



PONTIFICIA  
UNIVERSIDAD  
CATÓLICA  
DE CHILE



HAI  
—VIS

CENIA  
CENTRO NACIONAL DE INTELIGENCIA ARTIFICIAL

ihealth

# The road from Explainability in Recommender Systems to Visual XAI

Denis Parra

PUC Chile & CENIA, iHealth, IMF

IR seminar series, University Of Glasgow  
October 17th 2022



@denisparra

# Presents ...

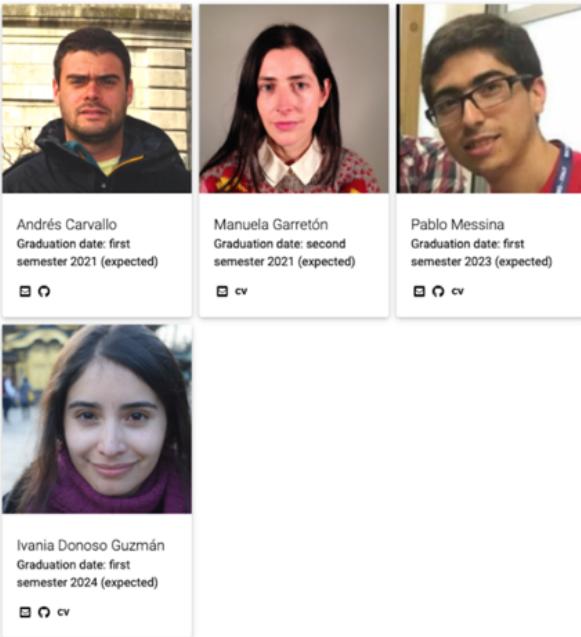
- Associate Professor, CS Department, PUC Chile
- Teaching: Recommender Systems, Information Visualization
- Research: HAIVis Lab ⊂ IA Lab
  - Machine learning applications (RecSys, Medical & Creative AI), intelligent user interfaces, information visualization
  - Supervising: 5 PhD, 5 Master, 2 Undergraduate students
- Principal Researcher at:
  - Chile's National Center on Artificial Intelligence (CENIA, 2021 - )
  - Millennium Institute for Healthcare Engineering (iHealth, 2021 - )
  - Millennium Institute Foundational Research on Data (IMFD, 2018 - )



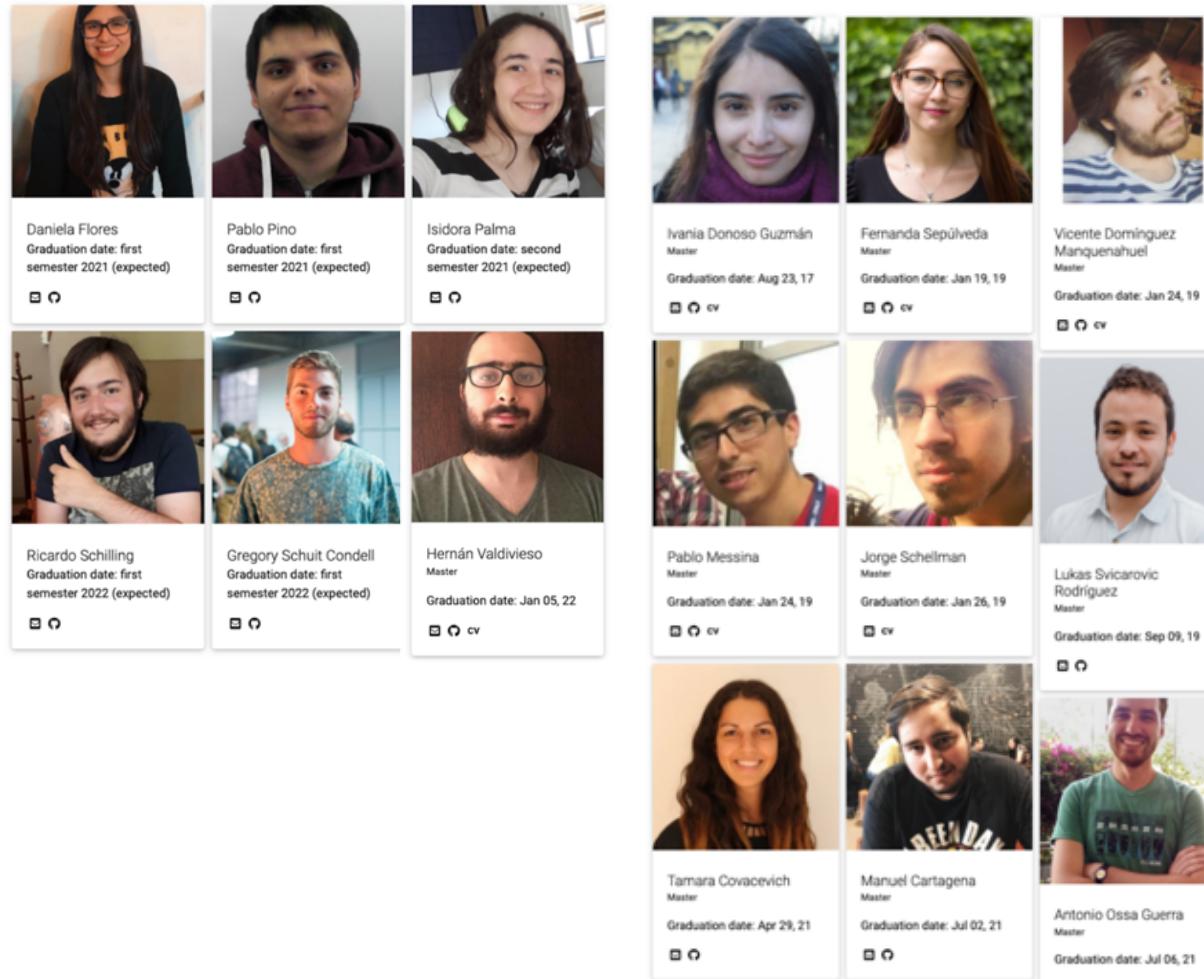
**Denis Parra**

# HAIVis UC: This research has been possible thanks to

PhD Students

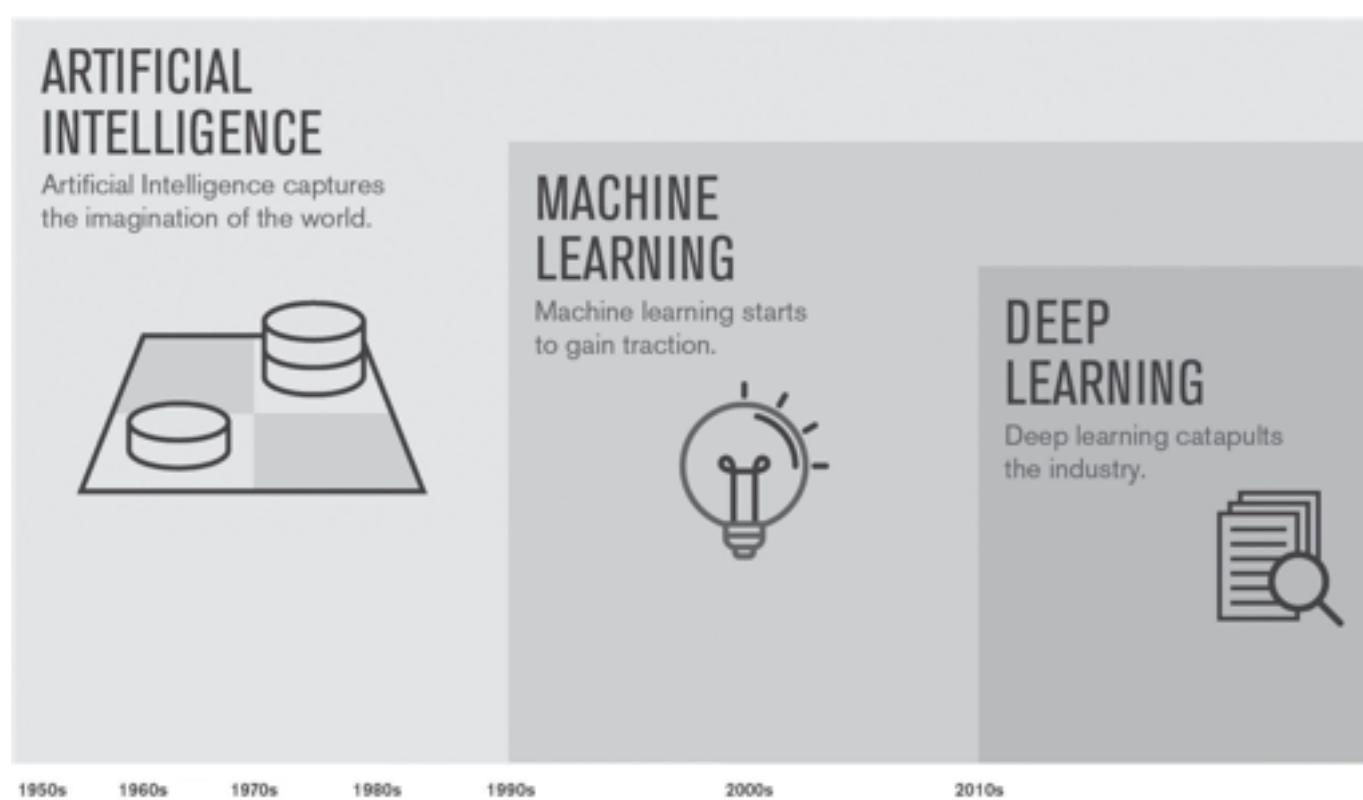


Master's students



# We are living incredible days...

- Technology is showing results which resemble science fiction, specially in the area of Artificial Intelligence.



