for Intranet s
query logs collection show that our personalised framework significantly improves the quality of suggested queries.
Search results ranked by Okapi BM25
Query suggestions based on document content, not search query
Top 10 Captions of Documents Displayed
Click on the caption to search
query suggestion
suggestive
results suggest
query
help users
search
search engine
generating
rdp
session
Intent Models for Contextualising and Diversifying Query Suggestions
Authors: Eugene Kharitonov, Craig Macdonald, Pavel Serdyukov, Iadh Ounis
Venue: arXiv Computer Science
Date: 05/12/2013
The query suggestion or auto-completion mechanisms help users to type less while interacting with a search engine. A basic approach that ranks
suggestions according to their frequency in the query logs is suboptimal. Firstly, many candidate queries with the same prefix can be removed as
redundant. Secondly, the suggestions can also be personalised based on the user's context. These two directions to improve the aforementioned
mechanisms' quality can be in opposition: while the latter aims to promote suggestions that address search intents that a user is likely to have, the
former aims to diversify the suggestions to cover as many intents as possible. We introduce a contextualisation framework that utilises
context using the user's behaviour within the current search session, such as the previous query, the documents examined, and the ca
suggestions that the user has discarded. This short-term context is used to contextualise and diversify the ranking of query suggestion
the user's information need as a mixture of intent-specific user models. The evaluation is performed offline on a set of approximately 1.
sessions. Our results suggest that the proposed approach sig
tions compared to the baseline approach.
Clicking query suggestion initiates new search
Infinite scroll
A Hierarchical Recurrent Encoder-Decoder
xt-Aware Query Suggestion
Authors: Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob G. Simonsen, Jian-Yun Nie
Venue: arXiv Computer Science
Date: 08/07/2015

SERP embedding model

- The SERP embedding model is an LSTM-based sequence-to-sequnce autoencoder
- LSTM encoder network outputs a **SERP embedding**
- Trained using ~70K SERPs from a corpus of CS papers from arXiv
- Used data augmentation to increase to ~300K SERPs



Alternative query generation



Expert assessment

- How well does our approach generalize to SERPs not present in the training data?

- Focused on situations where users are searching for documents related to multiple topics, e.g. "computer vision" + "autonomous driving"

- See paper for more details...

	Excess queries	Relevant/narrow	Relevant/in-between	Relevant/broad	Not relevant/too generic	Not relevant/unrelated	P@10	P@5
TF-IDF	2.4	308	88	375	301	178	0.62	0.77
χ^2	3.9	412	114	254	176	294	0.62	0.74
KL Divergence	2.4	301	97	403	311	138	0.64	0.76
Okapi BM25	2.5	301	86	441	284	138	0.66	0.81
SERP emb. (const.)	2.4	476	25	457	159	132	0.77	0.87
SERP emb.	2.0	553	45	391	118	143	0.79	0.85
Our method (const.)	2.4	468	28	488	144	121	0.79	0.89
Our method	**1.8**	505	55	478	117	95	**0.83**	**0.90**

User study

- **Baseline:** same system without query suggestions
- **Partipants:** 19 (8 female, 11 male) Computer Science students (8 MSc, 11 PhD)
- **Tasks and procedure:**
    - participants used both systems (within-subject study, system order was balanced)
    - write a short essay draft on an unfamiliar topic
    - document corpus was ~170K CS papers
    - 30 minutes max. search session + additional time to finalize draft
- **Data collected:**
    - After each system: SUS + modified ResQue
    - After both systems: post-experiment questionnaire + semi-structured interview
    - Search logs: queries issued, query suggestions, displayed documents, bookmarked documents, etc.
    - Essay grades: 1 (bad) - 5 (good), (Cohen's Kappa = 0.82)

Task performance and user behavior

- Participants used both systems, but when query suggestions were turned on:
    - they inspected **fewer documents per query** (7.8 vs 18.6, $p = 0.004$, Wilcoxon signed-rank)
    - they issued **more queries** overall (8.2 vs 3.7, $p = 0.0006$, Wilcoxon signed-rank)
    - they were **exposed to more documents** (55.3 vs 38.7, $p = 0.02$, Wilcoxon signed-rank)
    - they produced **higher quality essays** (3.37 vs 2.95, $p = 0.035$, Wilcoxon signed-rank)
- No difference in number of bookmarks
- Query suggestions account for ~50% of issued queries

Usability

- **SUS:** 76.8 vs 71.2 (p=0.136, Wilcoxon signed-rank)
- **ResQue:** 83.2 vs 67.8 (p=0.001, Wilcoxon signed-rank)

QS	B	p	Question
4.0	3.5	0.01	1. The documents recommended to me matched what I was searching for
3.7	3.1	0.0139	2. The system helped me discover new documents
3.6	2.8	0.1305	3. The documents recommended to me are diverse
3.6	2.8	0.0164	4. The system helped me find the ideal documents
4.2	4.1	0.7054	5. I became familiar with the system very quickly
3.8	2.8	0.0189	6. I found it easy to notice if the search results were not correct any more
3.9	3.2	0.011	7. I felt confident to modify my query
3.6	3.2	0.0522	8. Using the system to find what I like is easy
3.7	3.7	0.7192	9. I found it easy to re-find documents I had been recommended before
3.7	3.1	0.0079	10. The system gave me good suggestions
3.8	3.5	0.07	11. The system made me confident about the documents I bookmarked
3.6	3.2	0.0374	12. Overall, I am satisfied with the system

User perception

- Users preferred query suggestions being present during exploratory search
- Query suggestions reassured users that search results were relevant to their search goals
- ... but only half thought they were good followup queries

Prop. agree	p-value	Question
0.947	7.6e-05	1. Which system did you prefer to use?
0.737	0.063	2. I found it easier to perform the search with query suggestions
0.737	0.063	3. I found it easier to write the essay draft with query suggestions
0.895	0.0007	4. The labels of the query suggestion interface are clear
0.632	0.359	5. The bars of the query suggestion interface are clear
0.474	1.0	6. The query suggestions should be an optional function
0.895	0.0007	7. The query suggestions enhanced my search session
0.895	0.0007	8. The query suggestions were related to my search results
0.842	0.004	9. The query suggestions reassured me that my search results were relevant to my search goals
0.737	0.063	10. The query suggestions provided a good summary of my search results
0.526	1.0	11. The query suggestions provided good followup queries
0.0	3.8e-06	12. The query suggestions were distracting
0.158	0.004	13. The query suggestion animations were distracting
0.579	0.647	14. The system's confidence in each query suggestion was clearly indicated
0.895	0.0007	15. The system was better with the query suggestions than without
0.0	3.8e-06	16. There were too many query suggestions

Summary

- Previous studies related to using query suggestions in exploratory search were related to less cognitively demanding search tasks
- In scientific literature search, user behavior and perception results showed that query suggestions impacted users' search process
- Used as follow-on queries and for summarization

Sample, Nudge, and Rank: Exploiting Interpretable GAN Controls for Exploratory Search

Yang Liu, Alan Medlar and Dorota Głowacka
University of Helsinki

UNIVERSITY OF HELSINKI

Motivations

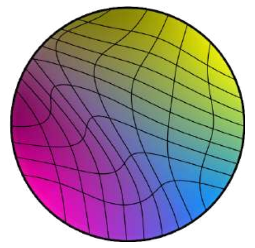
- Exploratory search is challenging

Not so sure what I need…

Uncertainty

Learning something new, but not so sure what I will learn…

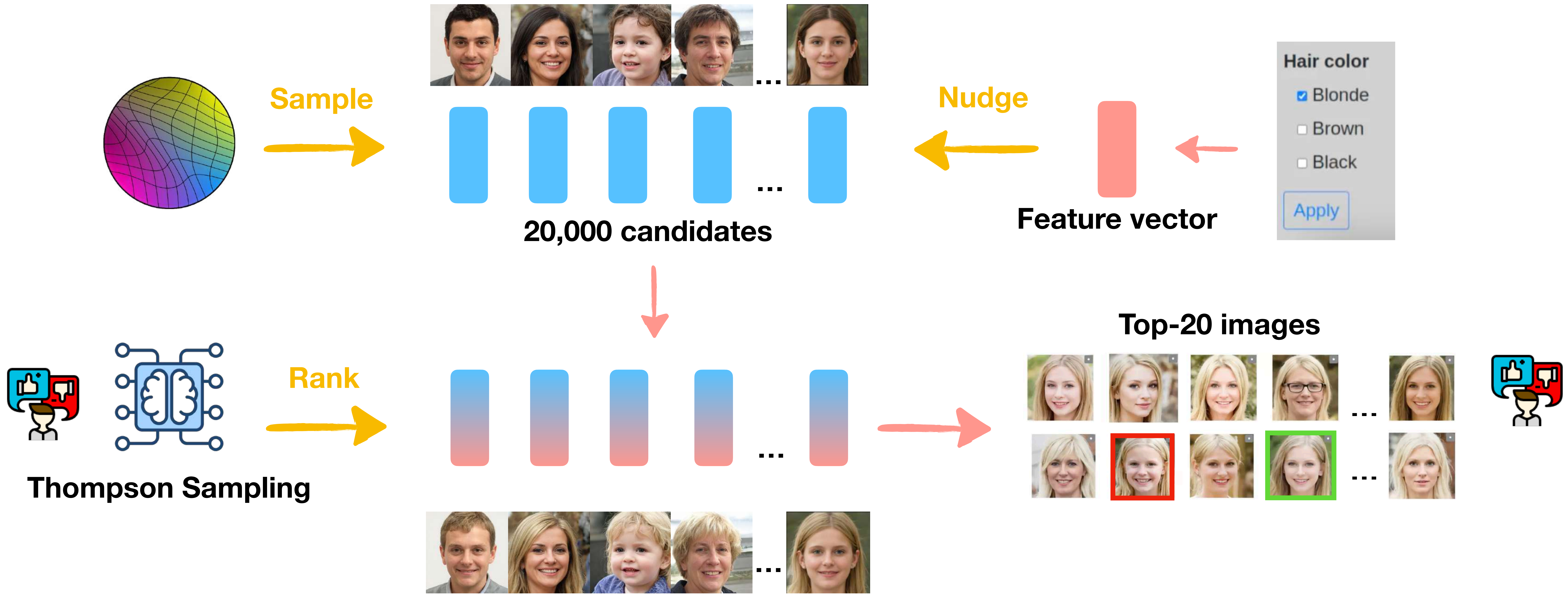
Open-endedness

- GANs present numerous opportunities
  - Expanded search space 
  - Interpretable GAN controls → truly satisfy users' search goals

