Transparency, Privacy, and Fairness in Recommender Systems

Habil Talk
@TU Graz, CSBME, 14.01.2025

Priv.-Doz. Dipl.Ing. Dr.techn. Dominik Kowald, BSc.

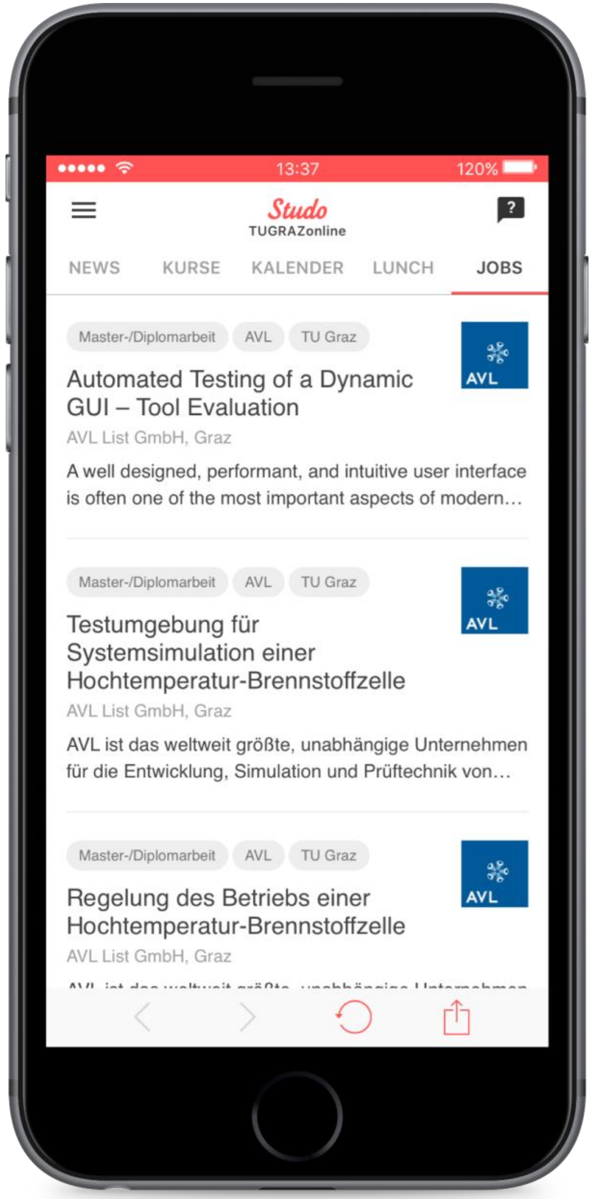
SCIENCE
PASSION
TECHNOLOGY

About Me

- **2021–now:** research area head **Fair-AI** @ Know-Center
  - **Habilitation** @ ISDS-TUG for scientific subject: *Applied Computer Science*
  - **Teaching:** *Databases, Data Management, Scientific Writing, theses and seminars*
- **2018–2020:** deputy research area head Social Computing @ Know-Center
  - **Projects:** *Horizon 2020 (e.g., AI4EU), FFG COMET modules (e.g., DDAI)*
- **2012–2017:** PhD student @ ISDS-TUG & researcher @ Know-Center
  - **Thesis:** *Modeling activation processes in human memory to improve tag recommendation*
  - **Supervisors:** *Prof. Stefanie Lindstaedt (ISDS-TUG), Assoc.Prof. Elisabeth Lex (ISDS-TUG), Prof. Tobias Ley (Tallinn University)*

Recommender Systems

- Recommender systems (RecSys) → **integral part of online experience**
  - **Analyze** past usage behavior to build **user models** and **suggest** new content



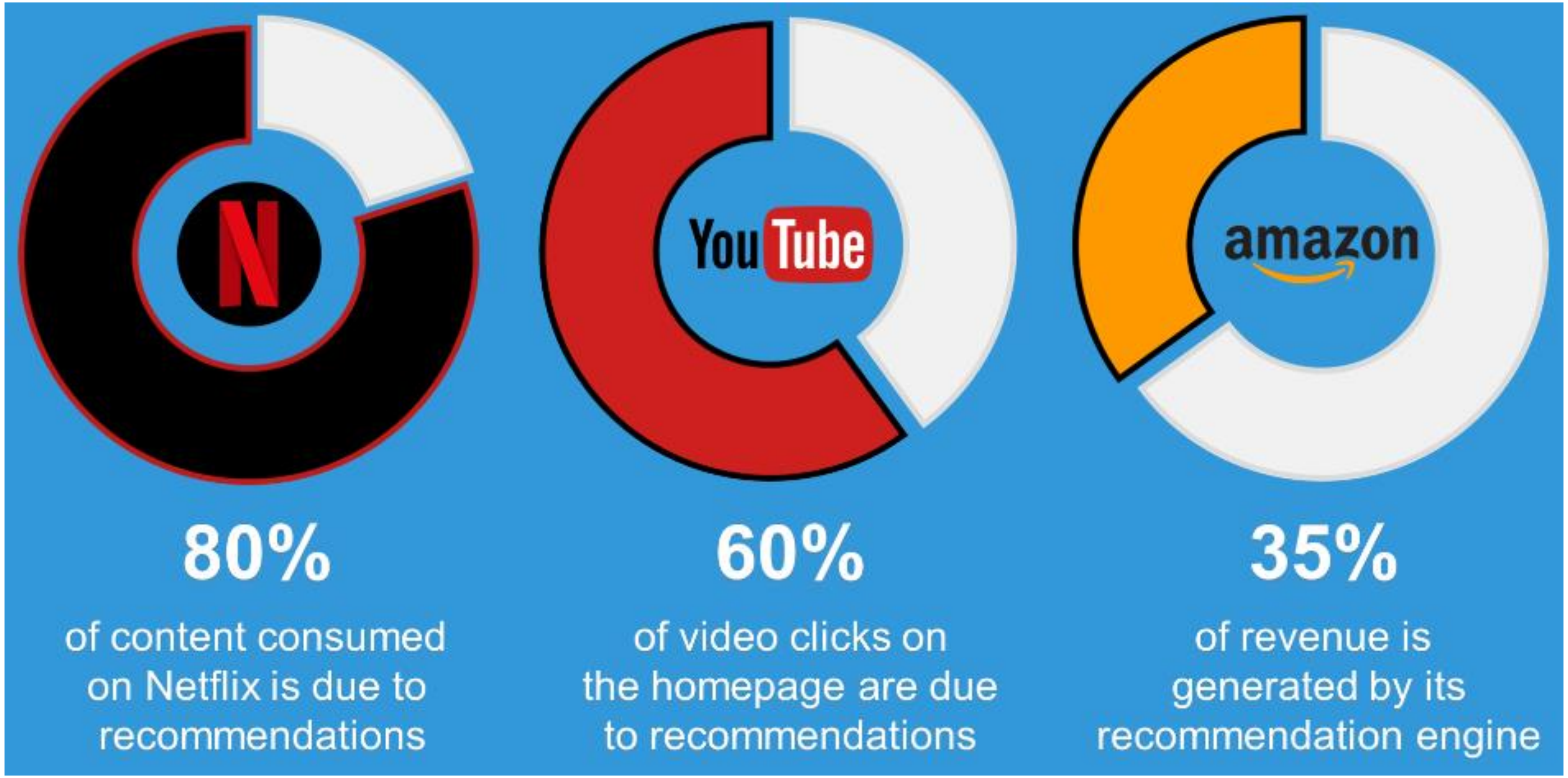
Recommender Systems

(Applied Computer Science aspects)

- **Network science**
  - Graph of users & items → collaborative filtering, graph neural networks
- **Natural language processing**
  - Similarities between items → content-based filtering, conversational systems
- **Machine learning**
  - Rating prediction & learning2rank → trustworthy AI and machine learning
- **User modeling**
  - Embeddings of users / items → alternative: cognitive-inspired modeling
- **Human-computer interaction**
  - Recommender SYSTEMS → user-centric design and evaluation
- **Data management**
  - Large user bases and item catalogs → efficient data storage for (near) real-time recommendations
- **Information retrieval**
  - Similarities to search systems → evaluation metrics (precision, recall, nDCG, diversity, etc.)