# Self-Driving Cars



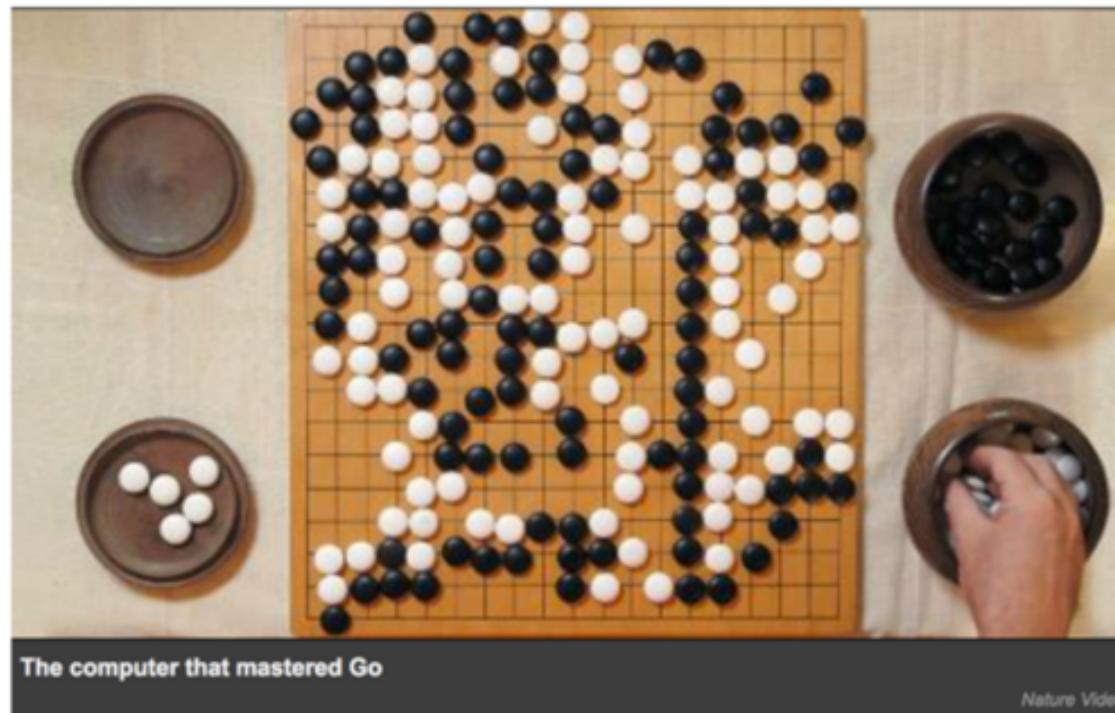
# Mastering Go

Google AI algorithm masters ancient game of Go

Deep-learning software defeats human professional for first time.

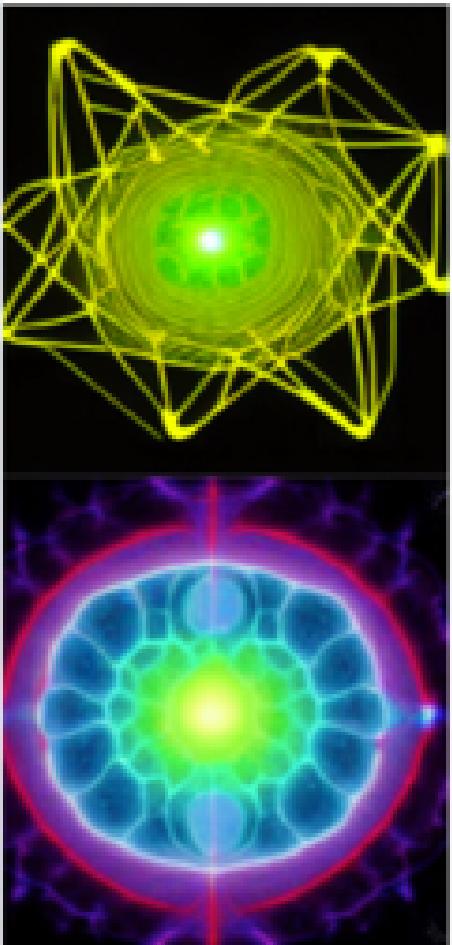
Elizabeth Gibney

27 January 2016



# Text-to-Image Generation

*'An illustration of a slightly conscious neural network'*



*'A painting of a squirrel eating a burger'*



*'A watercolor painting of a chair that looks like an octopus'*



*'A shirt with the inscription: "I love generative models!"'*



<https://github.com/CompVis/stable-diffusion/blob/main/assets/txt2img-preview.png>

# But there are some problems with AI



## Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica  
May 25, 2016

**O**N A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

- When the COMPAS system correctly predicts recidivism, it does it similarly to black and white,
- But, when it fails to predict correctly:

### Prediction Fails Differently for Black Defendants

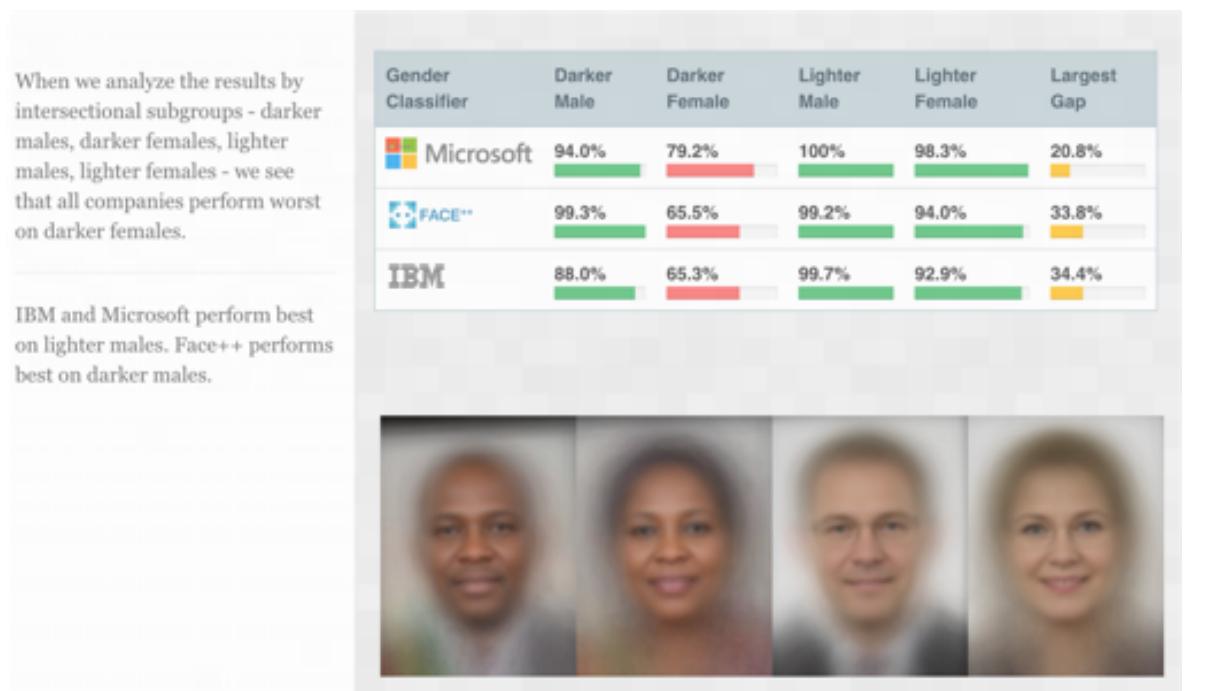
	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

# Other case: Gender Shades

- A Project by Joy Buolamwini, researcher at MIT Media Lab
- Examination of facial-analysis software shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women.



<http://gendershades.org/overview.html>

<https://www.media.mit.edu/projects/gender-shades/overview/>

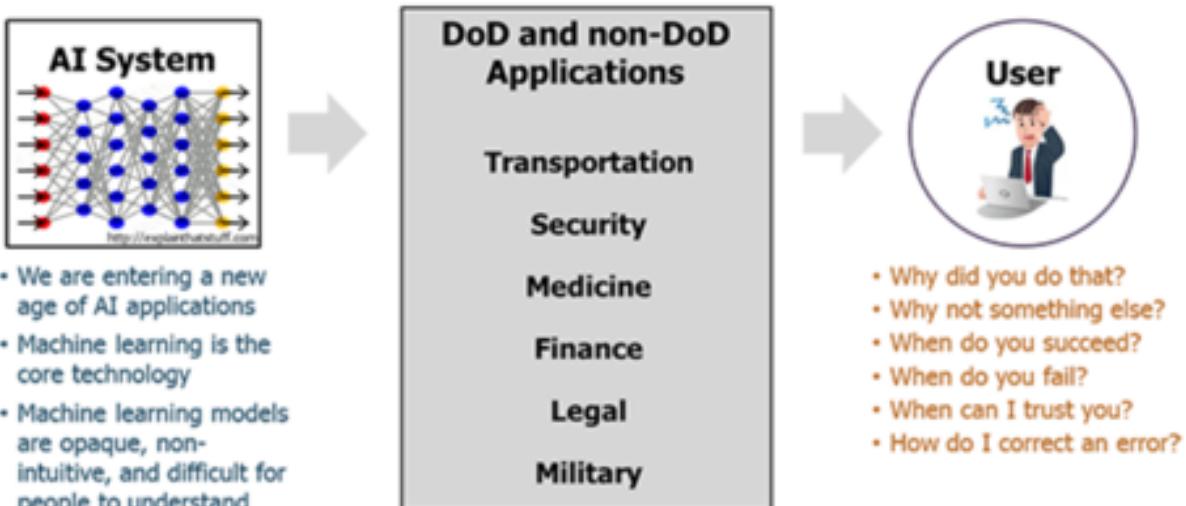
<https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>