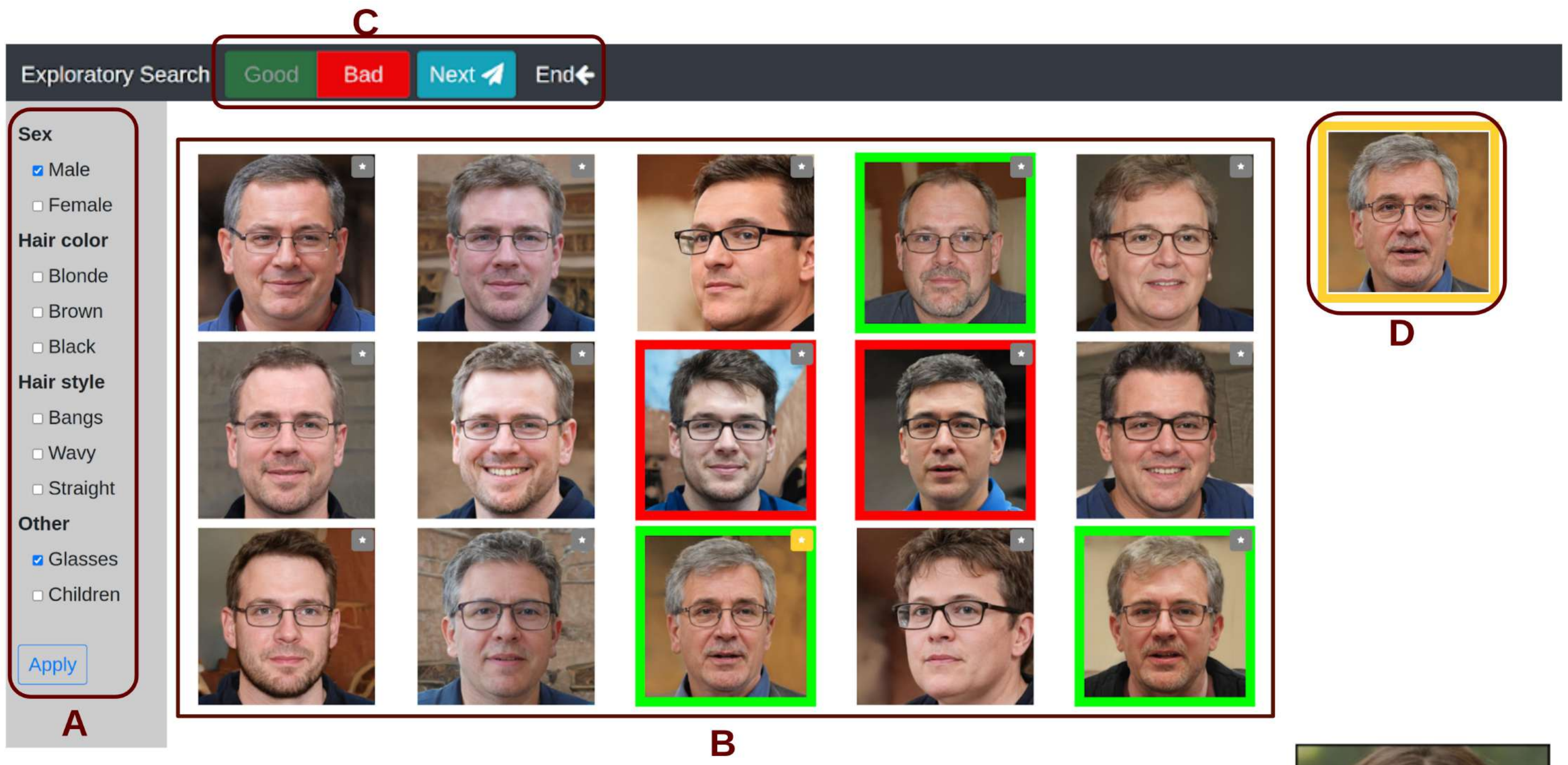
The number of unique images generated is exceptionally high

Sample, Nudge, and Rank

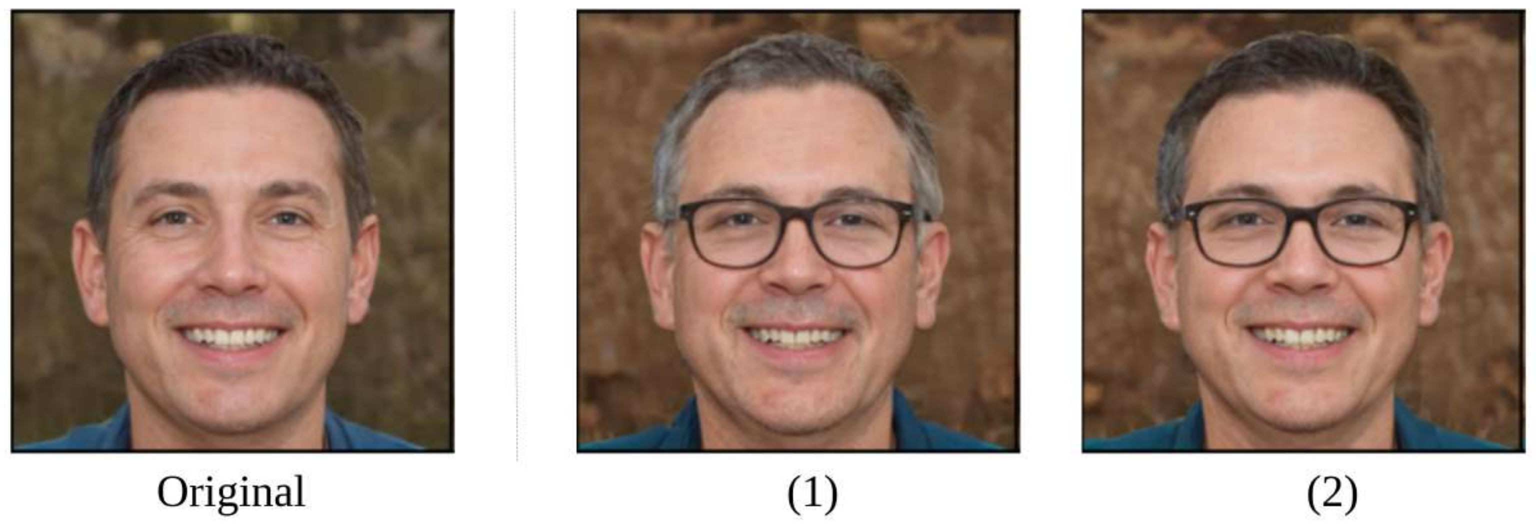
- Two interaction mechanisms: **faceted search** + **relevance feedback**



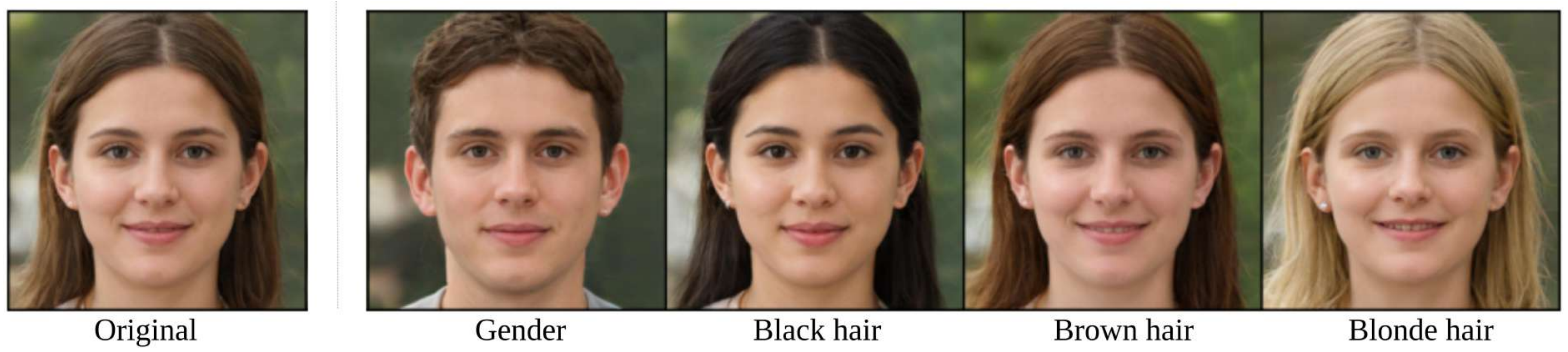
User Interface



Decoupling:



Nudging:



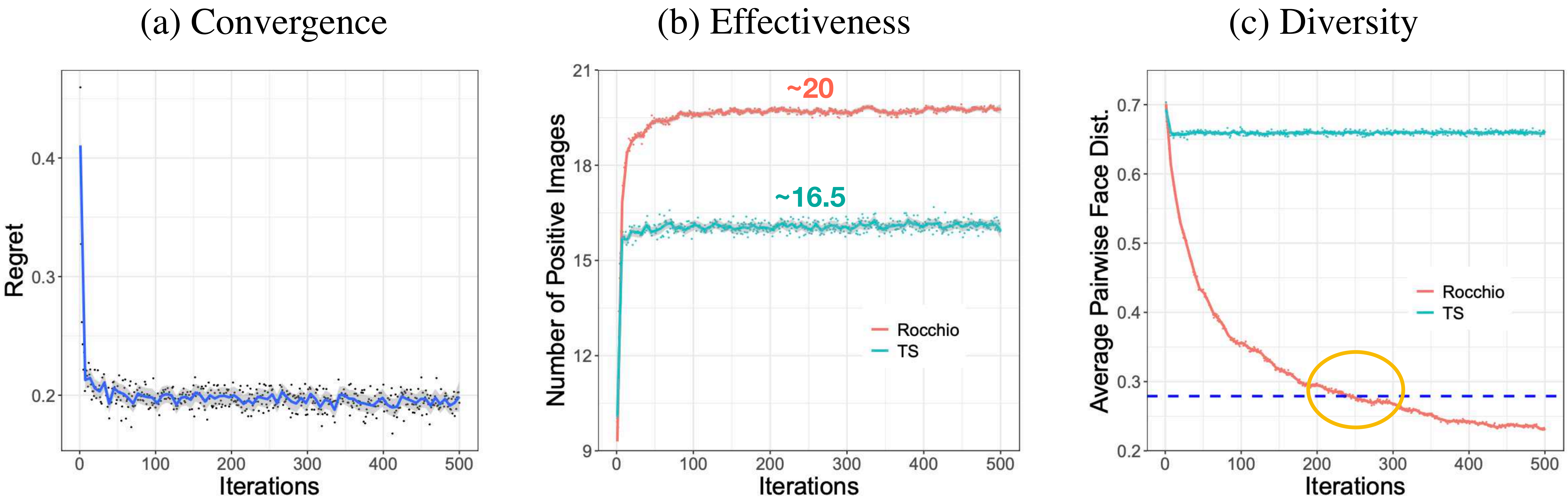
Evaluation

- Simulation study + User experiment
- Baseline approach: Rocchio algorithm[1]
  - Sampling images close to the centroid of relevance feedback
  - Only positive feedback + no facets
  - Warm start: selecting one seed image from 100 random images

1. Ukkonen et al. "Generating images instead of retrieving them: Relevance feedback on generative adversarial networks." *SIGIR*. 2020

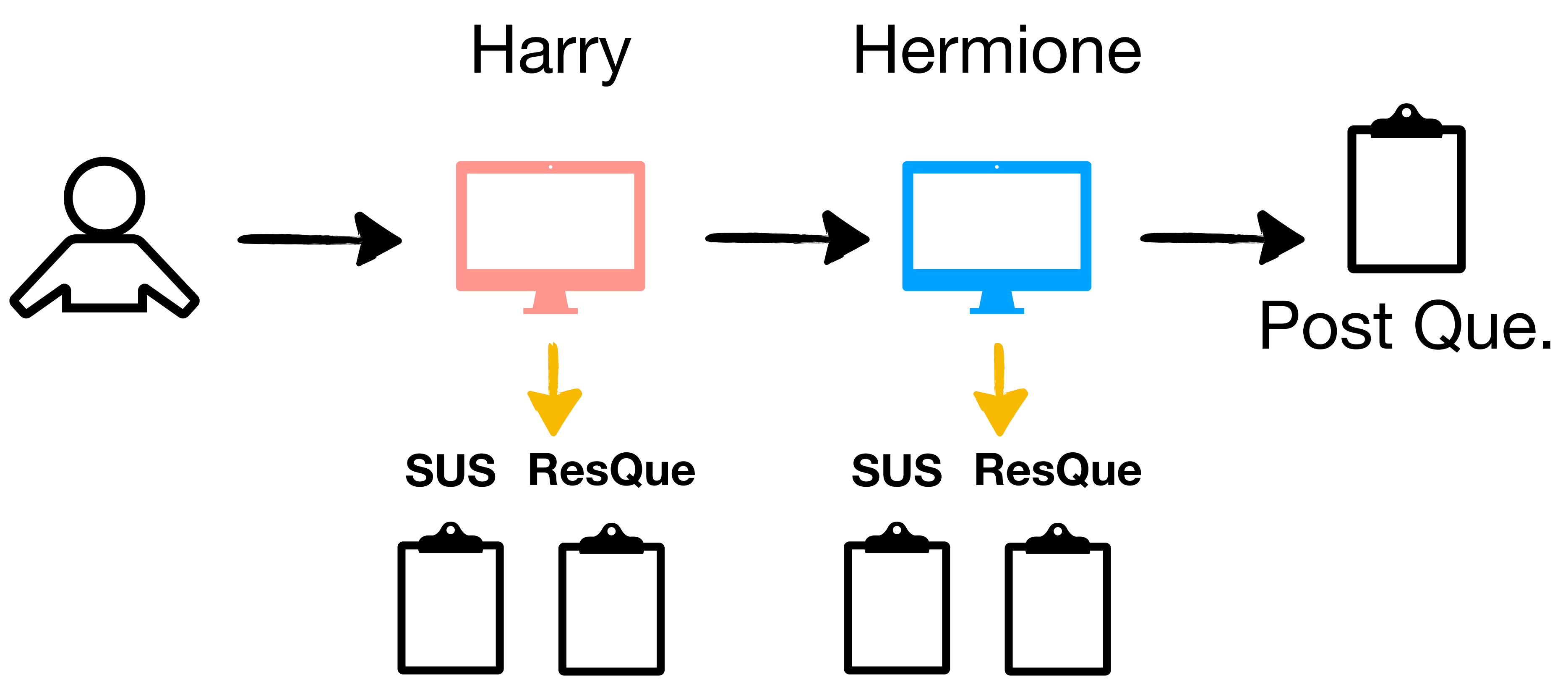
Evaluation: Simulations

Finding 1: Our approach efficiently **adapts to user preferences**, while **preserving** a high-level of **image diversity**



Evaluation: User Experiments

- 30 study participants
- Exploratory search task: casting for a fake Harry Potter movie
  - Two tasks: Harry Potter, Hermione Granger
- Within-subject study



Movie Plot

The movie takes place when Harry Potter and Hermione Granger are around 30 years old. Harry was framed for a crime he did not commit and was imprisoned in Azkaban (a prison for wizards). At the start of the movie, Harry escapes from Azkaban. His time in prison has been tough. Harry is angry and wants revenge. Hermione is now a teacher of the dark arts at Hogwarts, but is unhappy and disillusioned with the world of magic.

Evaluation: User Experiments

Finding 2: No significant difference found in **overall system usability** and **system satisfaction**

- Users of our system examined significantly fewer images (106.7 vs 188.7)

Finding 3: **23/30** participants preferred our system over baseline

Prop.	P-value	Question
0.767*	0.005	1. Which system did you prefer to use for finding an actor if you are a casting director?
0.733*	0.016	5. Which interface did you prefer?

- Diverse yet better recommendation provided in our system
- Very similar faces that were difficult to distinguish in baseline

Summary

- A novel approach to support exploratory search of GANs
- Implementation of faceted search and relevance feedback in GAN search
- Better performance of our approach in both simulations and the user study

User-centric Design and Evaluation of Exploratory Search and **Recommender Systems**

On the Negative Perception of Cross-domain Recommendations and Explanations

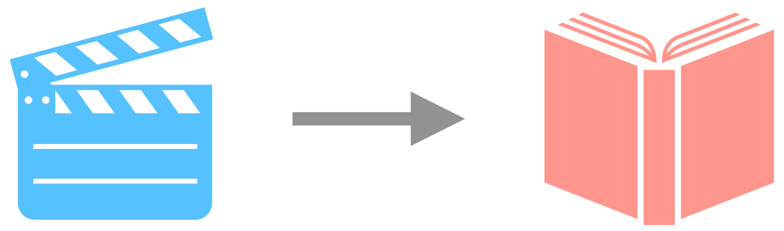
Denis Kotkov, Alan Medlar, **Yang Liu**, and Dorota Głowacka
University of Helsinki

UNIVERSITY OF HELSINKI

Motivations (1)

How do users perceive
cross-domain recommendations?

- Cross-domain recommendation
  - Knowledge sharing between source and target domains
  - Data sparsity, cold-start problems



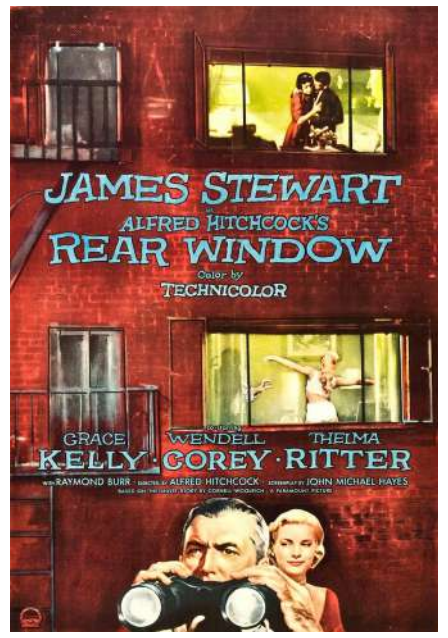
  - Higher Precision, Recall, MRR etc.
  - No prior studies on user perceptions

Motivations (2)

- Recommendation **explanations**
  - Increasing users' interest
  - Affecting user perceptions
  - Not explored in cross-domain settings

> "*…cross-domain models with explainability would be beneficial to **improve the transparency, persuasiveness, and trustworthiness** of CDRs.*"
>
> — A survey article by [Zang et al.](#) TOIS. 2022.

How do users perceive
CDR explanations?