Recommender Systems

(Business value, taken from https://scar.know-center.tugraz.at/)

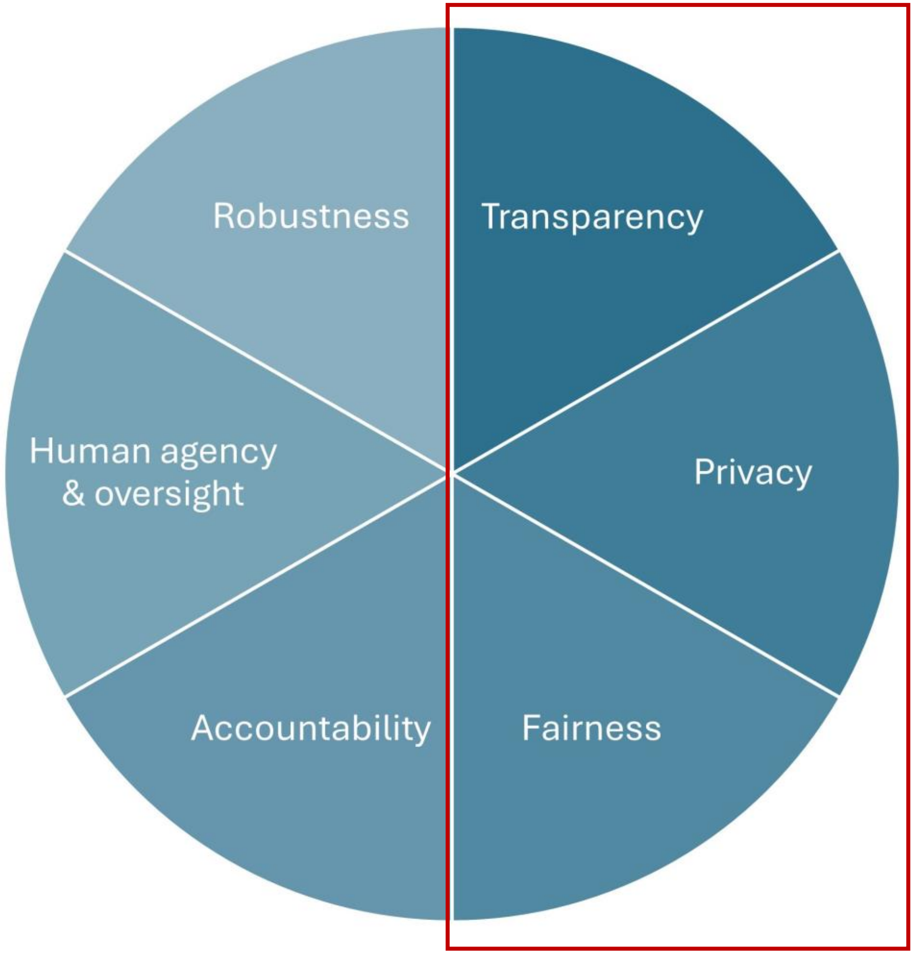


"… the combined effect of personalization and recommendations save us more than $1B per year." (Neil Hunt, Chief Product Officer, Netflix)

Motivation

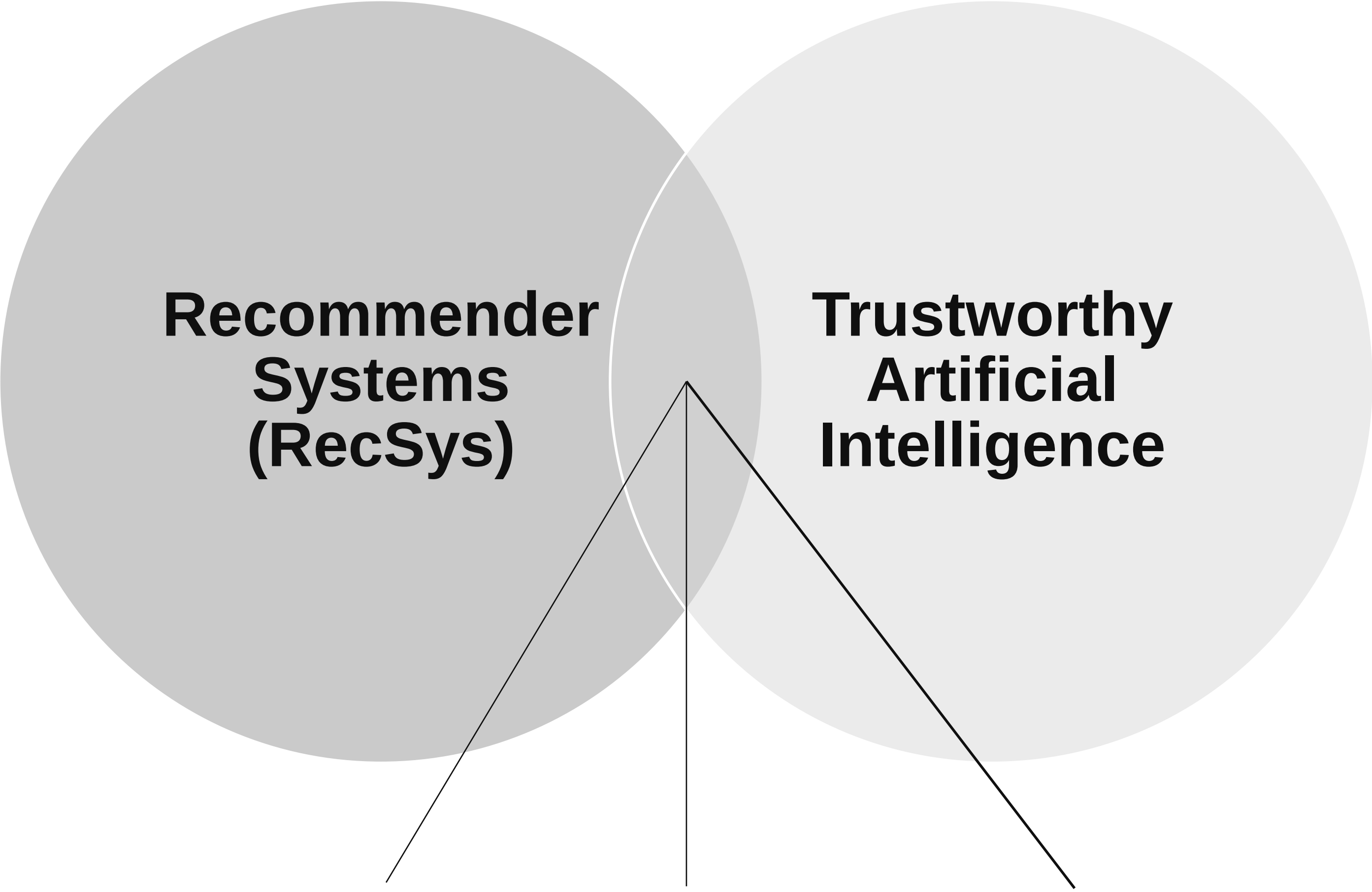
- Among most widely used **applications of AI** and **machine learning**
  - **User-centric nature** → humans directly interact with / are affected by RecSys

- **Regulations** and **requirements of Trustworthy AI** relevant for design of RecSys (e.g., AI Act)
  - **Transparency**
    - Explainable design and decisions of algorithms
  - **Privacy**
    - Responsible usage and protection of users' data
  - **Fairness**
    - Detect and prevent potential discrimination of users



Kowald, D. et al. (2024). Establishing and Evaluating Trustworthy AI: Overview and Research Challenges. *Frontiers in Big Data and AI.*

Research Fields



Transparency
& cognitive models
in RecSys

Privacy
& limited preference
information in RecSys

Fairness
& popularity bias
in RecSys

Regular article | Open Access | Published: 30 March 2021

Support the underground: characteristics of beyond-mainstream music listeners

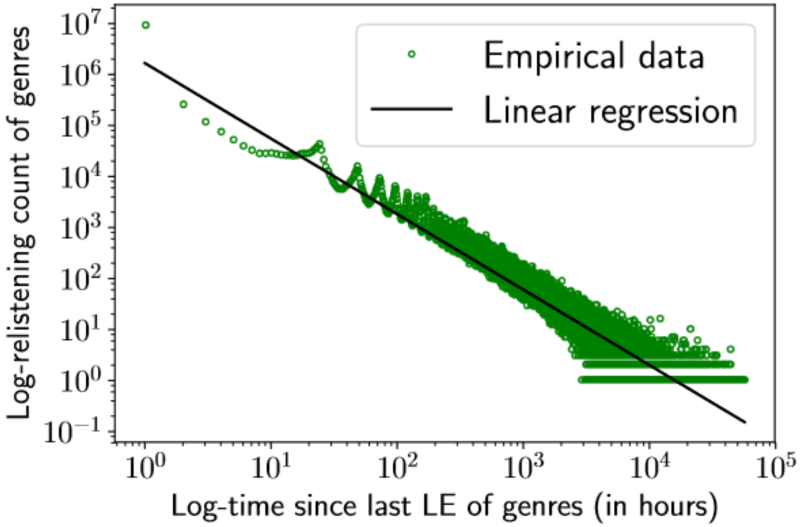
Award Winners 2022

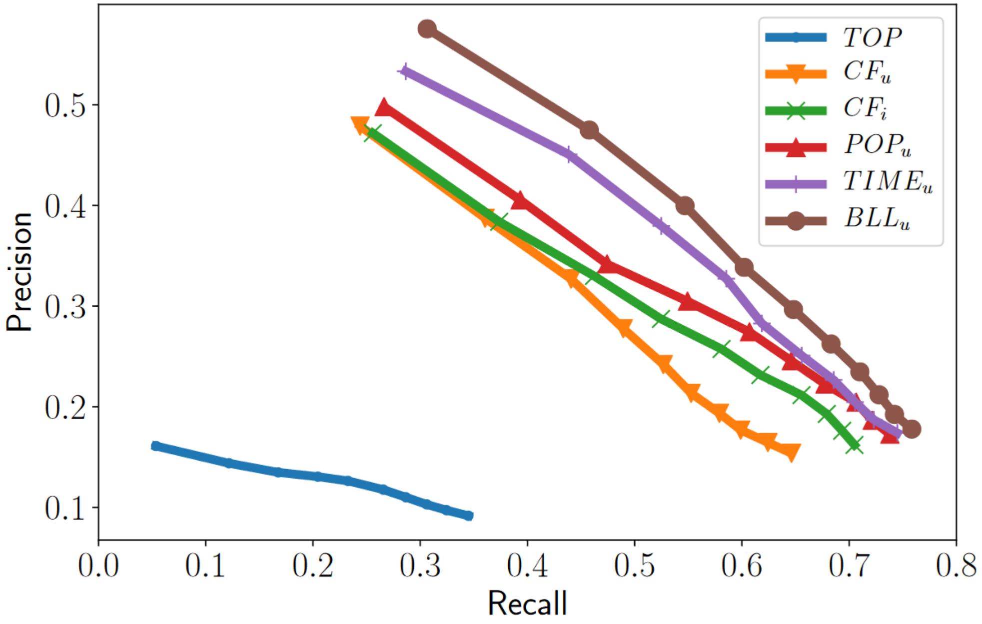
Publication Popularity Bias in Recommender Systems
by Dominik Kowald and Emanuel Lacić, Institute of Interactive Systems and Data Science and Know Center