# XAI (2017)

- XAI is a term coined by David Gunning, program manager at DARPA

**Explainable Artificial Intelligence (XAI)**  
Mr. David Gunning



Mr. David Gunning  
Information Innovation Office (I2O)  
Program Manager

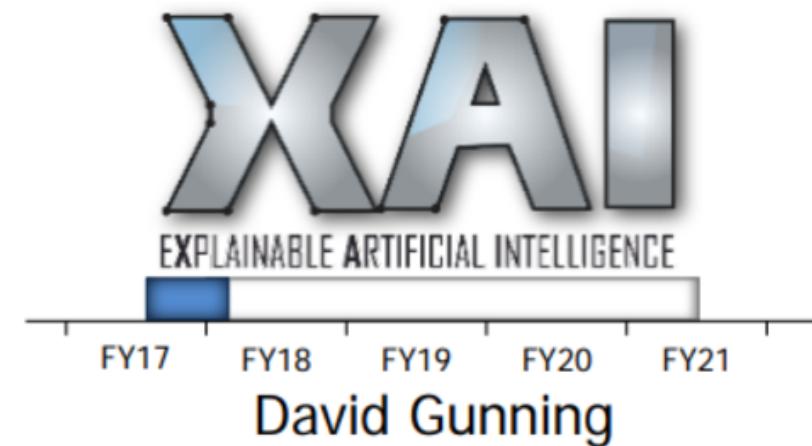


Figure 1. The Need for Explainable AI

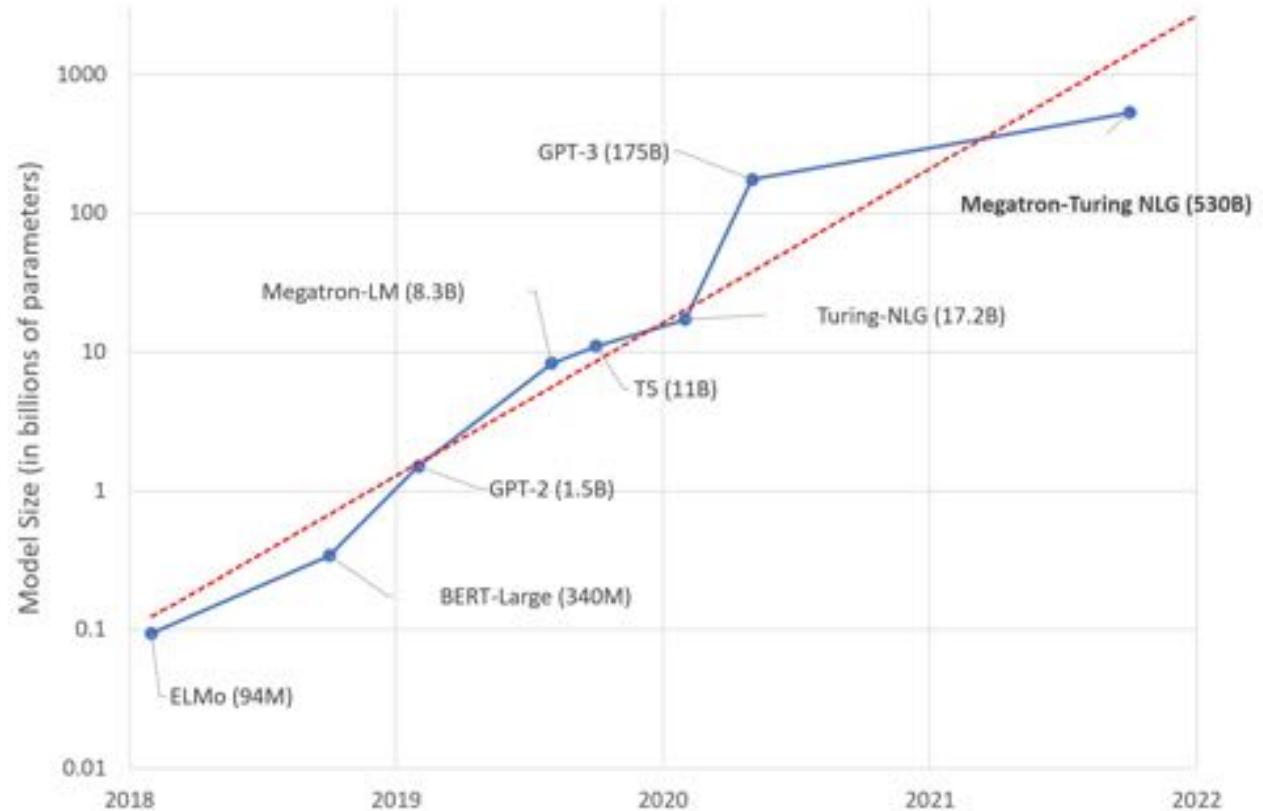
# What does XAI mean ?

- In Gunning et al (2019), an X-AI system is defined by its *Purpose*:
  - to make its behavior more intelligible to humans by providing explanations.
- Some general principles mentioned: The XAI system should be able to
  - explain its **capabilities** and **understandings**;
  - explain what it **has done**, what it is **doing now**, and what will **happen next**; and
  - **disclose the salient information** that it is acting on

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37), eaay7120.

# AI growing complexity

- With the growing complexity of AI models, the need for understanding their inner-workings has increased



**Simon, J. (2021) Large Language Models: A New Moore's Law?**  
<https://huggingface.co/blog/large-language-models>

# One approach: Visual XAI

- With the growing complexity of AI models, the need for understanding their inner-workings has increased
- Information Visualization can play an important role to address this need:
  - [visxai.io – 5<sup>th</sup> VisXAI at IEEE Vis]

## 5<sup>th</sup> Workshop on Visualization for AI Explainability

October 16th or 17th, 2022 at IEEE VIS in Oklahoma City, Oklahoma

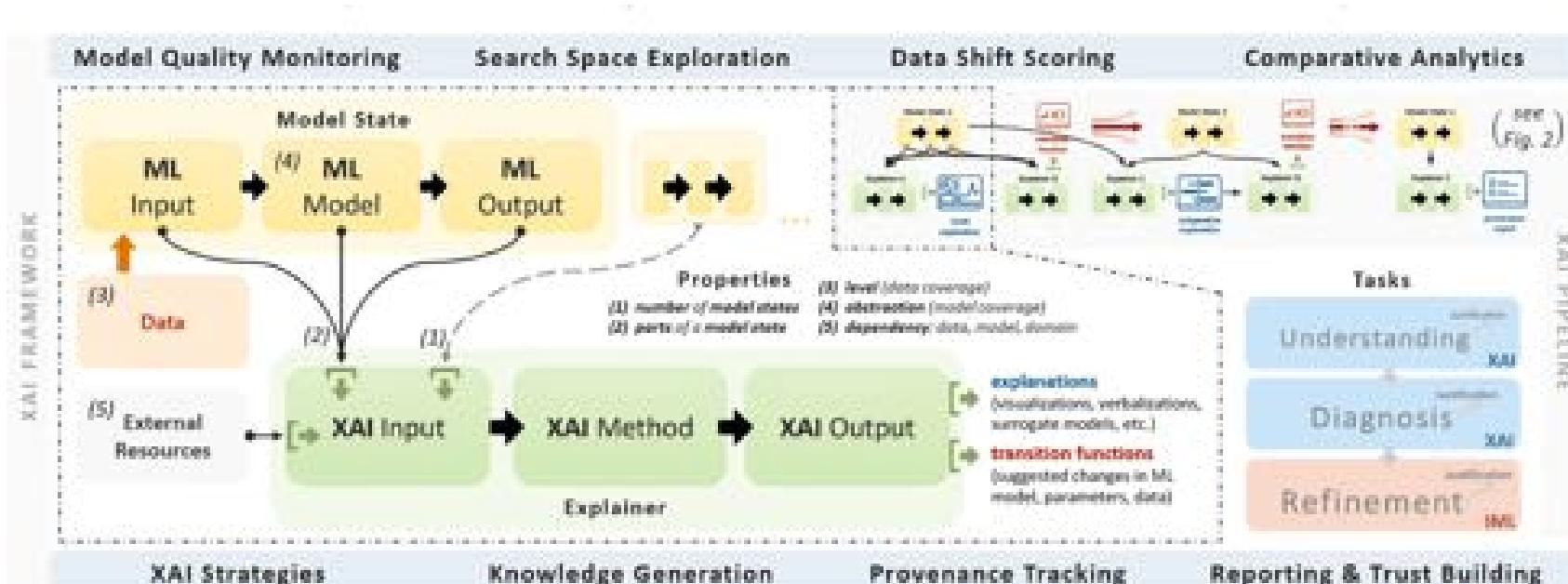
The role of visualization in artificial intelligence (AI) gained significant attention in recent years. With the growing complexity of AI models, the critical need for understanding their inner-workings has increased. Visualization is potentially a powerful technique to fill such a critical need.

The goal of this workshop is to initiate a call for "explainables" / "explorables" that explain how AI techniques work using visualization. We believe the VIS community can leverage their expertise in creating visual narratives to bring new insight into the often obfuscated complexity of AI systems.



# Visual XAI Frameworks

- Several approaches leveraging Visualization Systems have been introduced in the latest years, such as the **ExplAIner** framework proposed in 2019:

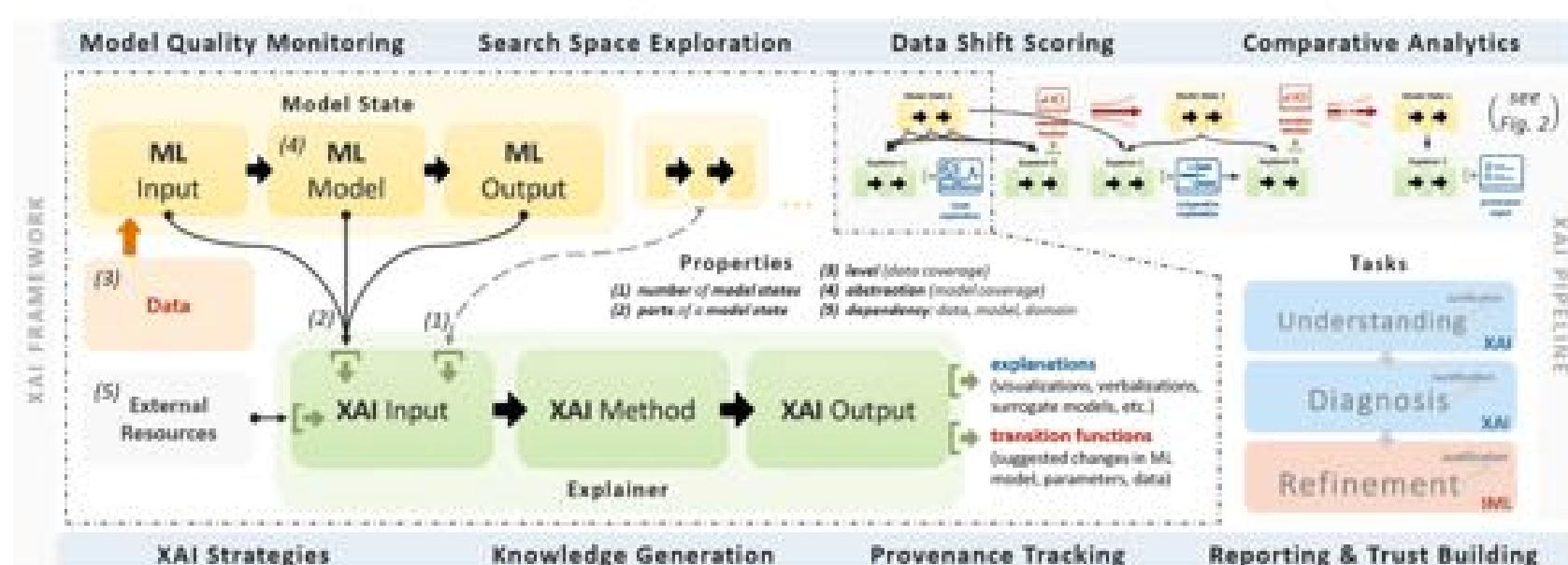


Spinner, T., et al (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE TVCG, 26(1), 1064-1074.

# Visual XAI Frameworks

- Several approaches leveraging Visualization Systems have been introduced in the latest years, and even a framework was proposed in 2019: ExplAIner

ExplAIner tells you the steps towards creating a Visual XAI, but Not How to **design** and **analyze** a visualization for XAI



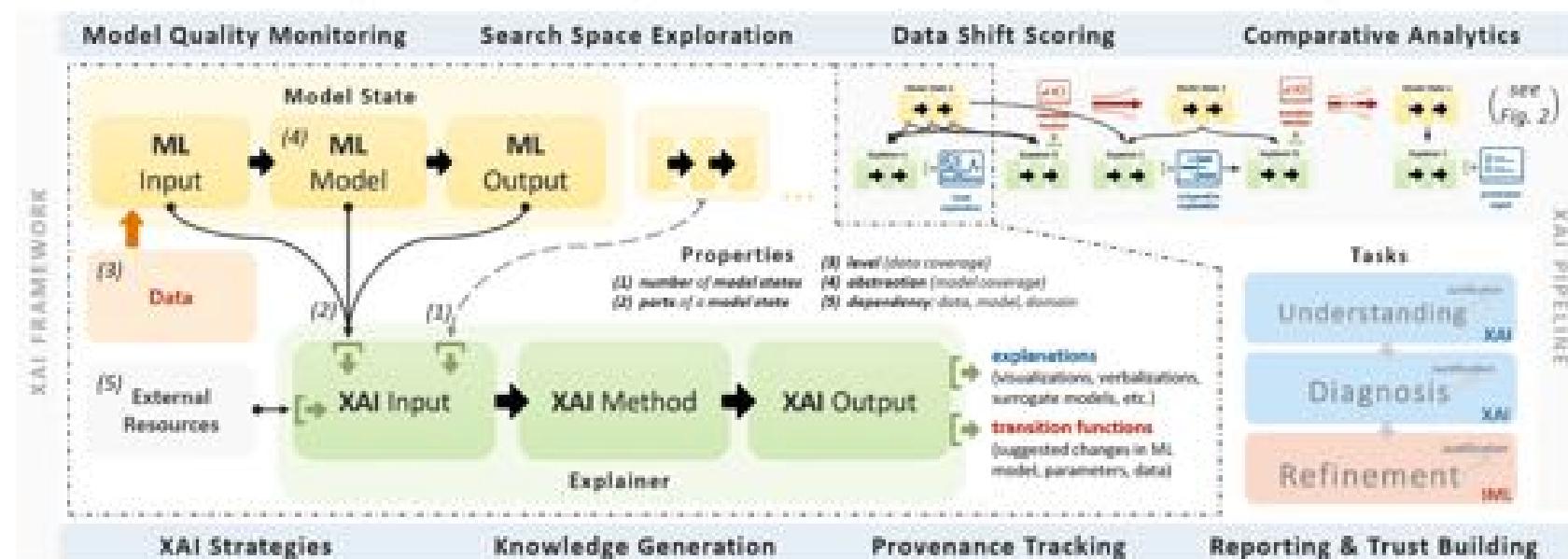
Spinner, T., et al (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE TVCG, 26(1), 1064-1074.

# Visual XAI Frameworks

- Several approaches leveraging Visualization Systems have been introduced in the latest years, and even a framework was proposed in 2019: ExplAIner

ExplAIner tells you the steps towards creating a Visual XAI, but Not How to **design** and analyze a visualization for XAI

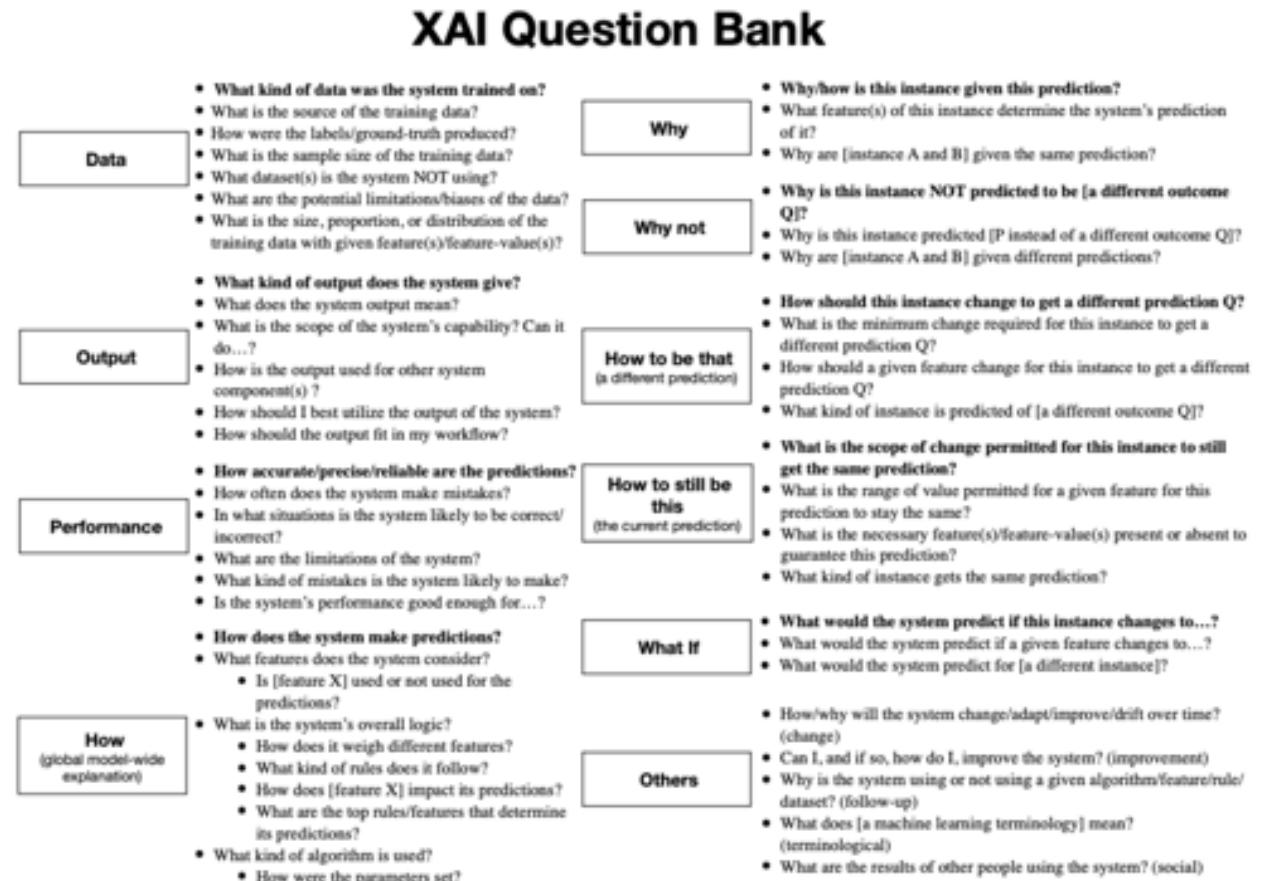
- Visual
- Interaction



Spinner, T., et al (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE TVCG, 26(1), 1064-1074.

# XAI question bank

- Liao, Gruen and Miller introduced a XAI question bank to guide the design process of XAI systems
- But they still do not provide clear guidelines for visual XAI applications



Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: informing design practices for explainable AI user experiences. CHI 2020.