Rear window

Your prediction is based on how MovieLens thinks you like this aspects of the film[1]:

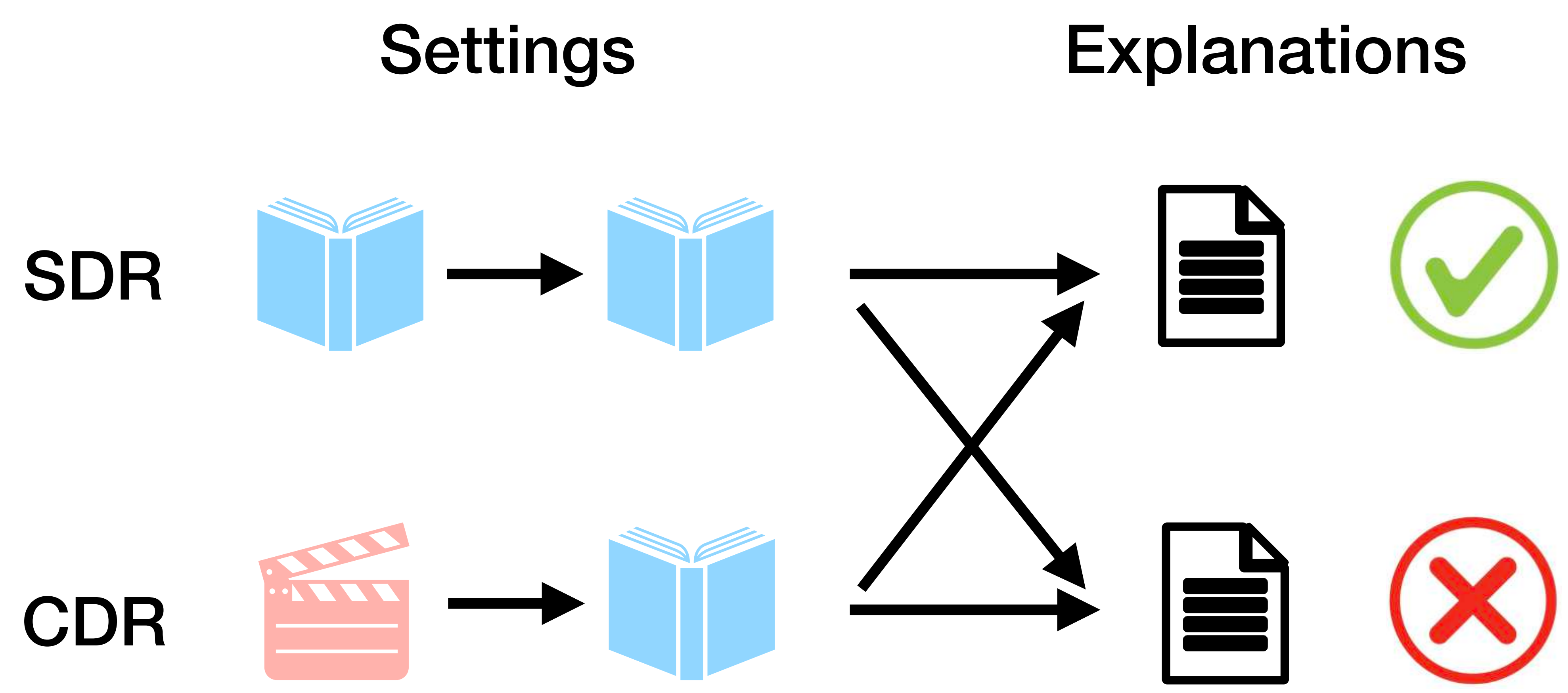
Relevance		Preference
	Hitchcock	☆☆☆☆☆
	Classic	☆☆☆☆☆
	Murder	☆☆☆☆

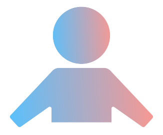
Explanations in SDR

1. Example adapted from: Vig et al. "Tagsplanations: explaining recommendations using tags" *IUI*. 2009

Study Design (1)

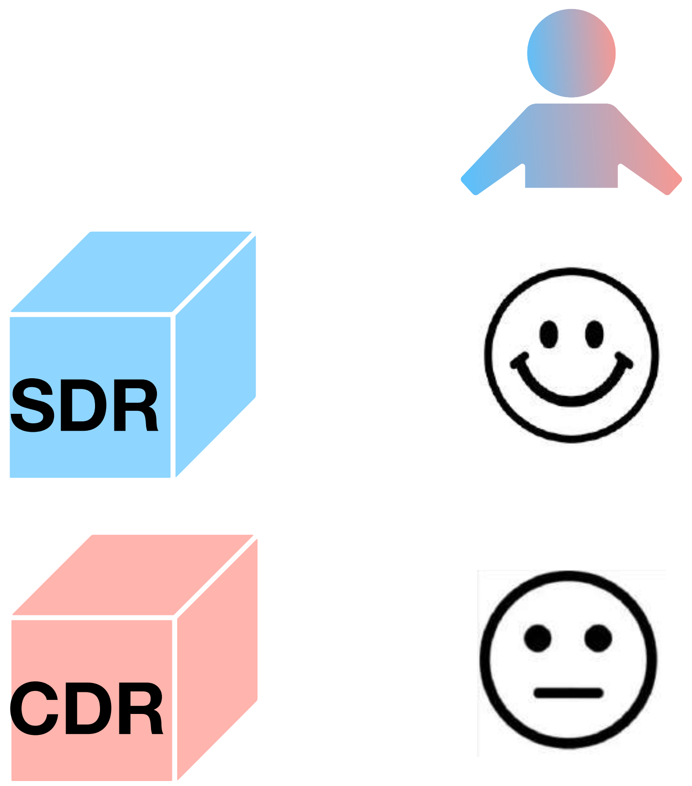
- Information availability for recommendations must be **unambiguous**



- **4** scenarios
- Between-subject design: each  only situated in one scenario

Study Design (2)

- <u>Recommendation quality</u> must be **consistent** across all scenarios
  - We wanted to focus on cross-domain recommendation **settings**



Possibilities

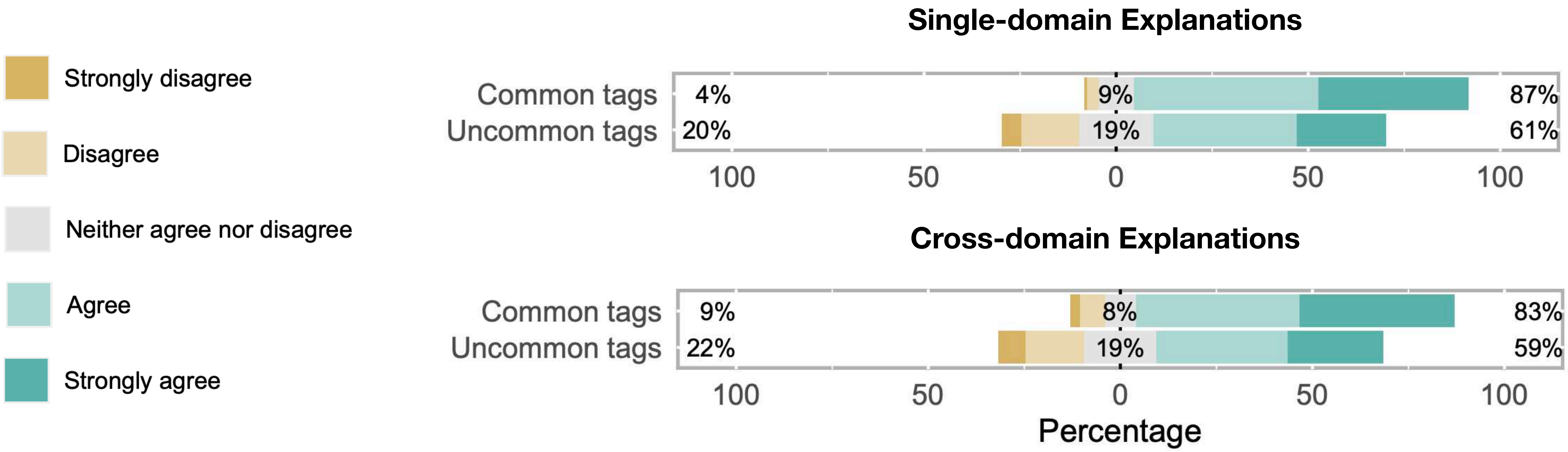
1. Bad CDR algorithm
2. CDR settings

...