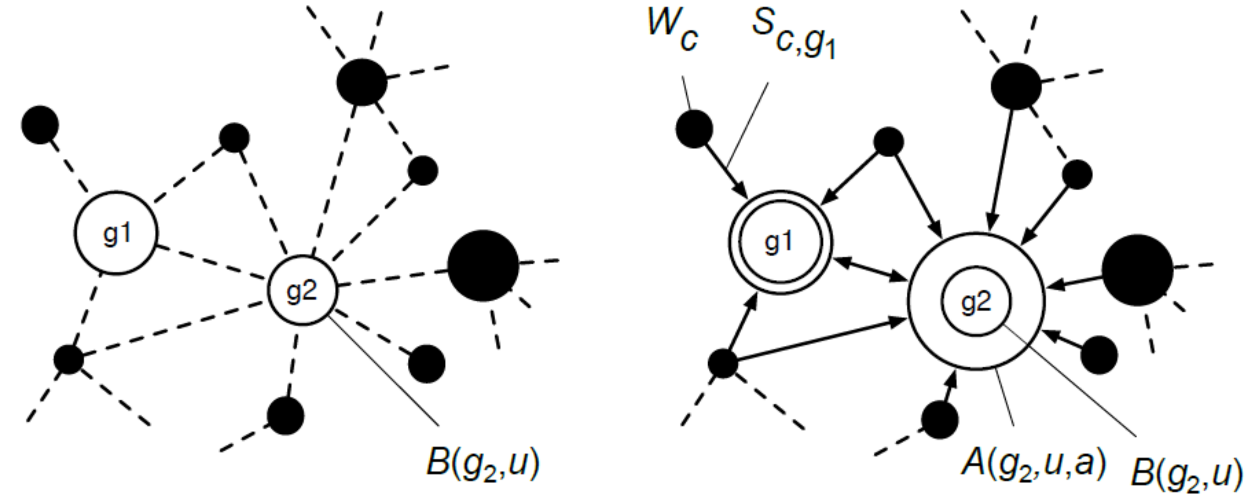
Transparency & Cognitive Models in RecSys

(ACT-R-based RecSys)

$$B(g,u) = ln\left(\sum_{j=1}^{n} t_{u,g,j}^{-d}\right)$$





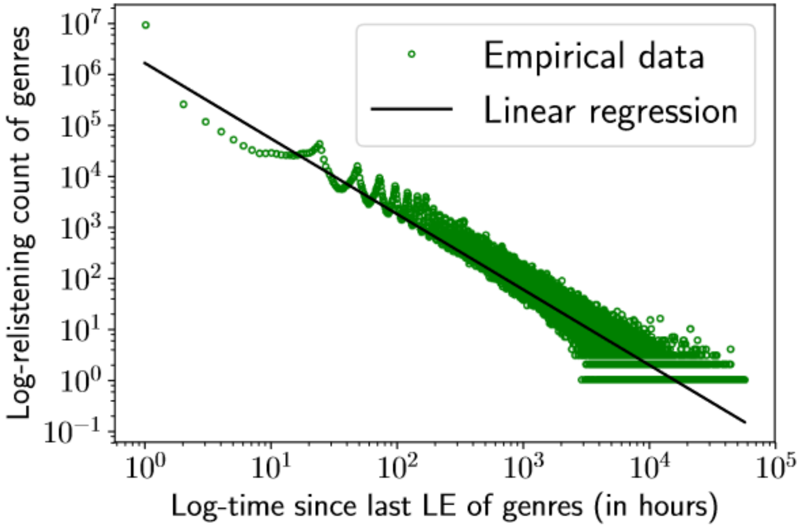


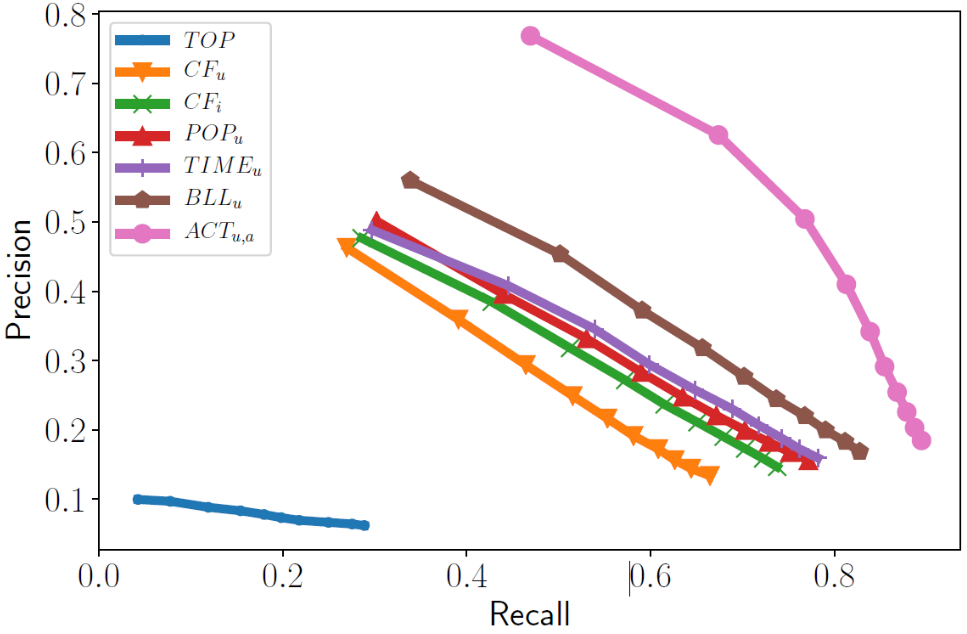
1. Kowald, D., Seitlinger, P., Trattner, C., & Ley, T. (2014). Long Time no See: The Probability of Reusing Tags as a Function of Frequency and Recency. In *Proceedings of the 23rd International World Wide Web Conference (WWW'2014)*. ACM.
2. Kowald, D., Pujari, S., & Lex, E. (2017). Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach. In *Proceedings of the 26th International World Wide Web Conference (WWW'2017)*. ACM.
3. Kowald, D., Reiter-Haas, M., Kopeinik, S., Schedl, M., & Lex, E. (2024). Music Preference Modeling and Recommendation with a Model of Human Memory Theory. *A Human-centered Perspective of Intelligent Personalized Environments and Systems*. Springer.

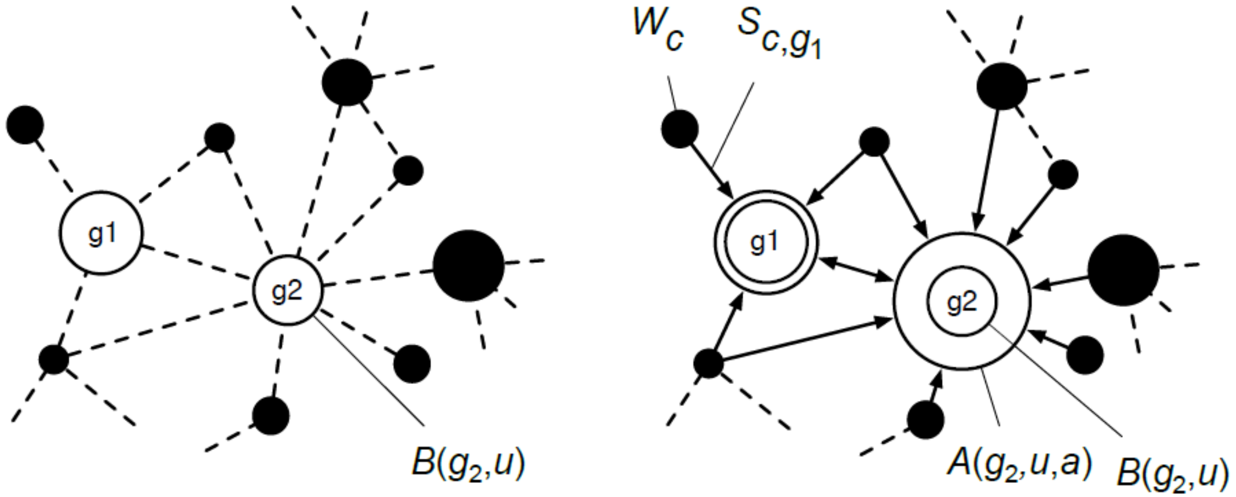
Transparency & Cognitive Models in RecSys