# XAI Challenges – Adaptation & Personalization

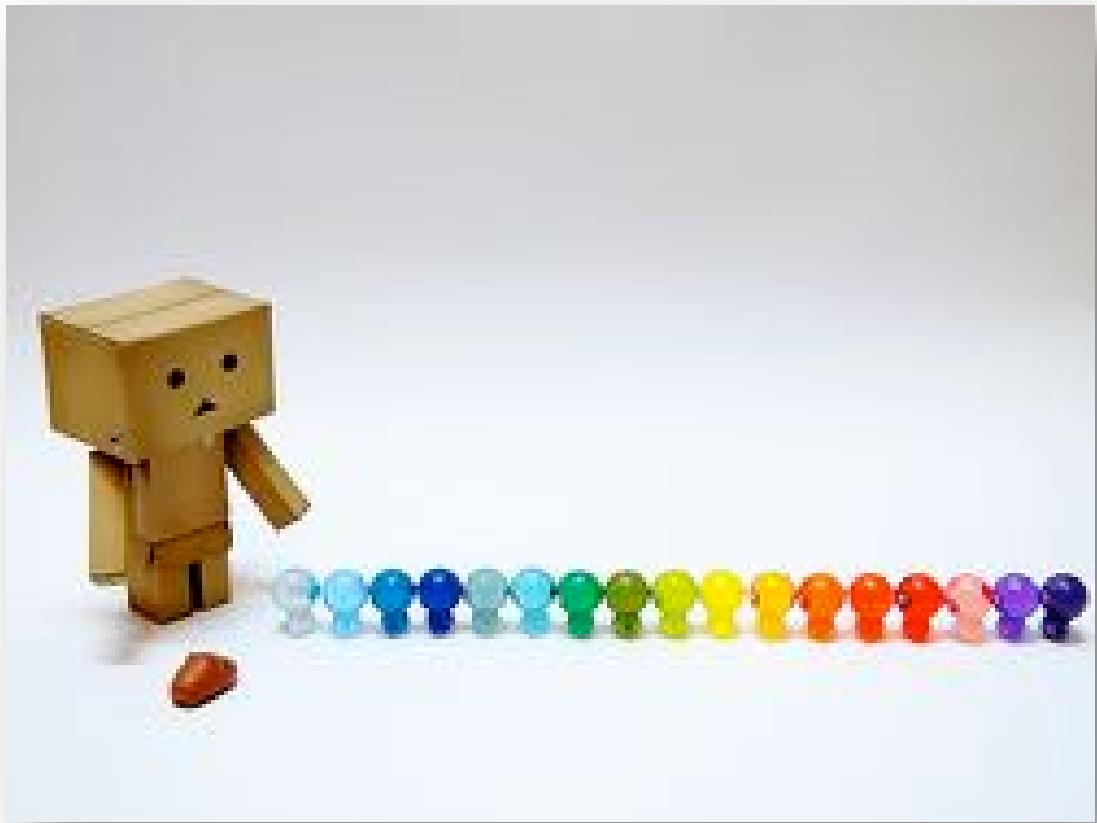
- An effective explanation will take the target user group of the system into account, **who might vary** in their **background knowledge** and **needs** for what should be explained.
  - How should we proceed under these circumstances ?
- Decision making for analysts, judges and operators: **Each user group may have a preferred explanation** type that is able to communicate information in the most effective way.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37), eaay7120.

# The role of RecSys in XAI

# Recommender Systems (RecSys) and XAI

- Recommender Systems are a type of systems/technology which strongly relies on AI algorithms
- Explainability and Transparency in Recommender Systems has been studied for at least 20 years
- What can we learn from such research for XAI ?



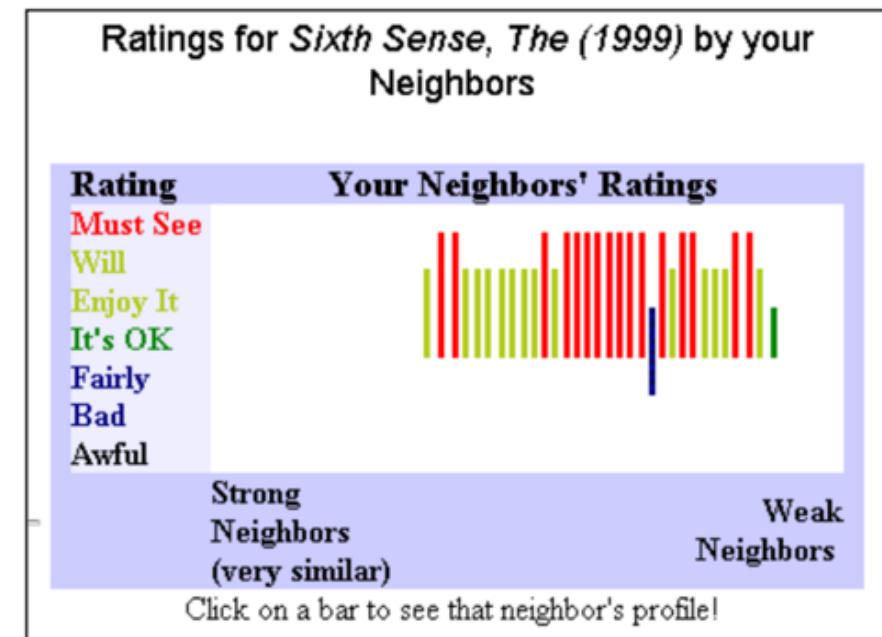
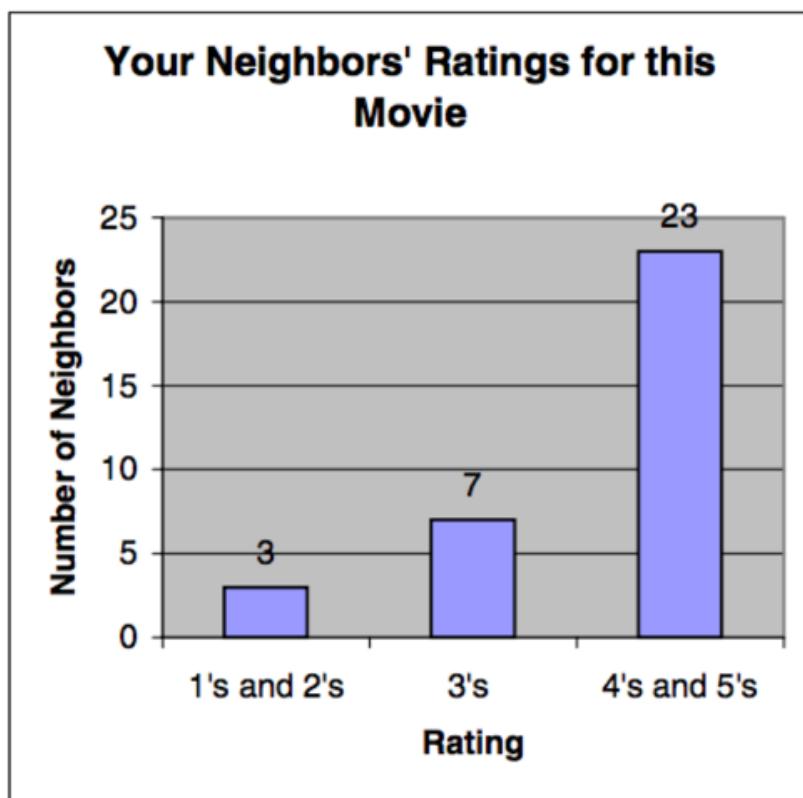
Picture licensed under CC  
<http://www.flickr.com/photos/meaganmakes/6769496875/sizes/m/>

# XAI in Recommender Systems: back to 2000s

- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). *Explaining collaborative filtering recommendations*. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 241-250). ACM.
- Sinha, R., & Swearingen, K. (2002). *The role of transparency in recommender systems*. In *CHI'02 extended abstracts on Human factors in computing systems* (pp. 830-831). ACM.

# XAI in Recommender Systems (movies)

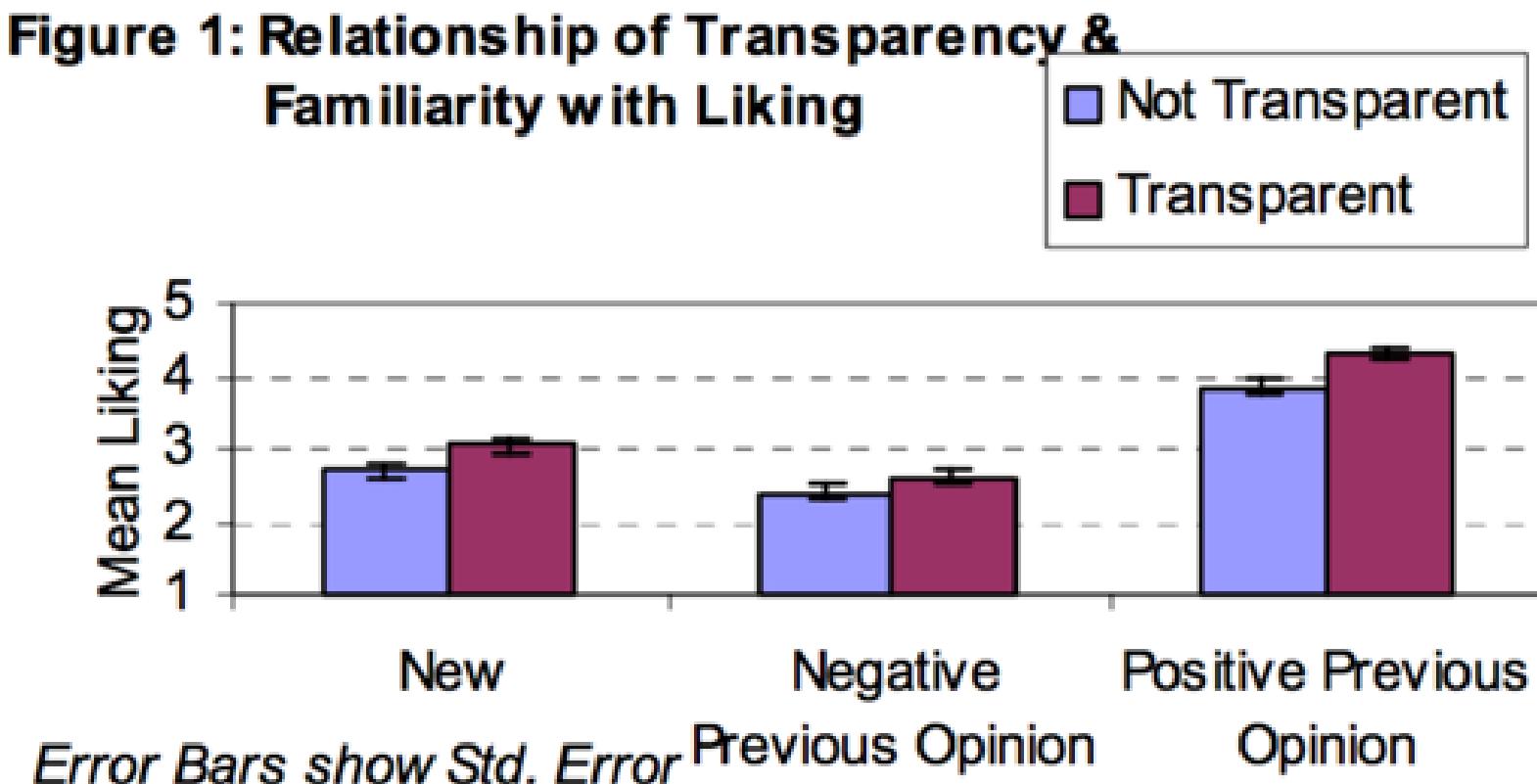
- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 241-250). ACM.