- Hence, we generate random recommendation lists (#4,209)
  - random, diverse, balanced

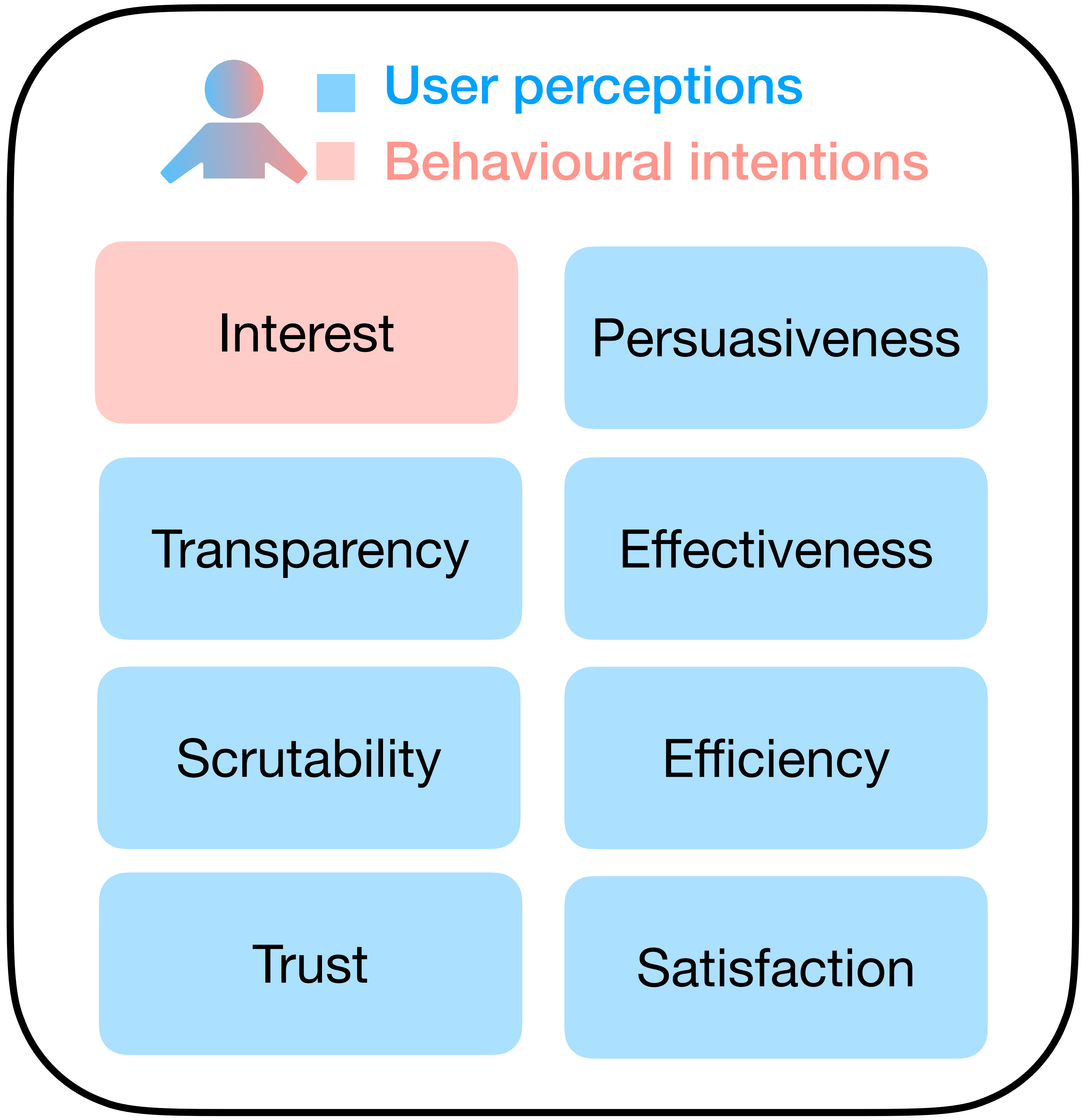
Study Design (3)

- Generated explanations must credibly **justify** recommendations
- Explanations: common + uncommon tags ← Tag Genomes
- Participants to help!



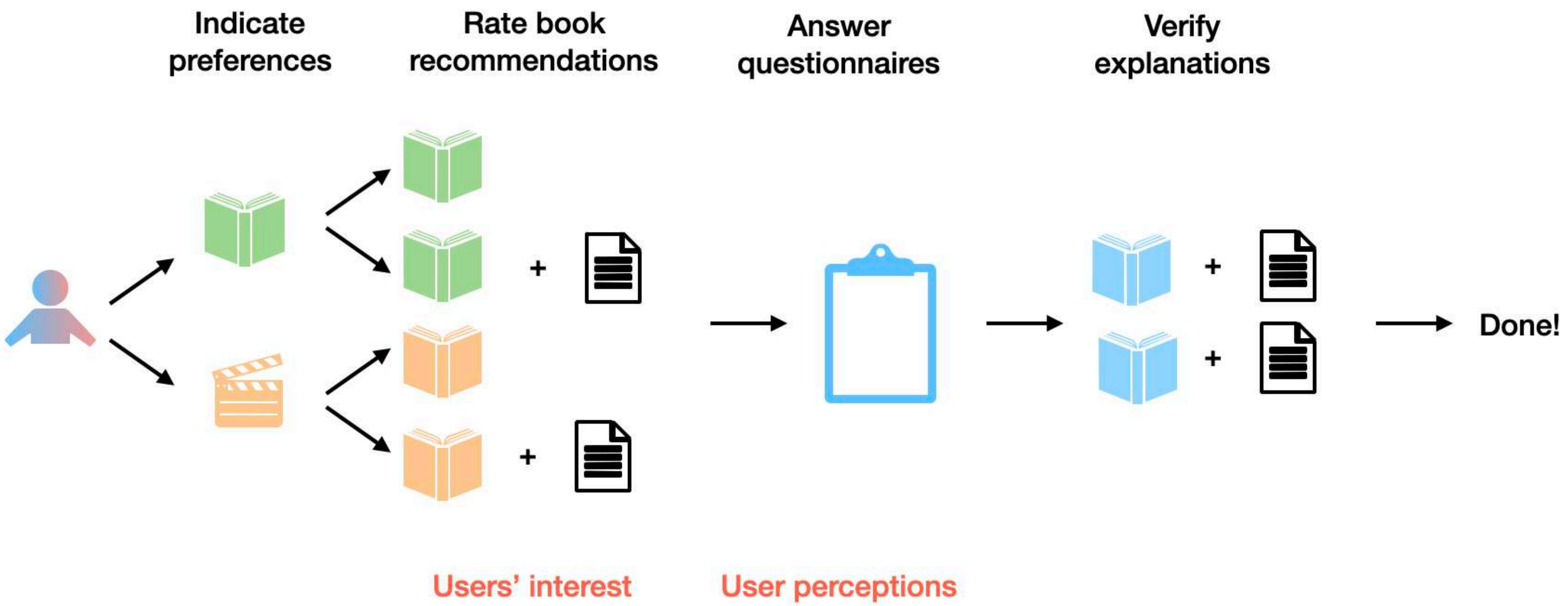
Measures

- **RQ:** How *cross-domain recommendations* and *explanations* affect user perceptions and behavioural intentions?



- Behavioural intentions
  - User's interest ← ratings
- User perceptions
  - **Seven** aspects
- 5-point Likert response scales

User Study



- 237 **valid** participants on Amazon Mechanical Turk
    - Between-subject design: 57-63 each scenario

Results - User Perceptions

Ind. Variables	Transparency	Scrutability	Trust	Efficiency	Persuasiveness	Effectiveness	Satisfaction
Familiar	1.155 ↑		1.156 ↑	0.896 ↓	1.219 ↑		
CDR			0.505 ↓				
Exp.	2.257 ↑	2.604 ↑					
CDR·Exp.							

Finding #1: CDR *decrease* perceived **trust** ↓

Finding #2: CDE *influence* user perceptions *the same as* SDE

Results - Behavioural Intentions

Familiar (know the plot)	Familiar (read the book)	CDR	Explanation	CDR•Exp.	
2.482 ↑	5.929 ↑	0.701 ↓ ×	0.783 ↓ ×	1.662 ↑	= 0.91 ↓

Finding #3: CDR *decrease* **interest** ↓

Finding #4: Explanations *decrease* **interest** ↓ in SDR

Finding #5: CDE *increase* **interest** ↑, but *lower than* SDR w/o Explanations

Summary

- The first study for user perceptions of CDR and Explanations

- Negative user perceptions

	Trust	Interest
CDR	↓	↓
CDR + Exp.		↑ (<SDR)

Offline and online evaluation may yield different results!

  - User experiments are important!

- Future work: different definitions of domains, different explanation styles

Temporal Consistency and Data Leakage in Offline Evaluation of Sequential Recommender Systems

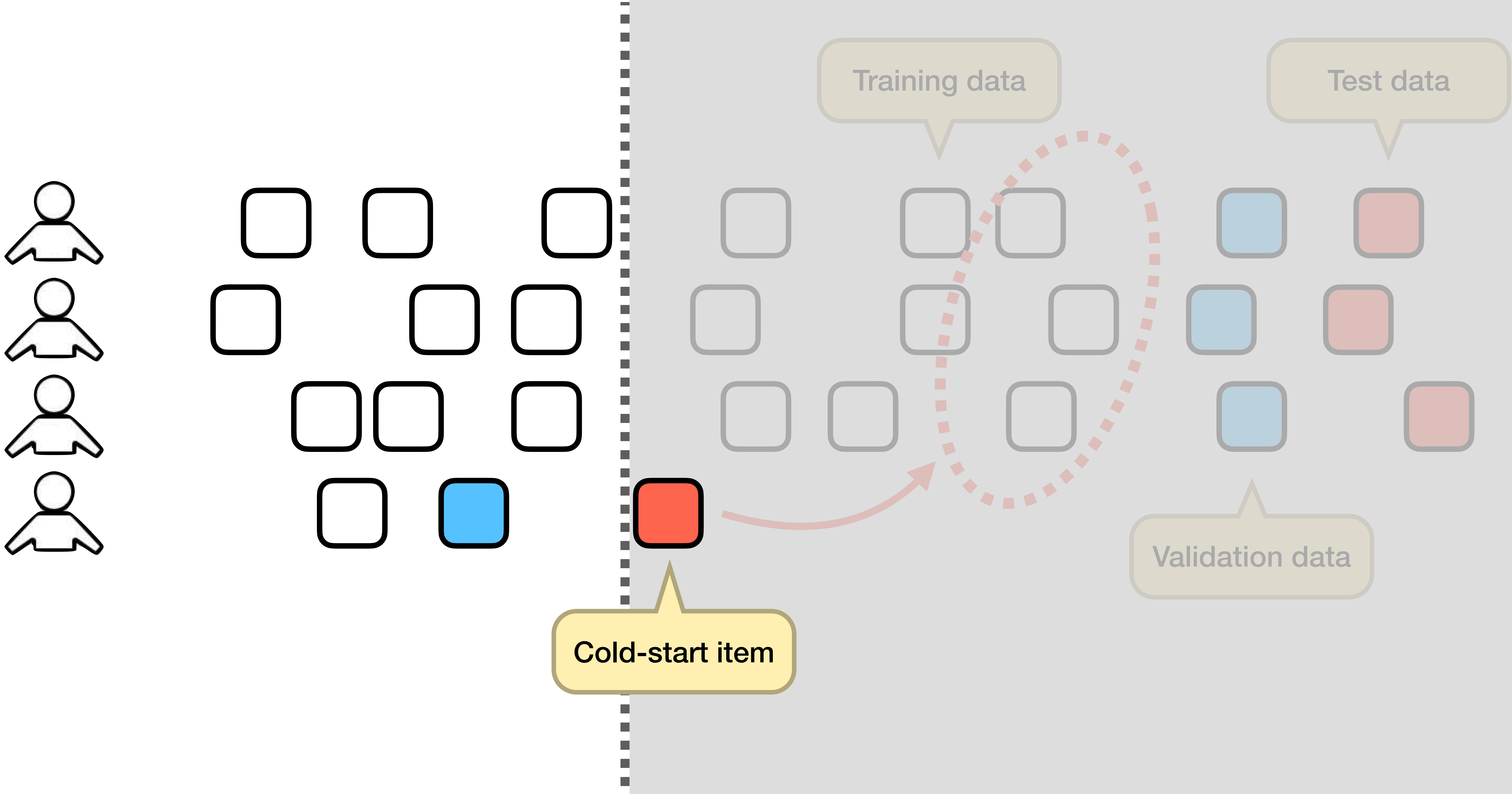
Huy Hong Le, Yang Liu, Dorota Głowacka and Alan Medlar
University of Helsinki