(ACT-R-based RecSys)

$$B(g,u) = ln\left(\sum_{j=1}^{n} t_{u,g,j}^{-d}\right)$$





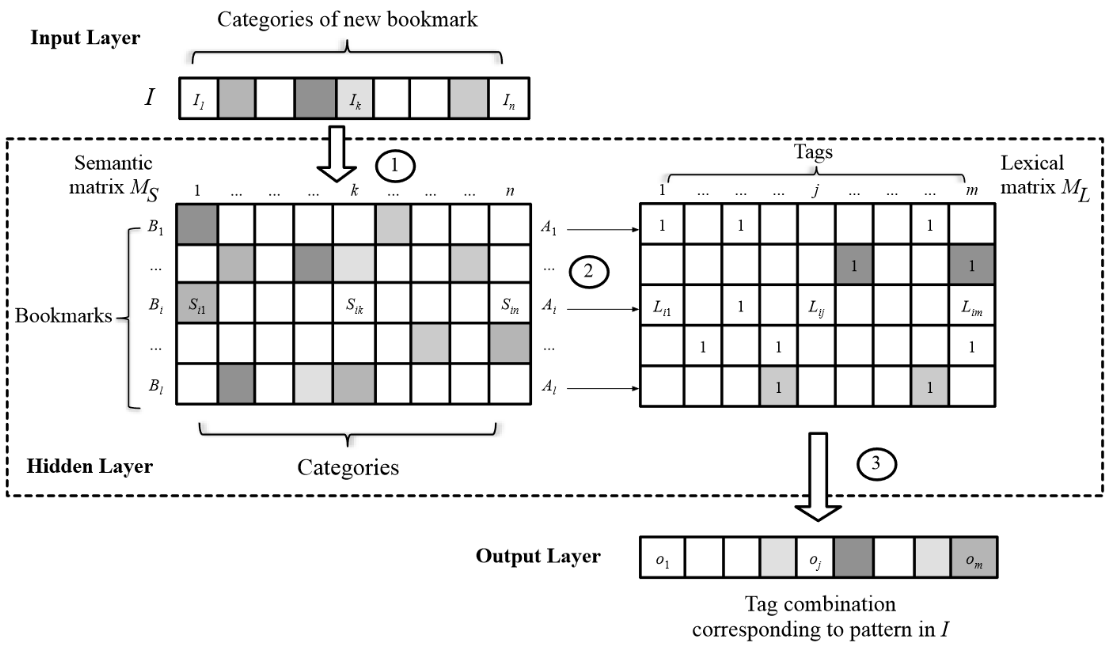


1. Kowald, D., Seitlinger, P., Trattner, C., & Ley, T. (2014). Long Time no See: The Probability of Reusing Tags as a Function of Frequency and Recency. In *Proceedings of the 23rd International World Wide Web Conference (WWW'2014)*. ACM.
2. Kowald, D., Pujari, S., & Lex, E. (2017). Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach. In *Proceedings of the 26th International World Wide Web Conference (WWW'2017)*. ACM.
3. Kowald, D., Reiter-Haas, M., Kopeinik, S., Schedl, M., & Lex, E. (2024). Music Preference Modeling and Recommendation with a Model of Human Memory Theory. *A Human-centered Perspective of Intelligent Personalized Environments and Systems*. Springer.

Transparency & Cognitive Models

(Psychology-informed RecSys)

- **Survey of publications on Psychology & RecSys:**
  - **Cognitive-inspired RecSys**
  - **Personality-aware RecSys**
  - **Affect-aware RecSys**
  - **Decision making & RecSys**
  - **User-centric evaluation of RecSys**

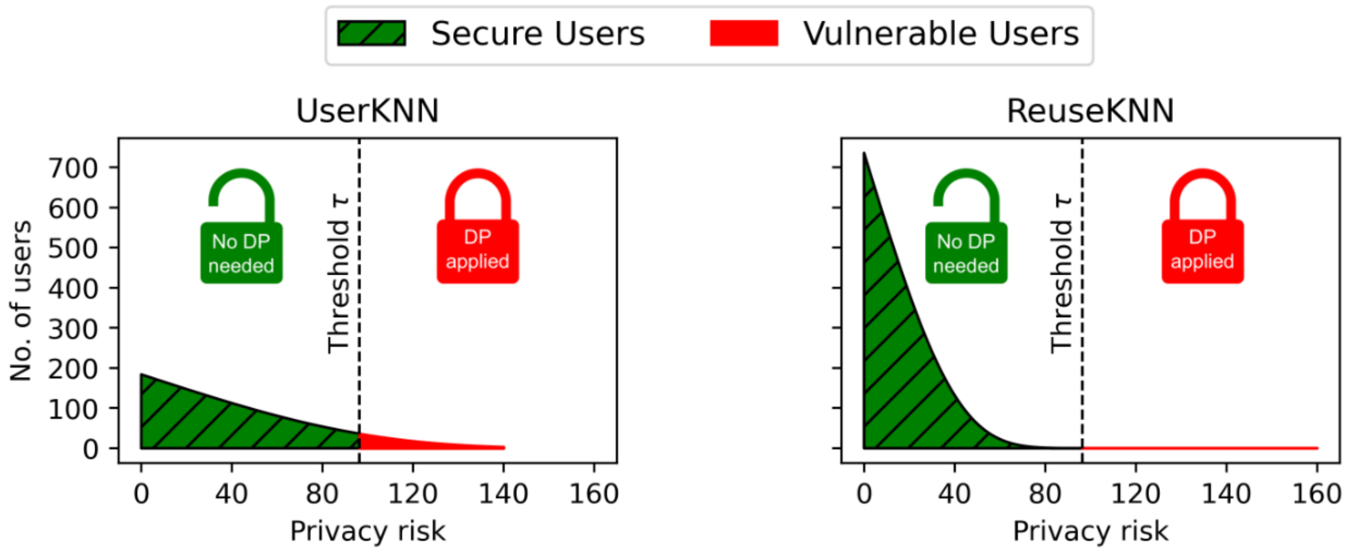


Recommended Track	Current obsession (BLL)	Current vibes (S)	Evergreens (V)	From similar listeners (SC)
From the Past Comes the Storms	0.471	0.248	0.281	0.000
Escape to the Void	0.306	0.353	0.341	0.000
To the Wall	0.294	0.359	0.347	0.000
R.I.P. (Rest in Pain)	0.264	0.374	0.362	0.000
The Abyss	0.263	0.375	0.362	0.000
Troops of Doom	0.000	0.000	0.000	1.000

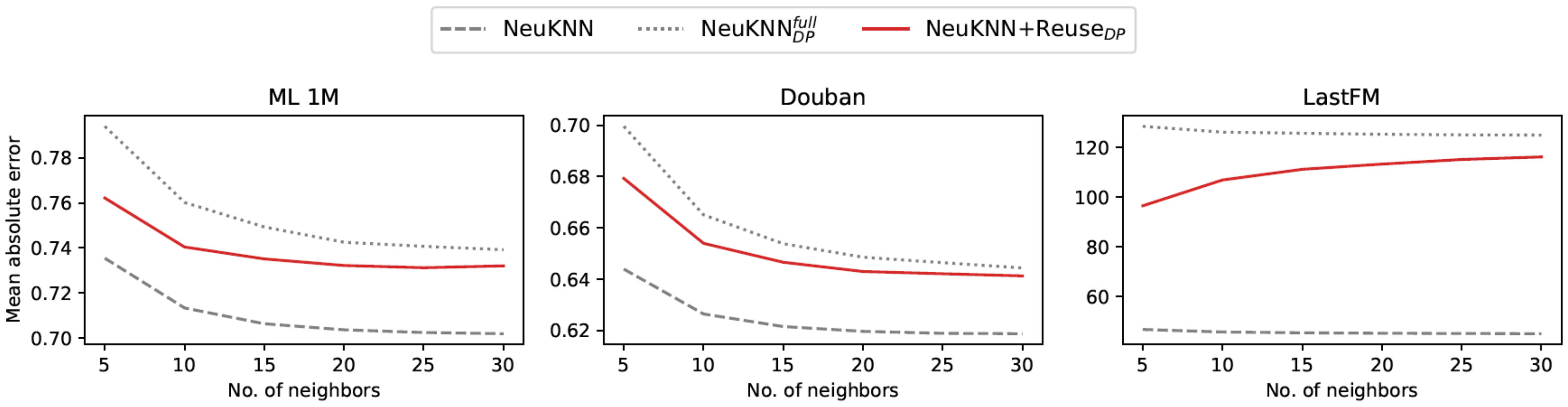
1. Seitlinger, P., Kowald, D., Trattner, C., & Ley, T. (2013). Recommending tags with a model of human categorization. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management (CIKM'2013)*. ACM.
2. Lex, E., Kowald, D., Seitlinger, P., Tran, T., Felfernig, A., & Schedl, M. (2021). Psychology-informed Recommender Systems. *Foundations and Trends in Information Retrieval*, Vol. 15, No. 2.
3. Moscati, M., Wallmann, C., Reiter-Haas, M., Kowald, D., Lex, E., & Schedl, M. (2023). Integrating the ACT-R Framework with Collaborative Filtering for Explainable Sequential Music Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys'2023)*. ACM.