**Figure 4.** A screen explaining the recommendation for the movie “The Sixth Sense.” Each bar represents a rating of a neighbor. Upwardly trending bars are positive ratings, while downward trending ones are negative. The x-axis represents similarity to the user.

# XAI in Recommender Systems (music)

- Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems* (pp. 830-831). ACM.



# XAI in Recommender Systems II

- Tintarev, N., & Masthoff, J. (2007). A survey of **explanations in recommender systems**. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 801-810). IEEE.
- Tintarev, N., & Masthoff, J. (2012). Evaluating the **effectiveness of explanations** for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5), 399-439.
- Tintarev, N., & Masthoff, J. (2015). **Explaining recommendations: Design and evaluation**. In *Recommender systems handbook* (pp. 353-382). Springer, Boston, MA.

# RecSys: Explanatory Goals and Definitions

Aim	Definition
Transparency (Tra.)	Explain how the system works
Scrutability (Scr.)	Allow users to tell the system it is wrong
Trust	Increase users' confidence in the system
Effectiveness (Efk.)	Help users make good decisions
Persuasiveness (Pers.)	Convince users to try or buy
Efficiency (Efc.)	Help users make decisions faster
Satisfaction (Sat.)	Increase the ease of usability or enjoyment

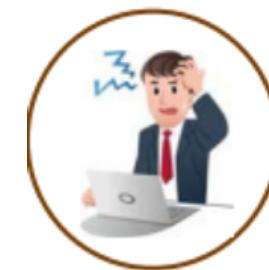
Tintarev, N., & Masthoff, J. (2007). A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 801-810). IEEE.

# RecSys: Explanatory Goals and Definitions

Aim	Definition
Transparency (Tra.)	Explain how the system works
Scrutability (Scr.)	Allow users to tell the system it is wrong
Trust	Increase users' confidence in the system
Effectiveness (Efk.)	Help users make good decisions
Persuasiveness (Pers.)	Convince users to try or buy
Efficiency (Efc.)	Help users make decisions faster
Satisfaction (Sat.)	Increase the ease of usability or enjoyment

2007

2017



User with a Task

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?



User with a Task

- I understand why
- I understand why not
- I know when you'll succeed
- I know when you'll fail
- I know when to trust you
- I know why you erred

Tintarev, N., & Masthoff, J. (2007). A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 801-810). IEEE.

# Did we do nothing between 2007-17 ?

- There were actually many works related to explainability, transparency, trust and user controllability in RecSys.
- In the coming slides, I will focus on research related **Visual User Interfaces** for recommendation Systems and how they contributed early to XAI and Visual XAI.

# Let's take a trip

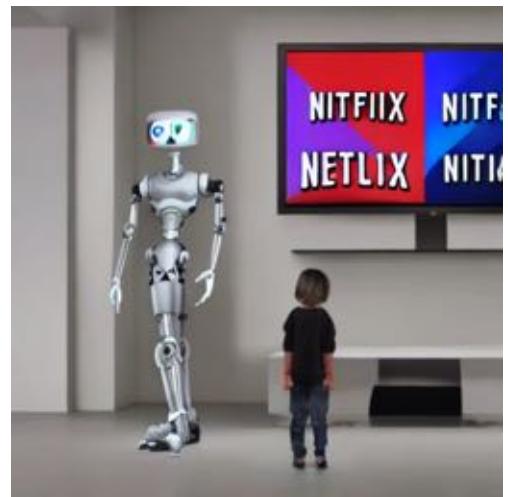
- ... through the history of explainability approaches **with and without** visualizations in RecSys
- (images generated with S.D.)

## Stable Diffusion Demo

Stable Diffusion is a state of the art text-to-image model that generates images from text.  
For faster generation and forthcoming API access you can try [DreamStudio Beta](#)

A robot explains netflix algorithm to a human using deep neural networks

Generate Image



# Non-visual approaches for XAI in RecSys

- A brief non-exhaustive list:

Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., & Ma, S. (2014). **Explicit factor models for explainable recommendation based on phrase-level sentiment analysis.** In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 83-92). ACM.

Seo, S., Huang, J., Yang, H., & Liu, Y. (2017). **Interpretable convolutional neural networks with dual local and global attention for review rating prediction.** RecSys 2017.

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. S. (2017). **Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention.** SIGIR 2017.

Adapting LIME to recommendation: Nóbrega, C., & Marinho, L. (2019). **Towards explaining recommendations through local surrogate models.** ACM/SIGAPP Symposium on Applied Computing.

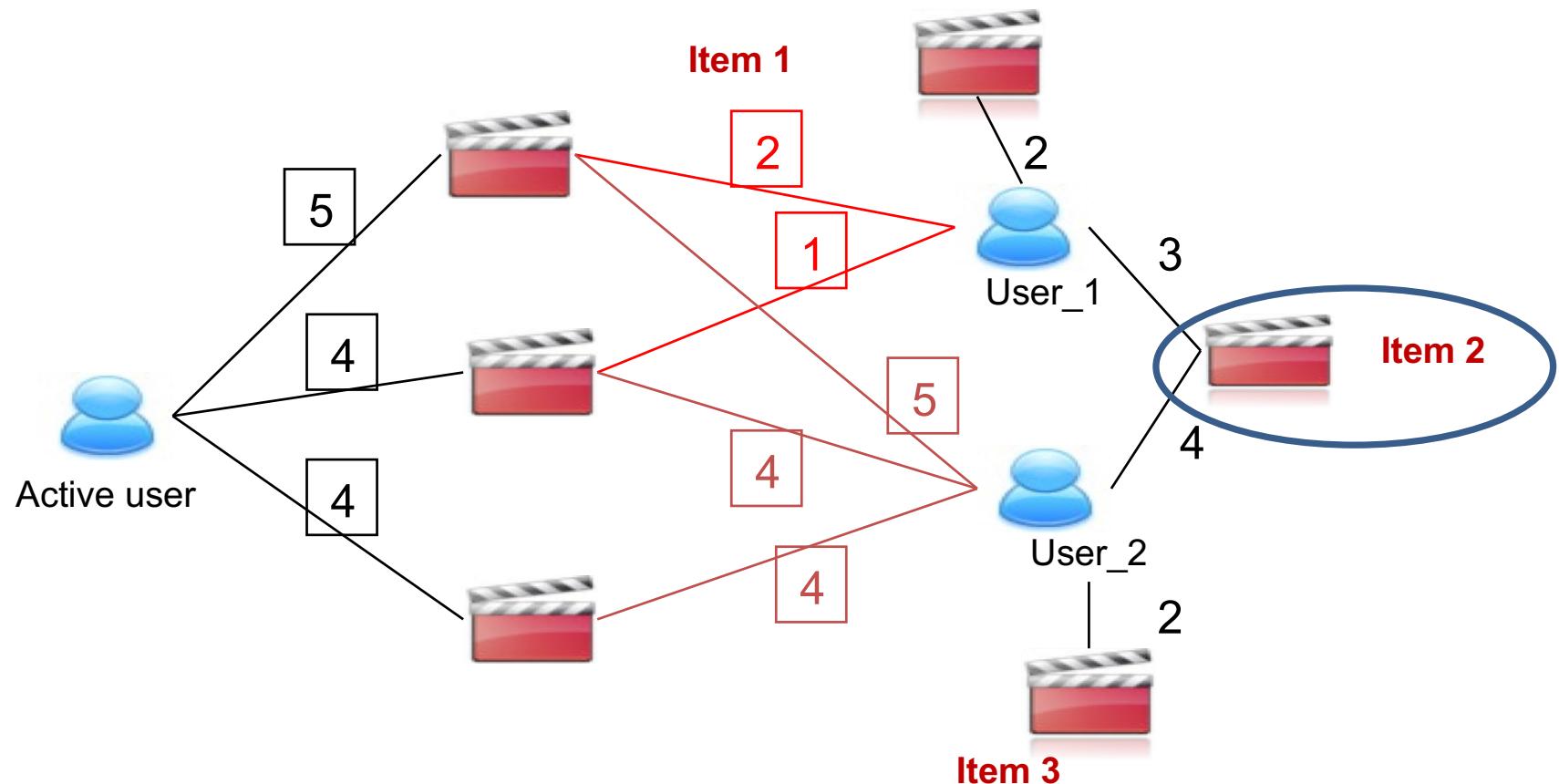
More information at this Tutorial web page: <https://sites.google.com/view/ears-tutorial/>

# 1<sup>st</sup> generation of RecSys

- Algorithms were simple and intuitive (User-based KNN, Item-Based KNN, Content-based, Case-based)
- Providing explanations for items recommended would not require a big engineering effort

# 1<sup>st</sup> generation of RecSys

- User Based KNN (collaborative filtering)



**Explanation:** Users who have similar ratings with you highly rated this item

# 1<sup>st</sup> generation of RecSys

- Content Based



**Explanation:** This items has similar content (features: description, actors, director, genre) to what you have liked in the past

# XAI for Recommender Systems

- First generation of approaches for Recommender Systems were easily to explain: User and Item based CF, Content-based, Rule-based
- The Second generation of RecSys, based on Matrix Factorization made the process more difficult: latent user and item representation
- The Third generation based on Deep Learning makes accountability and transparency even more difficult !