UNIVERSITY OF HELSINKI

Data leakage in Sequential Recommender Evaluation

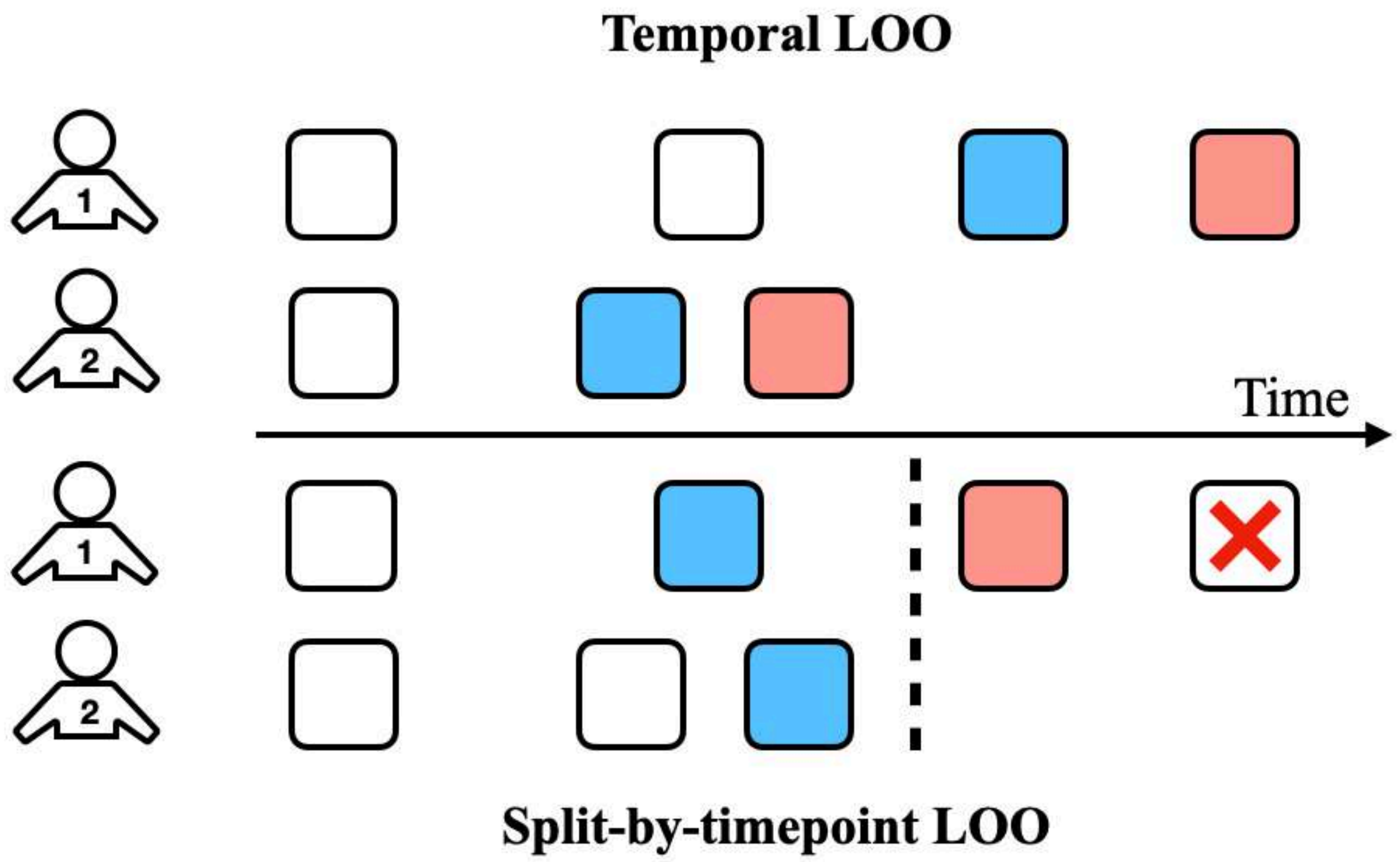
- Interactions are not i.i.d., they are a time-series
- Most data splitting strategies do not respect **temporal consistency** between training and test data
- Without temporal consistency there is **data leakage between users**
  - Recommending a movie that was just released using information from the future!

Temporal leave-one-out leaks information from the future



Temporal LOO vs Split-by-timepoint LOO

- Temporal LOO
  - For each user: last item in test set, second to last item in validation set
- Split-by-timepoint LOO
  - Find timestamp with the most active users, *t*
  - For each user: first interaction after *t* goes in test set, first interaction before *t* goes in validation set
  - Discard all other interactions after *t*



Results: Split-by-timepoint LOO vs Temporal LOO

- Split-by-timepoint LOO has much lower nDCG@10 than temporal LOO
- Median differences (unsampled nDCG):
  - ML-1m: -91.5%
  - Yelp: -54.2%
  - Steam: -19.5%
  - Beauty: -78.0%
- Similar results for recall@10…

	Model	Popularity-sampled			Unsampled		
		T-LOO (nDCG@10)	ST-LOO (nDCG@10)	Perf. diff (%)	T-LOO (nDCG@10)	ST-LOO (nDCG@10)	Perf. diff (%)
ML-1m	FPMC	0.3429	0.1118	-67.40%	0.1065	0.0158	-85.16%
	GRU4Rec	0.4748	0.1251	-73.65%	0.1624	0.0138	-91.50%
	Caser	0.3727	0.1078	-71.08%	0.1023	0.0081	-92.08%
	SASRec	**0.4921**	0.1250	-74.60%	**0.1814**	0.0174	-90.41%
	BERT4Rec	0.4654	0.0968	-79.20%	0.1613	0.0110	-93.18%
	S^3-Rec	0.4875	**0.1410**	-71.08%	0.1807	**0.0231**	-87.22%
	LightSANs	0.4592	0.1100	-76.05%	0.1457	0.0106	-92.72%
	SINE	0.2656	0.0782	-70.56%	0.0452	0.0064	-85.84%
	FEARec	0.4534	0.1032	-77.24%	0.1339	0.0106	-92.08%
Yelp	FPMC	0.3760	0.2395	-36.30%	0.0194	0.0082	-57.73%
	GRU4Rec	0.4278	0.3024	-29.31%	0.0232	0.0101	-56.47%
	Caser	0.3962	0.2688	-32.16%	0.0168	0.0091	-45.83%
	SASRec	0.4515	0.3066	-32.09%	**0.0374**	**0.0175**	-53.21%
	BERT4Rec	0.4081	0.2827	-30.73%	0.0207	0.0094	-54.59%
	S^3-Rec	-	-	-	-	-	-
	LightSANs	**0.4627**	0.3191	-31.04%	0.0355	0.0164	-53.80%
	SINE	0.4313	0.3215	-25.46%	0.0295	0.0155	-47.46%
	FEARec	0.4528	**0.3243**	-28.38%	0.0349	0.0155	-55.59%
	FPMC	0.0848	0.0745	-16.45%	0.0547	**0.0556**	+1.65%
	GRU4Rec	0.0988	0.0791	-19.94%	0.0622	0.0501	-19.45%
	Caser	0.0923	0.0761	-17.55%	0.0640	0.0547	-14.53%
	SASRec	0.1017	0.0789	-22.42%	0.0669	0.0533	-20.33%