Privacy & Limited Preference Information in RecSys

(Differentially-Private RecSys)



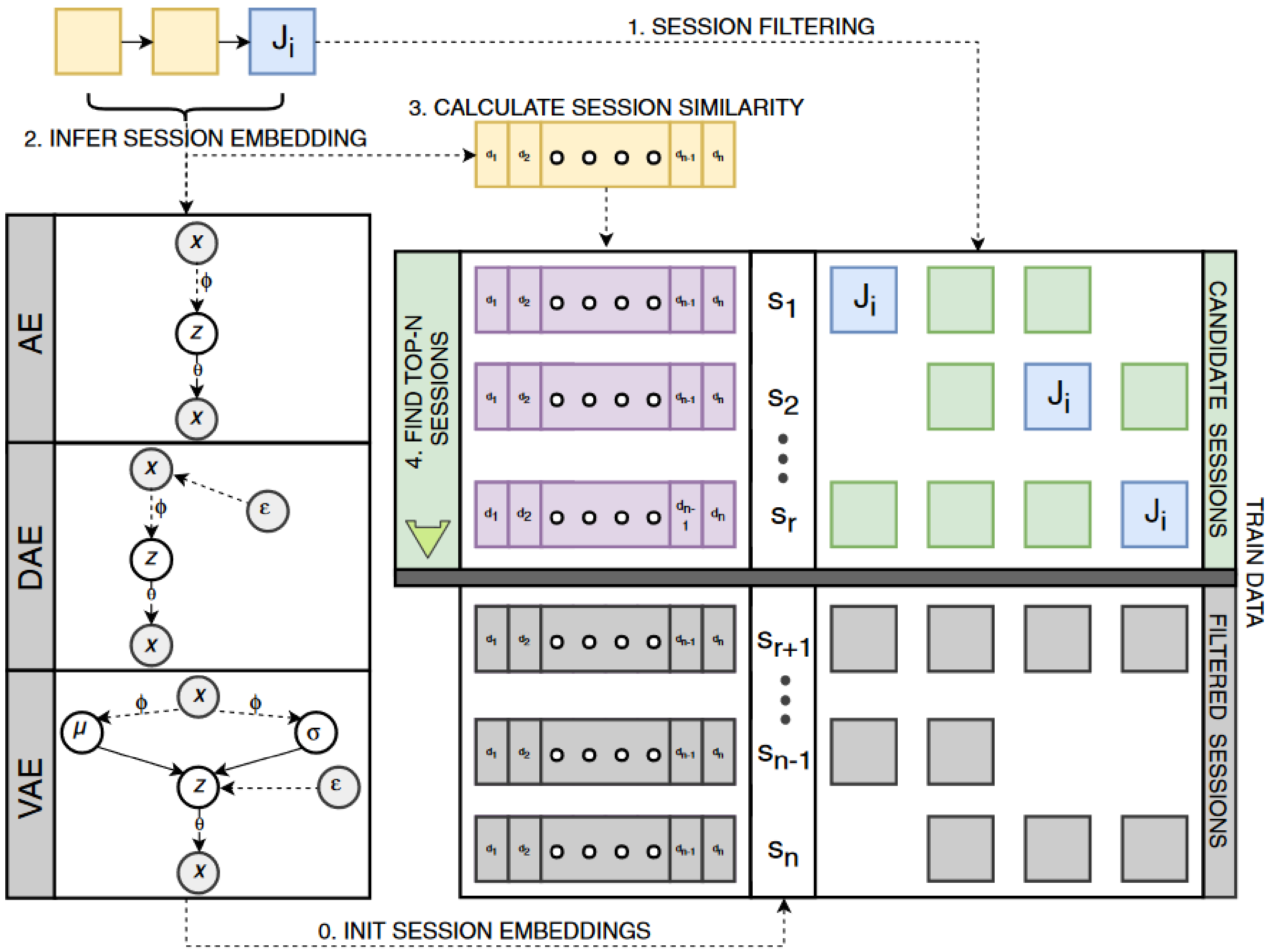
Method	ML 1M	Douban	LastFM	Ciao	Goodreads
UserKNN	80.39%	96.68%	99.89%	8.02%	65.00%
UserKNN+Reuse	84.64%	87.37%	98.90%	7.91%	52.29%
Expect	**24.13%**	**34.40%**	**68.20%**	**7.88%**	**29.12%**
Gain	25.09%	37.43%	80.28%	8.19%	40.51%



1. Muellner, P., Kowald, D., & Lex, E. (2021). Robustness of Meta Matrix Factorization Against Strict Privacy Constraints. In *Proceedings of the 43rd European Conference on Information Retrieval (ECIR'2021)*. Springer.
2. Muellner, P., Lex, E., & Kowald, D. (2021). Position Paper on Simulating Privacy Dynamics in Recommender Systems. In *Simulation for Recommender Systems Workshop (SimuRec'2021) co-located with ACM Conference on Recommender Systems (RecSys'2021)*. ACM.
3. Muellner, P., Lex, E., Schedl, M., & Kowald, D. (2023). Differential Privacy in Collaborative Filtering Recommender Systems: A Review. *Frontiers in Big Data - Reviews in Recommender Systems*.
4. Muellner P., Schedl, M., Lex, E., & Kowald, D. (2023). ReuseKNN: Neighborhood Reuse for Differentially-Private KNN-Based Recommendations. *Transactions on Intelligent Systems and Technology (TIST)*. ACM.

Privacy & Limited Preference Information in RecSys

(Session-based RecSys)

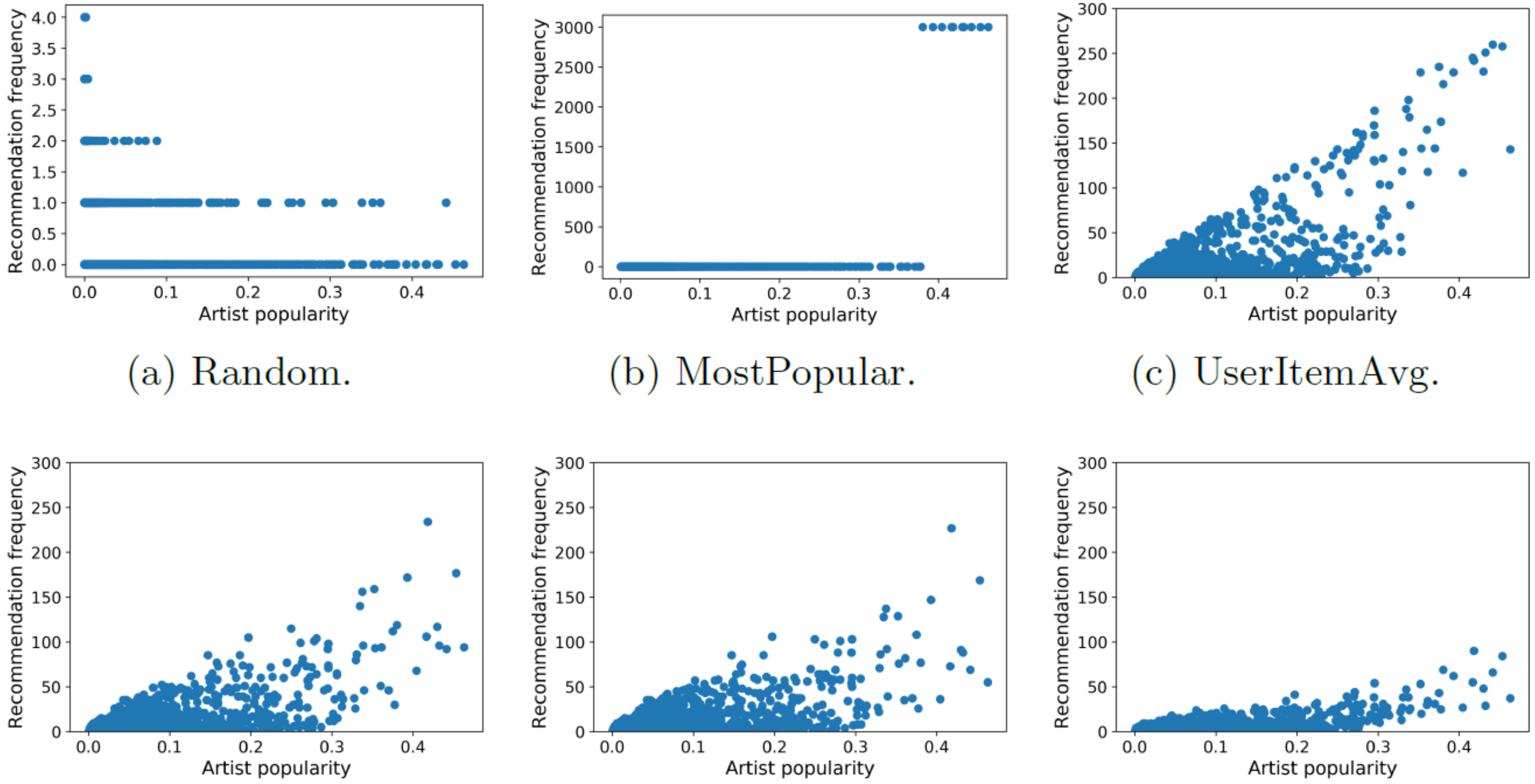


	Accuracy	Beyond Accuracy	Coverage
VAE_{Int}	++	++	++
VAE_{Comb}	+	++	++
sKNN	+	o	+
V-sKNN	++	+	++
S-sKNN	++	+	+
GRU4Rec	++	+	+
pRNN	- -	- -	- -
Bayes	- -	- -	o
iKNN	o	-	+
BPR-MF	-	- -	++
POP	- -	- -	- -