# 2<sup>nd</sup> generation of RecSys

- Alternatives: try to assign explicit meaning to latent factor models:
  - Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., & Ma, S. (2014). **Explicit factor models for explainable recommendation based on phrase-level sentiment analysis.** In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 83-92). ACM.
  - Chen, X., Qin, Z., Zhang, Y., & Xu, T. (2016). **Learning to rank features for recommendation over multiple categories.** In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 305-314). ACM.
  - Wang, N., Wang, H., Jia, Y., & Yin, Y. (2018). **Explainable recommendation via multi-task learning in opinionated text data.** In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 165-174). ACM.

# 2<sup>nd</sup> generation of RecSys

- Zhang et al (2014) “Explicit factor models for explainable recommendation based on phrase-level sentiment analysis”

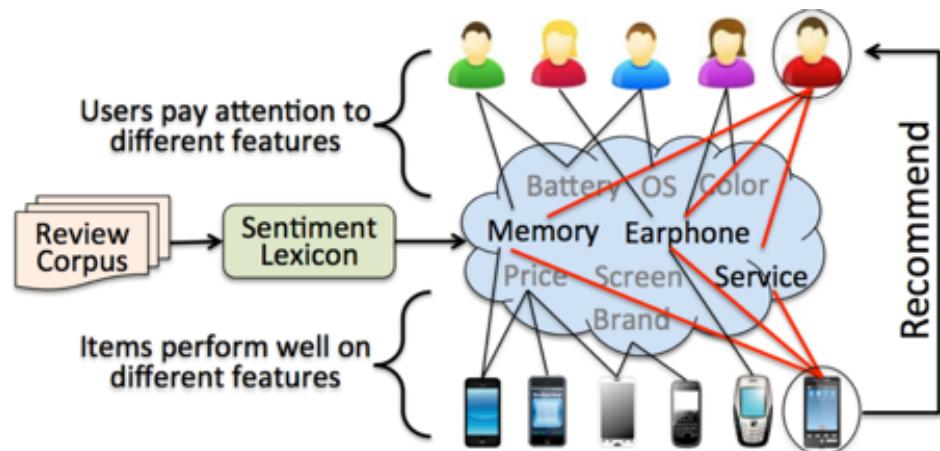


Figure 1: The product feature word and user opinion word pairs are extracted from user review corpus to construct the sentiment lexicon, and the feature word set further serves as the explicit feature space. An item would be recommended if it performs well on the features that a user cares.

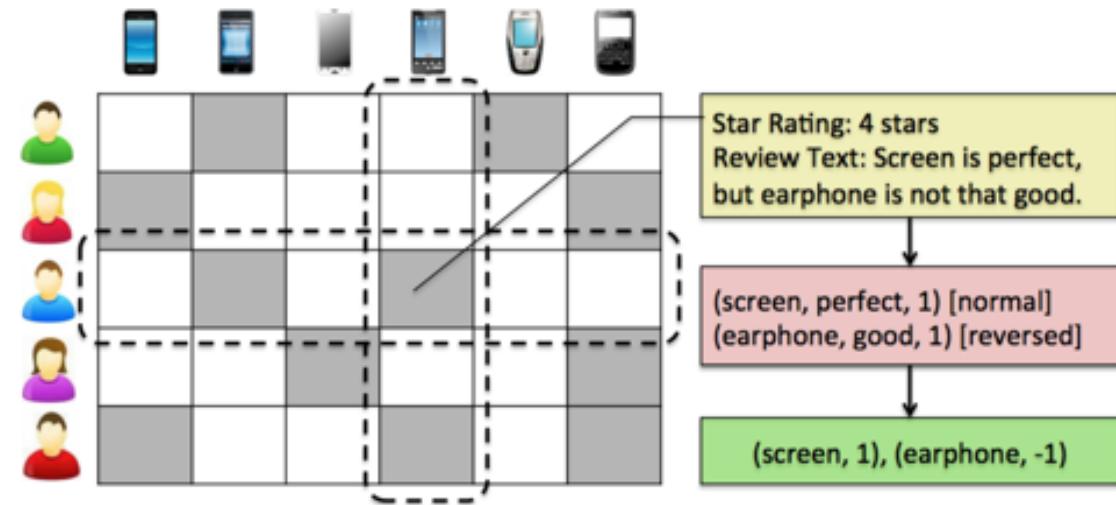


Figure 2: An example of user-item review matrix, where each shaded block is a review made by a user towards an item; the entries included in the review are extracted, and further transformed to feature scores while considering the negation words.

# XAI for Recommender Systems

- First generation of approaches for Recommender Systems were easily to explain: User and Item based CF, Content-based, Rule-based
- The Second generation of RecSys, based on Matrix Factorization made the process more difficult: latent user and item representation
- The Third generation based on Deep Learning makes accountability and transparency even more difficult !

# 3<sup>rd</sup> Generation of RecSys

- In Matrix factorization we had one level of interactions, with deep learning we can have many! Making explanations more complex

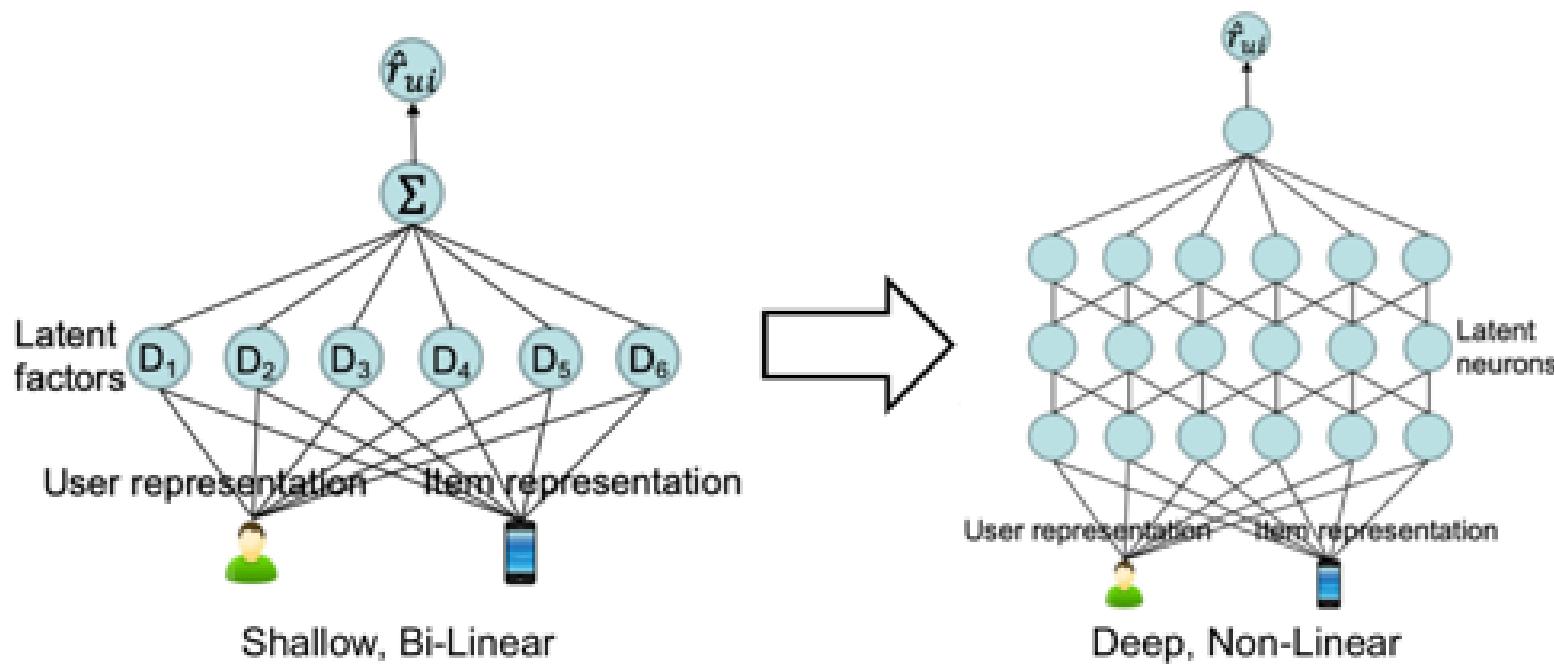


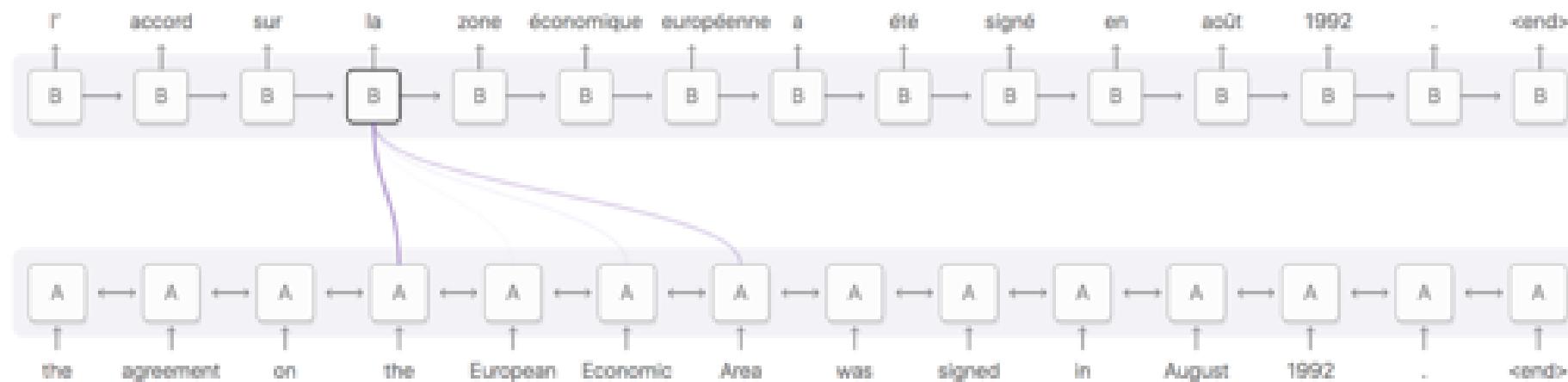
Image from Zhang et al. (2019) Tutorial on Explainable Recommendation and Search

# 3<sup>rd</sup> Generation of RecSys

- Alternatives: use attention mechanism within the neural architecture (over text or images)
- Generate explanations directly (Natural Language Generation)
- Use a model agnostic approach: generate explanations after recommendation (LIME, SHAP, etc.)