Results: Data Leakage in Temporal LOO

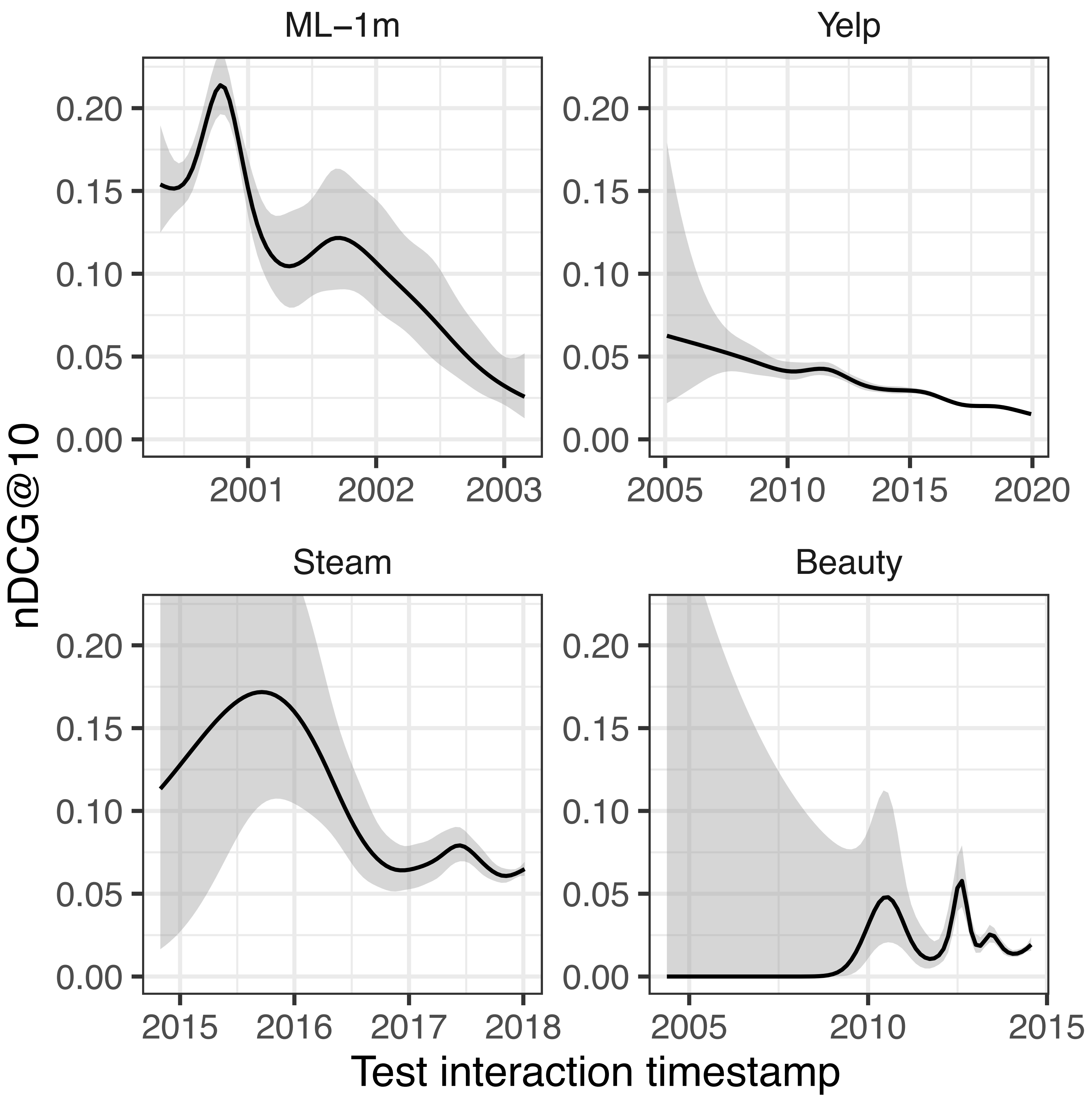
- Lower nDCG in split-by-timepoint LOO could be due to **data leakage** or **model quality** (1.2–2.3x more training data in temporal LOO)
- Evidence data leakage > training set size:
    - Performance of test items in T-LOO drop over time
    - Validation performance drop in ST-LOO much lower

      ML-1m: -91.5% → -5.8% (median diff. nDCG@10)

      Yelp: -54.2% → -2.9%

      Steam: -19.5% → +3.9%

      Beauty: -78.0% → -10.9%



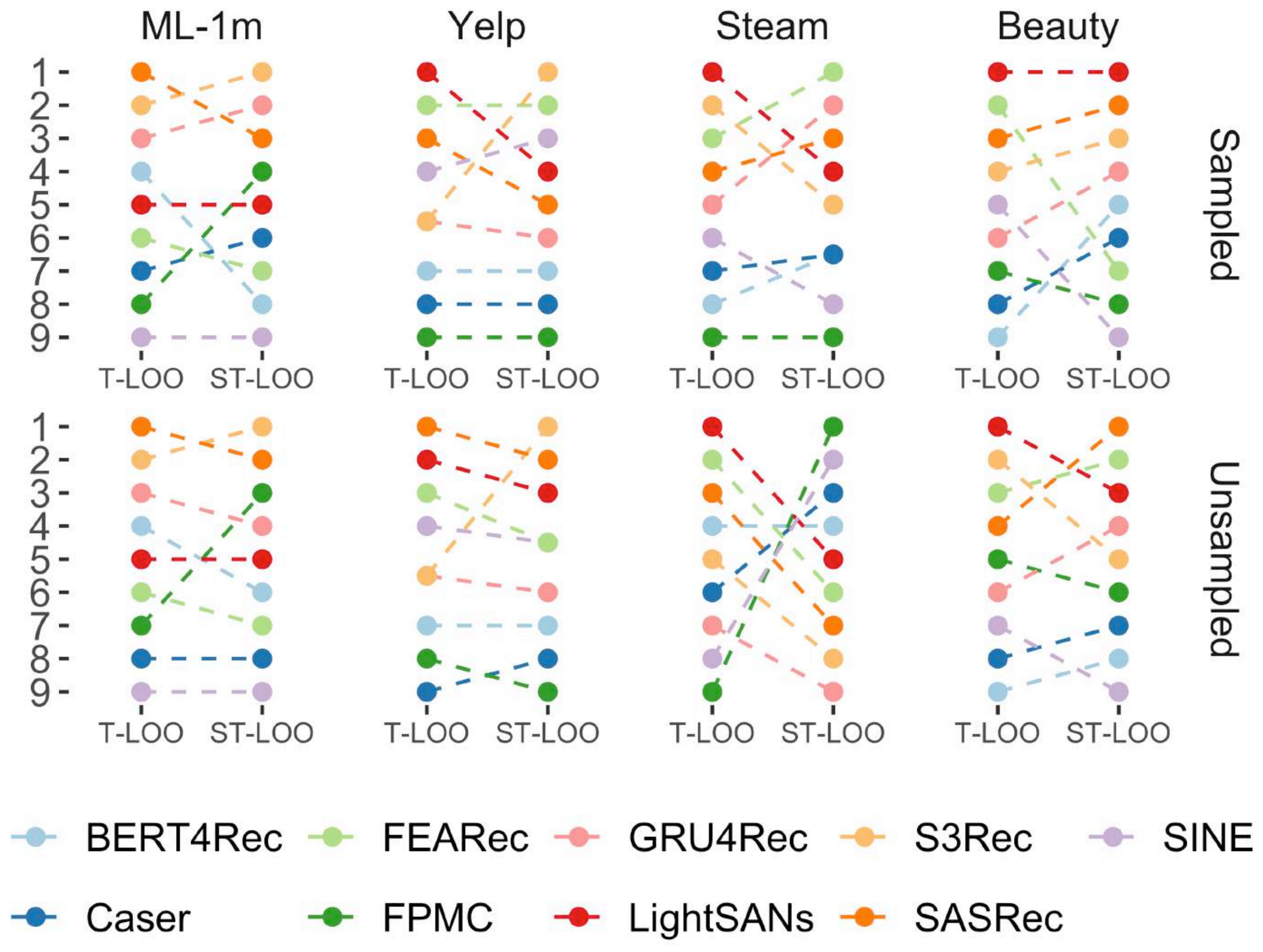
Results: Comparison with general recommenders using ST-LOO

- General recommenders outperform the best performing sequential recommenders in **ML-1m**, **Steam** and **Beauty** (unsampled nDCG)

Model	ML-1m		Yelp		Steam		Beauty	
	Pop. sampled (nDCG@10)	Unsampled (nDCG@10)	Pop. sampled (nDCG@10)	Unsampled (nDCG@10)	Pop. sampled (nDCG@10)	Unsampled (nDCG@10)	Pop. sampled (nDCG@10)	Unsampled (nDCG@10)
Pop.	0.0892	0.0249	0.0229	0.0017	0.0614	0.0383	0.0164	0.0052
ItemKNN	0.1006	0.0263	0.1971	0.0106	0.0690	0.0430	0.0895	0.0059
BPR	0.1249	0.0290	0.2249	0.0077	0.0738	0.0564	0.0611	0.0079
SLIM	0.1072	**0.0305**	-	-	**0.0885**	0.0285	0.0521	0.0007
NeuMF	0.1222	0.0219	0.2216	0.0085	0.0724	**0.0581**	0.0573	0.0057
NGCF	0.1302	0.0108	0.2472	0.0018	0.0676	0.0405	0.0891	0.0036
LightGCN	0.1345	0.0301	0.2470	0.0138	0.0714	0.0579	0.0970	**0.0129**
NCL	0.1282	0.0250	0.2574	0.0114	0.0788	0.0579	0.0684	0.0115
FPMC	0.1118	0.0158	0.2395	0.0082	0.0745	0.0556	0.0598	0.0063
S^3-Rec	**0.1410**	0.0231	-	-	0.0777	0.0510	0.0936	0.0079
SASRec	0.1250	0.0174	0.3066	**0.0175**	0.0789	0.0533	**0.1002**	0.0102
FEARec	0.1032	0.0106	**0.3243**	0.0155	0.0793	0.0535	0.0628	0.0093

Summary

- Temporal leave-one-out (1) **exaggerates the performance of sequential recommenders due to data leakage**, which (2) changes the model ranking
- Split-by-timepoint leave-one-out does not suffer from data leakage, but performance is slightly lower due to smaller training set size
- General recommenders can outperform sequential recommenders in 3/4 data sets



Thank you!

https://glowacka.org
dorota.glowacka@helsinki.fi

User-centric Design and Evaluation of Exploratory Search and Recommender Systems **and more!**