1. Duricic, T., Lacic, E., Kowald, D., & Lex, E. (2018). Trust-Based Collaborative Filtering: Tackling the Cold Start Problem Using Regular Equivalence. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys'2018)*. ACM.
2. Lacic, E., Reiter-Haas, M., Kowald, D., Dareddy, M., Cho, J., & Lex, E. (2020). Using Autoencoders for Session-based Job Recommendations. *User Modeling and User-Adapted Interaction (UMUAI)*. Springer.
3. Duricic, T., Kowald, D., Lacic, E., & Lex, E. (2023). Beyond-Accuracy: A review on Diversity, Serendipity, and Fairness in Recommender Systems based on Graph Neural Networks. *Frontiers in Big Data - Reviews in Recommender Systems*.

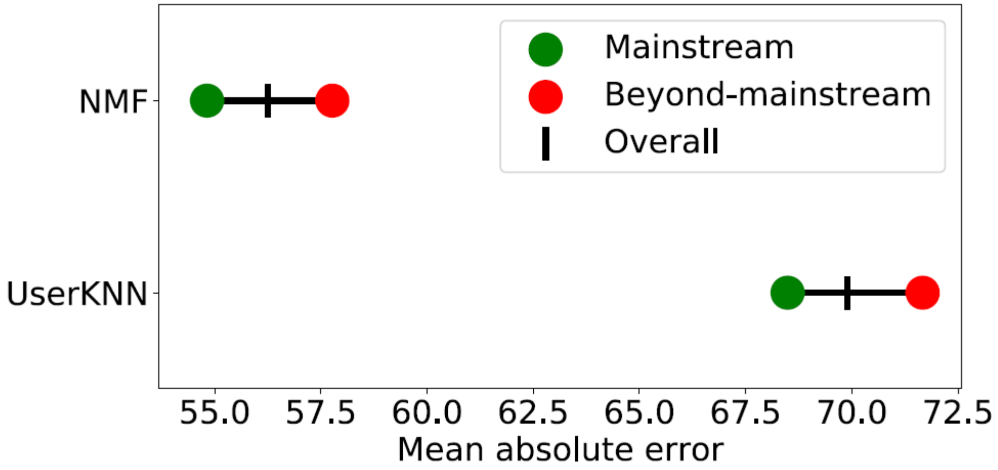
Fairness & Popularity Bias in RecSys

(Music RecSys)



(a) Random. (b) MostPopular. (c) UserItemAvg.

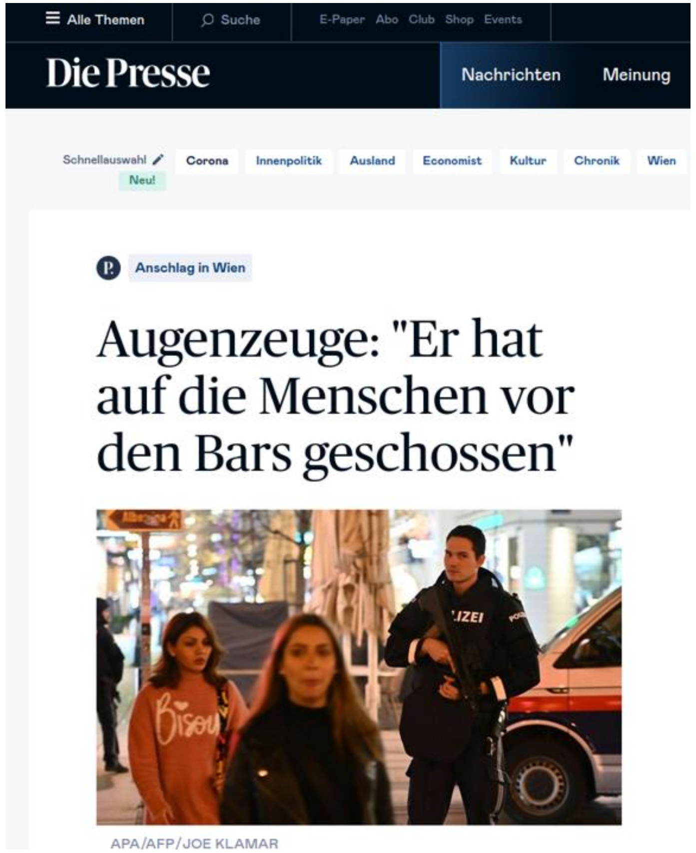
(d) UserKNN. (e) UserKNNAvg. (f) NMF.

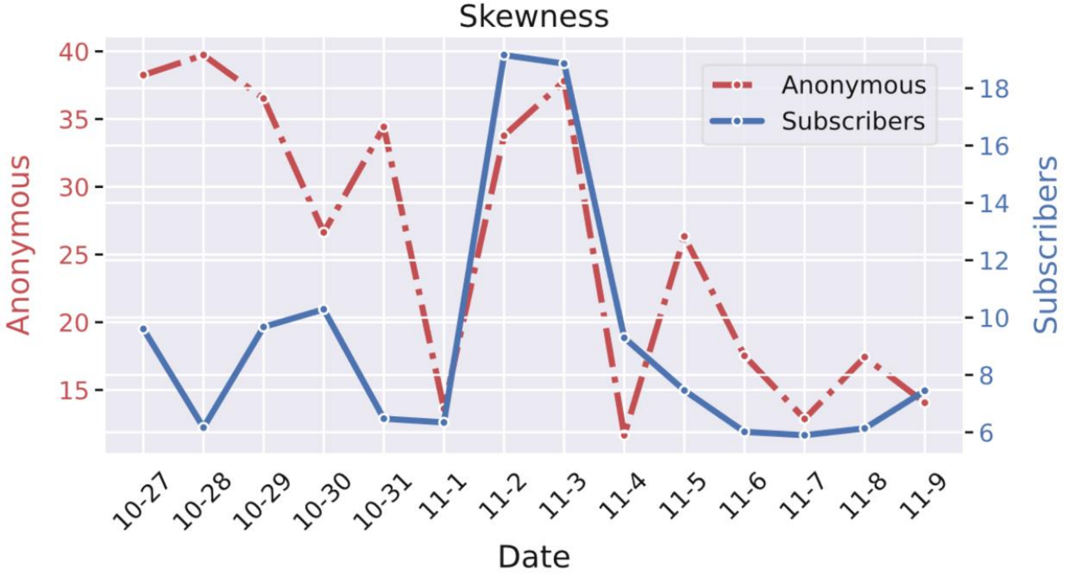


1. Kowald, D., Schedl, M., & Lex, E. (2020). The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study. In *Proceedings of the 42nd European Conference on Information Retrieval (ECIR'2020)*. Springer.
2. Lesota, O., Melchiorre, A., Rekabsaz, N., Brandl, S., Kowald, D., Lex, E., & Schedl, M. (2021). Analyzing Item Popularity Bias of Music Recommender Systems: Are Different Genders Equally Affected?. In *Proceedings of the 15th ACM Conference on Recommender Systems (RecSys'2021).* ACM.
3. Kowald, D., Muellner, P., Zangerle, E., Bauer, C., Schedl, M. & Lex, E. (2021). Support the Underground: Characteristics of Beyond-Mainstream Music Listeners. *EPJ Data Science*. Springer.
4. Kowald, D., & Lacic, E. (2022). Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems. In *Advances in Bias and Fairness in Information Retrieval (BIAS)*. Springer.
5. Kowald, D., Mayr, G., Schedl, M., & Lex, E. (2023). A Study on Accuracy, Miscalibration, and Popularity Bias in Recommendations. In *Advances in Bias and Fairness in Information Retrieval (BIAS)*. Springer.