Behind the Scenes

Adapting Cinematography and Editing Concepts to Navigation in Virtual Reality

Alan Medlar, Mari Lehtikari, Dorota Głowacka

alan.j.medlar@helsinki.fi
dorota.glowacka@helsinki.fi


Teleportation
Pro: No VR sickness
Con: Reduced spatial awareness


METROPOLIS
MANUSKRIPT:
THEA von HARBOU
MUSIK:
GOTTFRIED HUPPERTZ
EIN FILM VON FRITZ LANG
JN DEN HAUPTROLLEN:
BRIGITTE HELM · GUSTAV FRÖHLICH,
ALFRED ABEL, RUDOLF KLEIN-ROGGE, THEODOR LOOS, FRITZ RASP, HEINRICH GEORGE
AN DER KAMERA: KARL FREUND, GÜNTHER RITTAU
UFA
UFA FILM IM VERLEIH DER
PARUFAMET

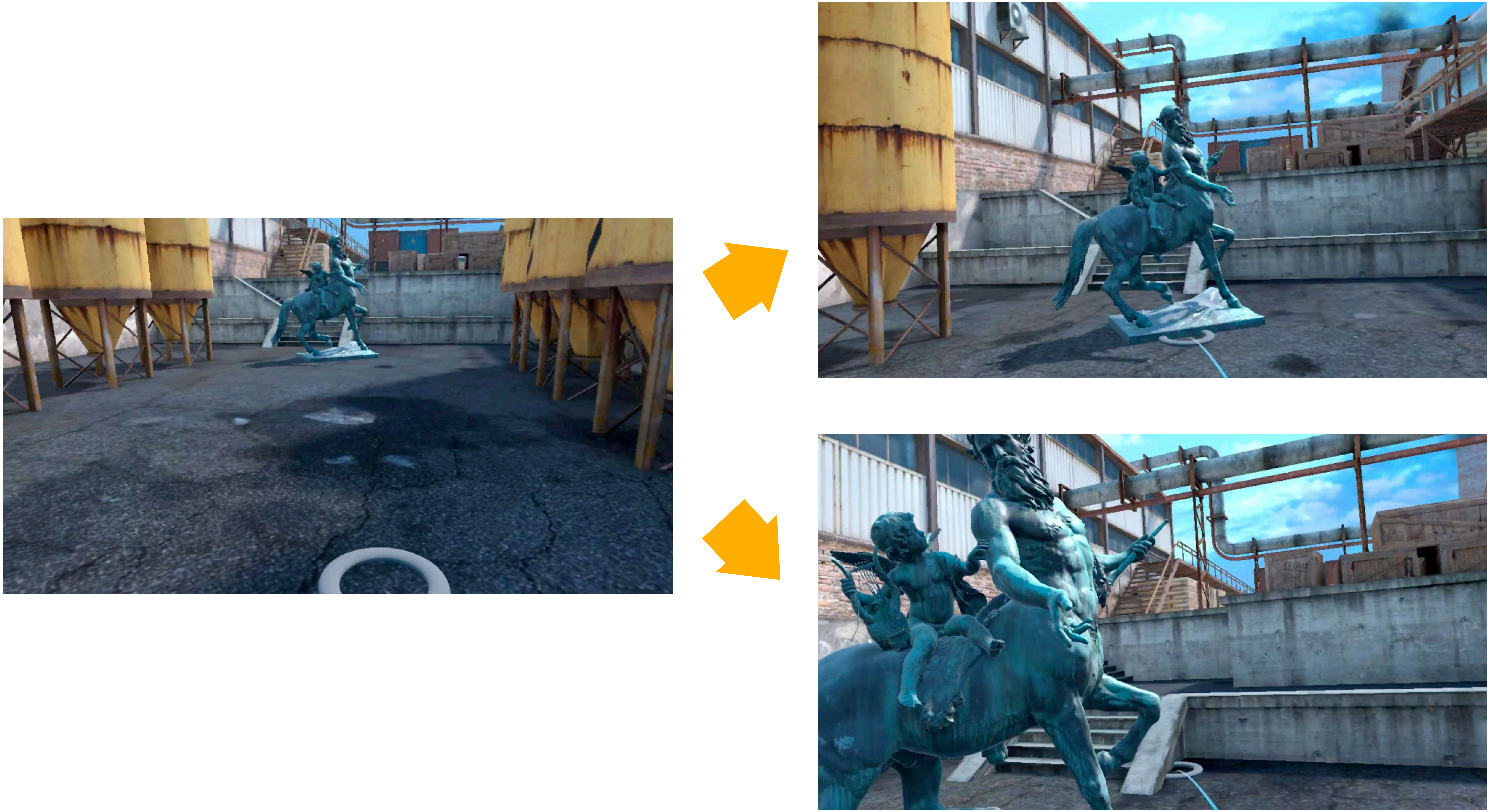
Cut

Cut Cut Cut Cut

ACTIVE

- We reconceptualize teleportation as a cut and apply the rules of continuity editing
- Procedure:
  - Select target position + teleport
  - Reposition camera
  - Reorient camera

Rule of Thirds
Establishing Shot
Cutting Closer
180 Degree Rule
Graphic Vectors



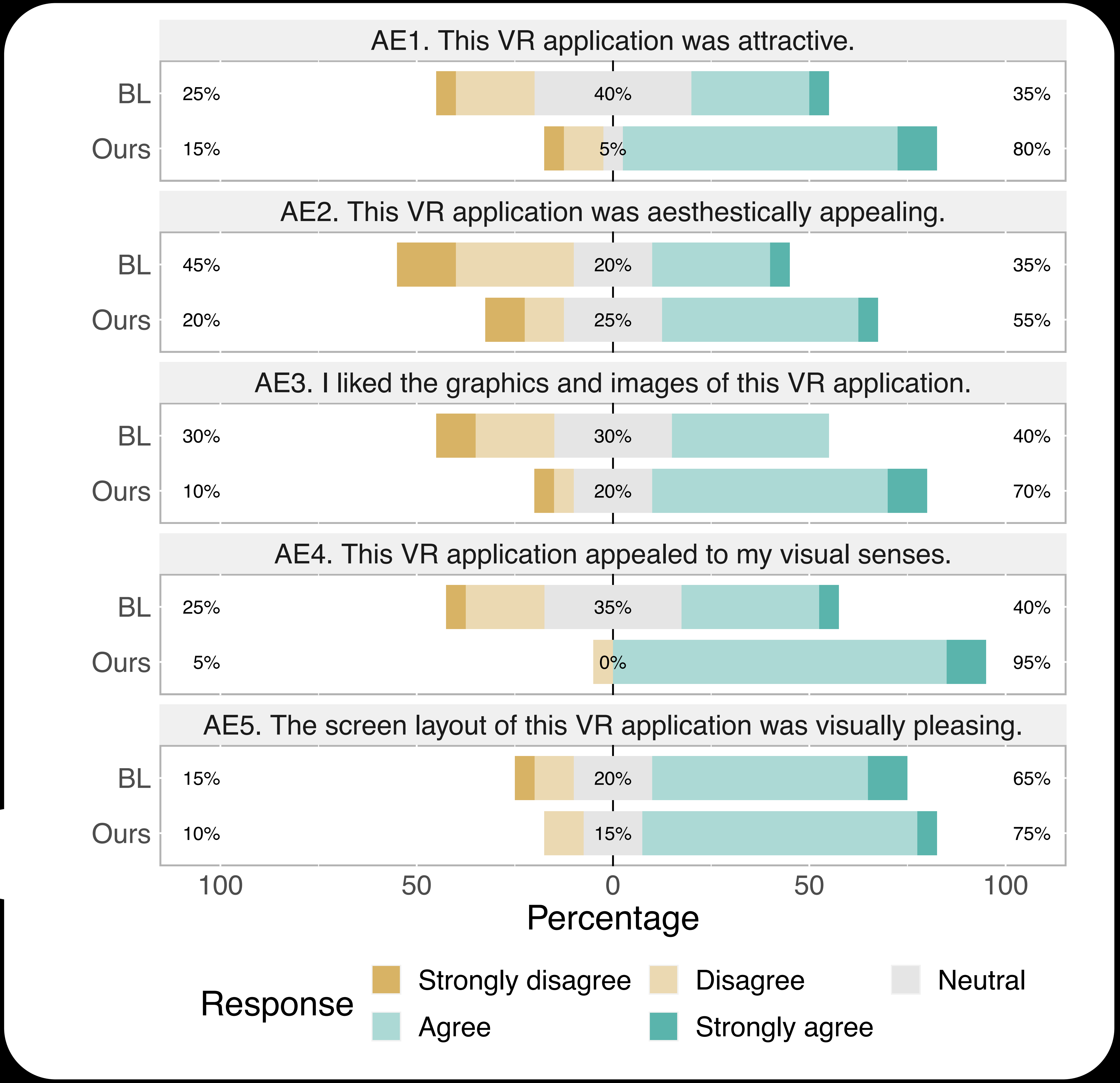
How does ACTIVE affect…

- …user **engagement** in virtual environments?
- …**recall** of the contents of the virtual environment?
- …symptoms of **VR sickness**?
- …perception of **involvement/control** in VR?


Setup: Industrial setting, 20 statues, Meta Quest 2
Task: Explore a virtual environment using teleporting or ACTIVE
Participants: 40, between-subject design
Measures: UES, cued recall, SSQ, PQ

Results

- VR sickness: no difference
- Presence: no difference
- Cued recall: no difference
- User engagement: +8.6%
  - Aesthetic appeal: +17.6% (10-55% points)



Concepts from continuity editing make VR more engaging

Increased engagement does not improve recall

No loss of presence, no VR sickness

Thank you!

https://glowacka.org
dorota.glowacka@helsinki.fi