Fairness & Popularity Bias in RecSys

(News RecSys)





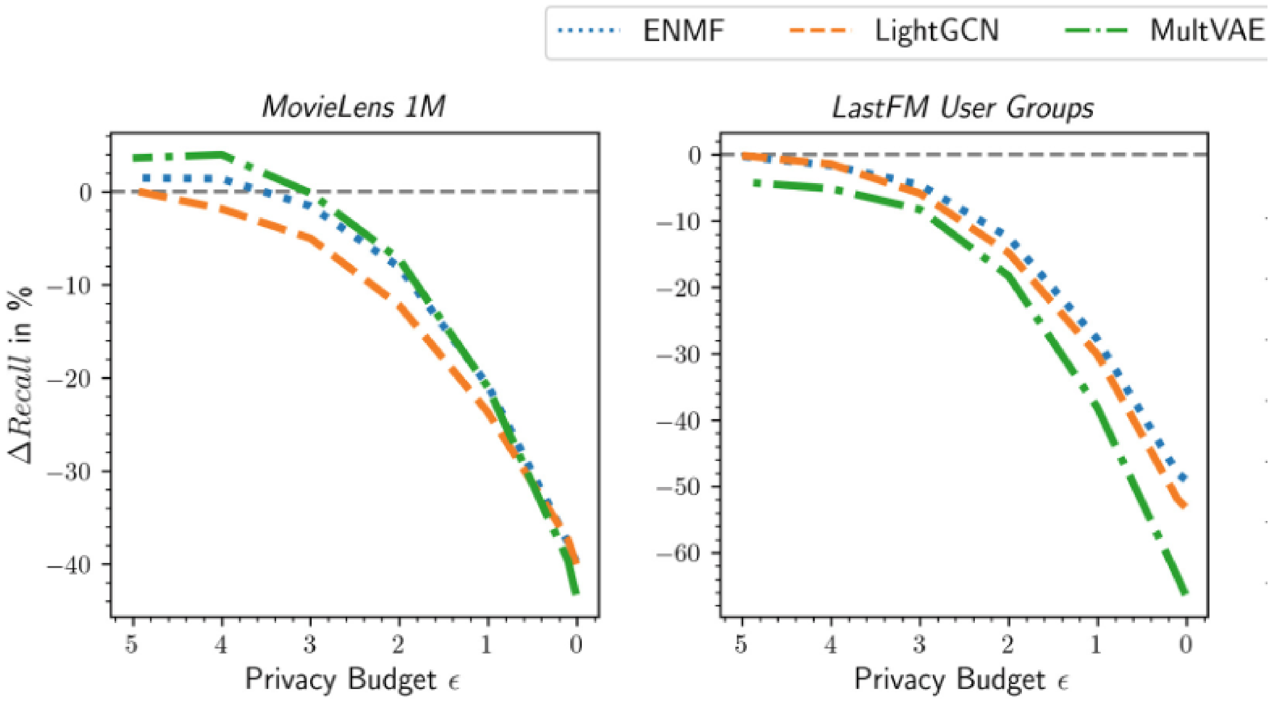
Measure	User group	Sum
No. of (users) / sessions	Anonymous	1, 182, 912
	Subscribers	(15, 910) 23, 721
	Sum	1,206,633
No. of distinct news articles	Anonymous	17, 028
	Subscribers	3, 238
	Sum	17, 372
No. of reads	Anonymous	2, 371, 451
	Subscribers	295, 416
	Sum	2, 666, 887

1. Lacic, E., Fadljevic, L., Weissenboeck, F., Lindstaedt, S., & Kowald, D. (2022). What Drives Readership? An Online Study on User Interface Types and Popularity Bias Mitigation in News Article Recommendations. In *Proceedings of the 44th European Conference on Information Retrieval (ECIR'2022)*. Springer.
2. Scher, S., Kopeinik, S., Truegler, A., & Kowald, D. (2023). Long-Term Dynamics of Fairness: Understanding the Impact of Data-Driven Targeted Help on Job Seekers. *Sci. Reports.* Nature.
3. Lesota, O., Geiger, J., Walder, M., Kowald, D., & Schedl, M. (2024). Oh, Behave! Country Representation Dynamics Created by Feedback Loops in Music Recommender Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys'2024)*. ACM.

Impact of Differential Privacy on Popularity Bias in RecSys

(Trade-Offs between Aspects)

ϵ	Model	*MovieLens 1M*		*LastFM User Groups*	
		No. Users ↓	Avg. J. ↓	No. Users ↓	Avg. J. ↓.
5	*ENMF*	99.41%	0.5118	98.06%	0.4988
	LightGCN	97.40%	0.4207	99.14%	0.5112
	MultVAE	99.71%	0.5903	99.68%	0.6983
2	*ENMF*	99.85%	0.5974	99.64%	0.5757
	LightGCN	99.86%	0.6252	99.92%	0.6518
	MultVAE	99.93%	0.6828	**100.00%**	0.7950
1	*ENMF*	99.99%	0.7006	99.95%	0.6858
	LightGCN	**99.99%**	0.7352	99.99%	0.7464
	MultVAE	**100.00%**	0.7592	**100.00%**	0.8408
0.1	*ENMF*	**100.00%**	**0.8183**	**100.00%**	**0.8058**
	LightGCN	**99.99%**	**0.8300**	**100.00%**	**0.8490**
	MultVAE	**100.00%**	**0.8447**	**100.00%**	**0.9250**

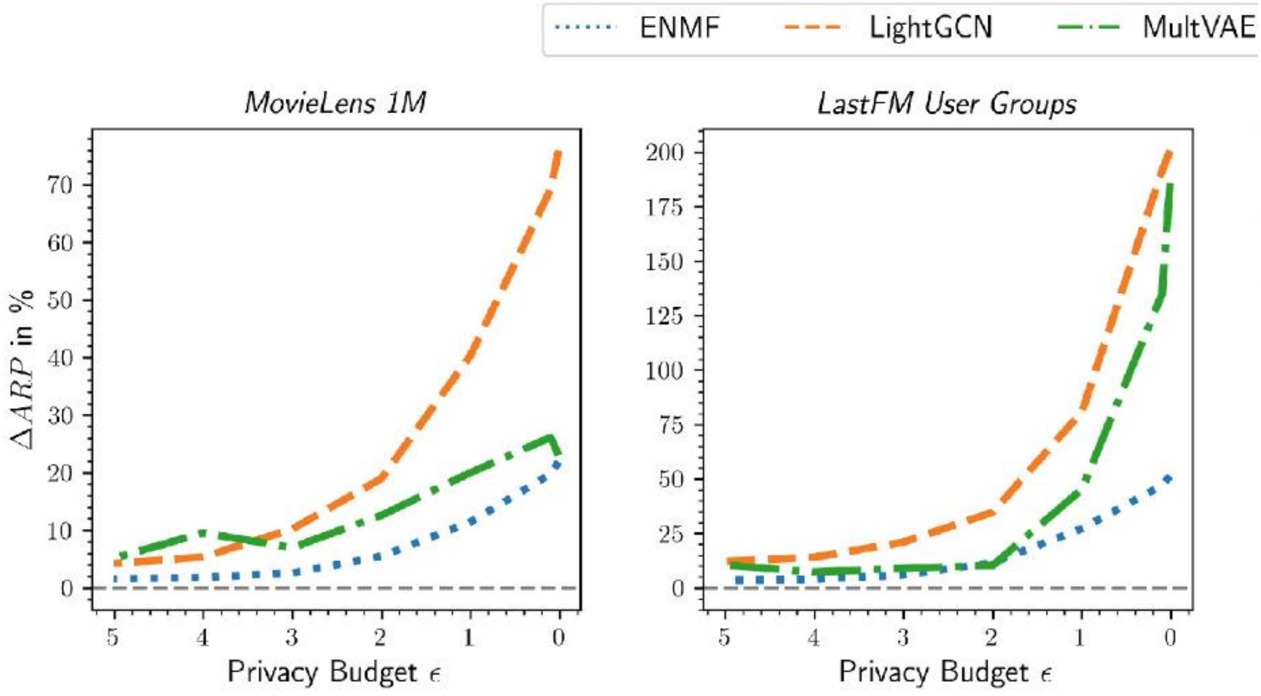


Muellner, P., Lex, E., Schedl, M., & Kowald, D. (2024). The Impact of Differential Privacy on Recommendation Accuracy and Popularity Bias. In *Proceedings of the 46th European Conference on Information Retrieval (ECIR'2024)*. Springer.

Impact of Differential Privacy on Popularity Bias in RecSys

(Trade-Offs between Aspects)

ϵ	Method	MovieLens 1M		LastFM User Groups	
		$PopLift \downarrow$ U_{low}/U_{high}	$Gap \downarrow$	$PopLift \downarrow$ U_{low}/U_{high}	$Gap \downarrow$
No DP	*ENMF*	1.0923/0.4800	0.6124	4.1028/1.1578	2.9450
	LightGCN	0.4225/0.5296	0.1072	2.7848/1.3273	1.4576
	MultVAE	0.6247/0.4901	0.1347	0.7441/0.9092	0.1651
5	*ENMF*	1.0940/0.4903	0.6037	4.0972/1.1629	2.9343
	LightGCN	0.4625/0.5566	0.0940	2.7952/1.2790	1.5162
	MultVAE	0.6538/0.5227	0.1311	0.7244/0.8928	0.1685
2	*ENMF*	1.2088/0.5147	0.6941	4.5492/1.2334	3.3158
	LightGCN	0.7447/0.6206	0.1241	3.1516/1.2894	1.8623
	MultVAE	0.8409/0.5814	0.2595	0.1894/1.0524	0.8629
1	*ENMF*	1.3658/0.5612	0.8046	5.2311/1.3309	3.9001
	LightGCN	1.1633/0.7265	0.4368	3.8118/1.5267	2.2851
	MultVAE	1.0044/0.6233	0.3811	0.3395/1.3139	**0.9744**
0.1	*ENMF*	**1.5276/0.6445**	**0.8831**	**5.7217/1.4448**	**4.2769**
	LightGCN	**1.7767/0.8415**	**0.9352**	**5.7233/1.6460**	**4.0773**
	MultVAE	**1.1595/0.6370**	**0.5225**	**1.0760/1.7873**	0.7113



Muellner, P., Lex, E., Schedl, M., & Kowald, D. (2024). The Impact of Differential Privacy on Recommendation Accuracy and Popularity Bias. In *Proceedings of the 46th European Conference on Information Retrieval (ECIR'2024).* Springer.

Summary

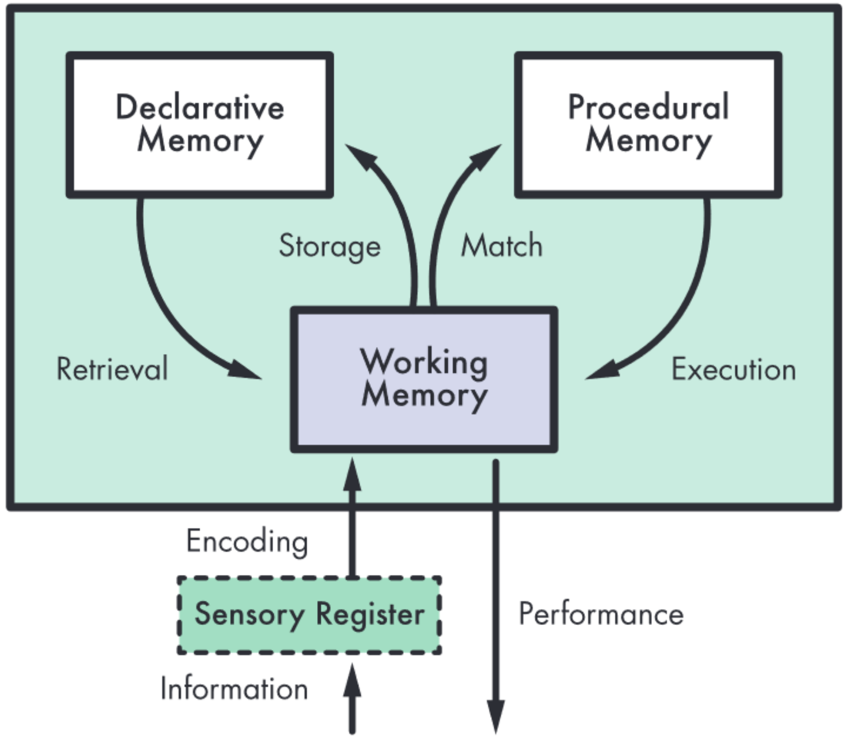
- **Transparency and cognitive models in RecSys**
  - Using cognitive models for a transparent design of RecSys
  - Using components of ACT-R to explain RecSys
- **Privacy and limited preference information in RecSys**
  - Addressing the accuracy-privacy trade-off via neighborhood reuse
  - Using variational autoencoders for session-based RecSys
- **Fairness and popularity bias in RecSys**
  - Measuring popularity bias for user groups differing in mainstreaminess
  - Analyzing popularity bias mitigation in news recommendations

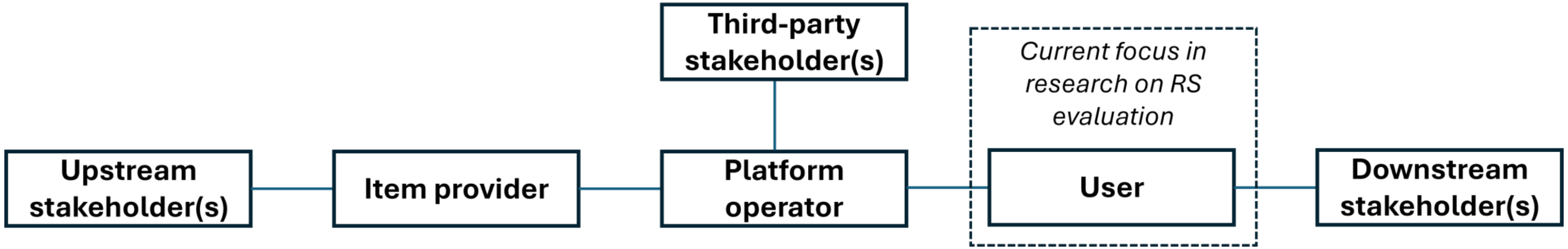
- **Impact of differential privacy on popularity bias in RecSys**
  - Trade-offs between privacy, accuracy, popularity bias, and fairness

Future Research Directions

- **Trade-offs between objectives**
- **Long-term aspects of trustworthiness**
- **Reproducibility of AI/RecSys research**

- **Additional ACT-R components**
  - **Hybrid AI (subsymbol/symbolic)**
- **Multistakeholder aspects**



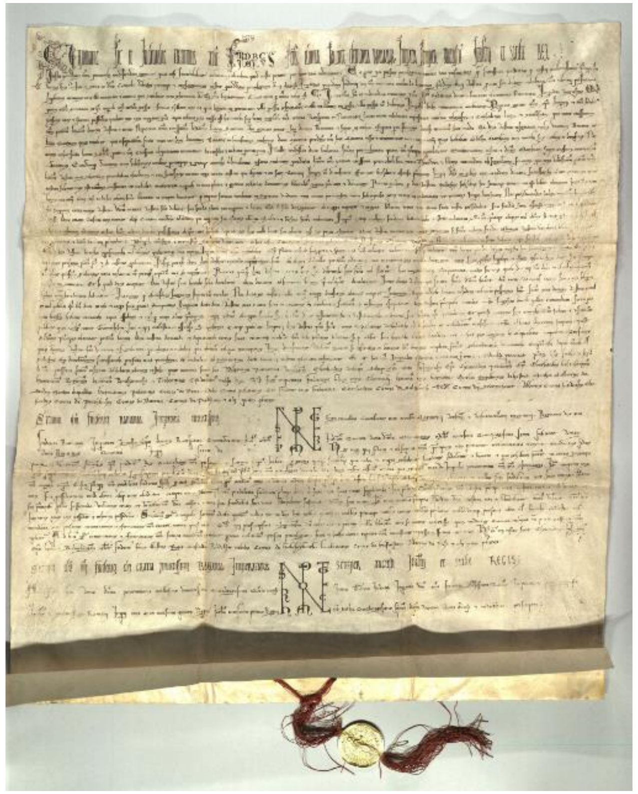


1. Kowald, D. (2017). Modeling Activation Processes in Human Memory to Improve Tag Recommendations. *PhD. Thesis.* TU Graz.
2. Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Habler, A., Thalmann, S. & Kowald, D. (2023). Reproducibility in Machine Learning-based Research: Overview, Barriers and Drivers. *AI Magazine (in revision)*
3. Burke, R., Adomavicius, G, Bogers, T., Di Noia, T., Kowald, D., Neidhardt, J., Özgöbek, Ö., Pera, M. S., Tintarev, N., Ziegler, J. (2024). Multistakeholder and Multimethod Evaluation. *Dagstuhl Seminar Report on "Evaluation Perspectives of Recommender Systems"*

Future Research Directions

(Multistakeholder RecSys for Humanities and Historical Research)

- **Monasterium.net**
  - **650k historical documents**



Category	Example	Objectives and Values
Upstream	Archives, museums	Ensure high-quality curations Impact scholarly work through recommendations Achieve visibility and prestige
Providers	Digitization services	Provide accurate and comprehensive data Facilitate efficient data access Increase exposure of items through recommendations
System	Platform owners	Drive strategic growth of the platform Foster collaboration with stakeholders Increase user satisfaction with recommendations
Consumers / users	Researchers	Obtain relevant recommendations Rely on accurate and useful results Generate new insights and research questions
Downstream	Publishers (e.g., journals)	Increase novelty of published works Publish accurate and impactful research Generate revenue through publications
Third Parties	Funding agencies	Support sustainability and accessibility of data Promote extensive use of funded curations Ensure maximum impact of funded projects

Atzenhofer-Baumgartner, F., Geiger, F., Vogeler, G., & Kowald, D. (2024). Value Identification in Multistakeholder Recommender Systems for Humanities and Historical Research: The Case of the Digital Archive Monasterium net. *NORMalize workshop co-located with ACM Conference on Recommender Systems (RecSys'24).* ACM.

Thank you for listening!

Questions? Comments? Want to collaborate?

Priv.-Doz. Dipl.Ing. Dr.techn. Dominik Kowald, BSc

Know Center Research GmbH (Fair-AI) & TU Graz (ISDS)

Sandgasse 34/2, 8010 Graz

- Mail: *dkowald [AT] know-center [.] at / dominik.kowald [AT] tugraz [.] at*
- Web: https://dominikkowald.info/
- Scholar: https://scholar.google.at/citations?user=qQ-L8rUAAAAJ&hl=en
- GitHub: https://github.com/domkowald/