# Neural Attention

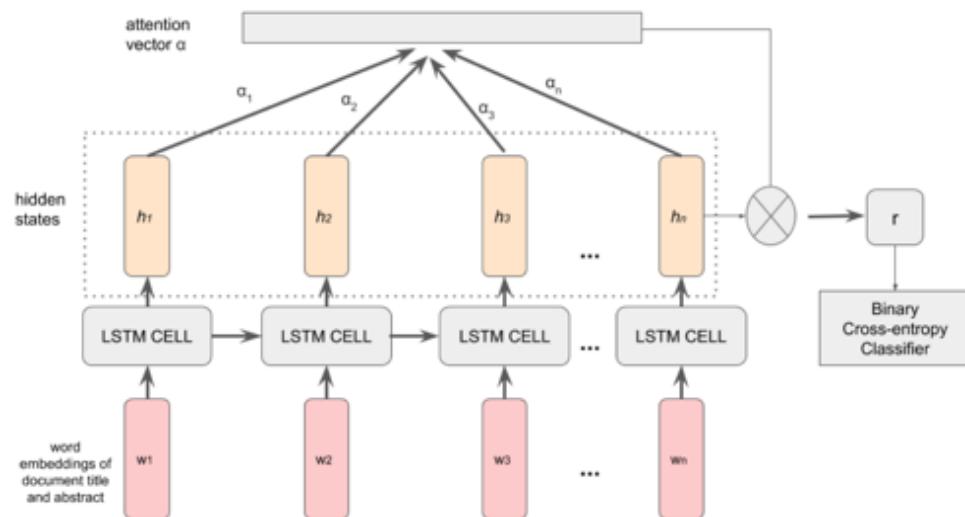
- Attention in neural networks is a mechanism which allows the model to focus selectively during the learning process.
- Eventually, we can observe where the network was attending to in order to make a prediction.



Olah, C., & Carter, S. (2016). Attention and augmented recurrent neural networks. *Distill*, 1(9), e1. <http://doi.org/10.23915/distill.00001>

# Neural Attention

- Example of document classification: Does the model attends to reasonable words ?



A meta analysis of birth origin effects on reproduction in diverse captive environments

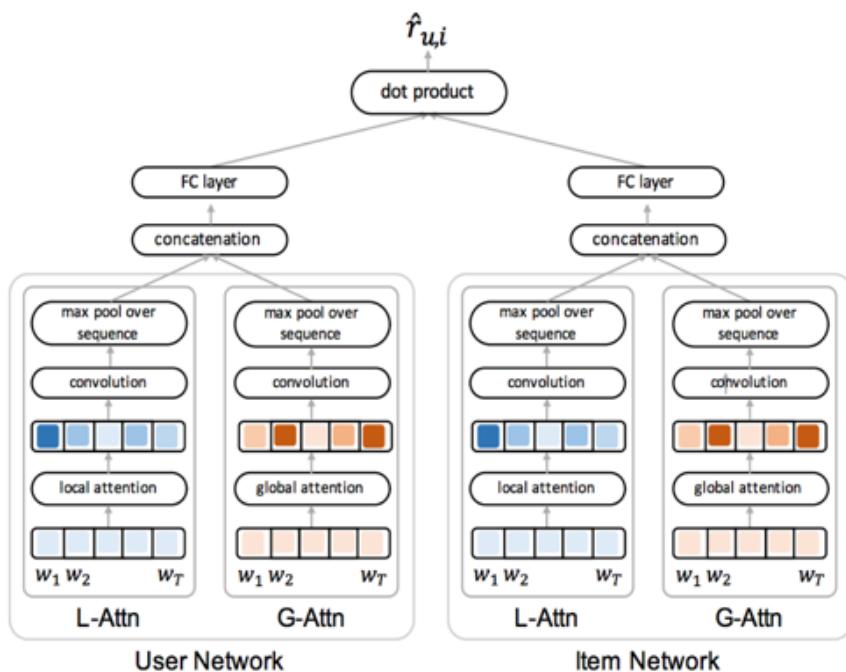
Prediction: Not Relevant (NR)

Ground truth: Not Relevant (NR)

Title: a meta analysis of birth origin effects on reproduction in diverse **captive** environments  
Abstract: successfully establishing **captive** breeding programs is priority across diverse industries to address food security demand for ethical laboratory research animals and prevent extinction differences in reproductive success due to birth origin may threaten the long term sustainability of **captive** breeding our meta analysis examining effect sizes from species of invertebrates fish **birds** and mammals shows that overall **captive** born animals have decreased odds of reproductive success in captivity compared to their wild born counterparts the largest effects are seen in commercial aquaculture relative to conservation or laboratory settings and offspring survival and offspring quality were the most sensitive traits although somewhat weaker trend reproductive success in conservation and laboratory research breeding programs is also in negative direction for **captive** born animals our study provides the foundation for future investigation of non genetic and genetic drivers of change

# 3<sup>rd</sup> Generation of RecSys

- Seo, S., Huang, J., Yang, H., & Liu, Y. (2017). **Interpretable convolutional neural networks with dual local and global attention for review rating prediction.** RecSys 2017.



**Yelp (user), L-Attn-only model: local attention**

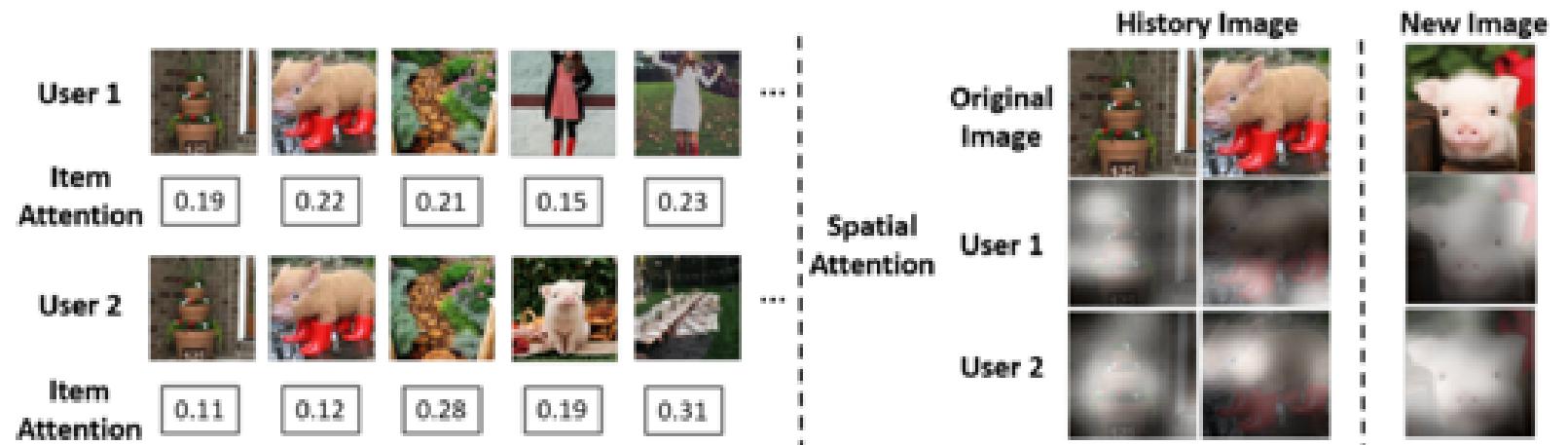
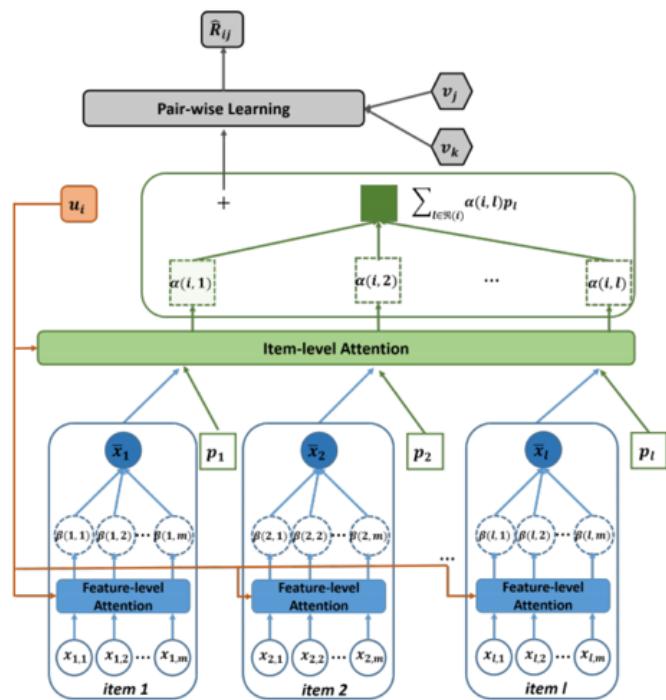
They carry some rare things that you can't find anywhere else. The staff is pretty damn cool too best in Arizona. I prefer ma-and-pa. They treat you the best and they value your business extreme. They are good people great atmosphere and music. I definitely believe that Lux has the best coffee I've ever had at this point. Screw all my previous reviews. This place has coffee down, they make damn good toast too.

**Yelp (user), G-Attn-only model: global attention**

They carry some rare things that you can't find anywhere else. The staff is pretty damn cool too best in Arizona. I prefer ma-and-pa. They treat you the best and they value your business extreme. They are good people great atmosphere and music. I definitely believe that Lux has the best coffee I've ever had at this point. Screw all my previous reviews. This place has coffee down, they make damn good toast too.

# 3<sup>rd</sup> Generation of RecSys

- Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. S. (2017). Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. SIGIR 2017.



# 3<sup>rd</sup> Generation of RecSys

- Li, P., Wang, Z., Ren, Z., Bing, L., & Lam, W. (2017,). Neural rating regression with abstractive tips generation for recommendation.
- SIGIR 2017

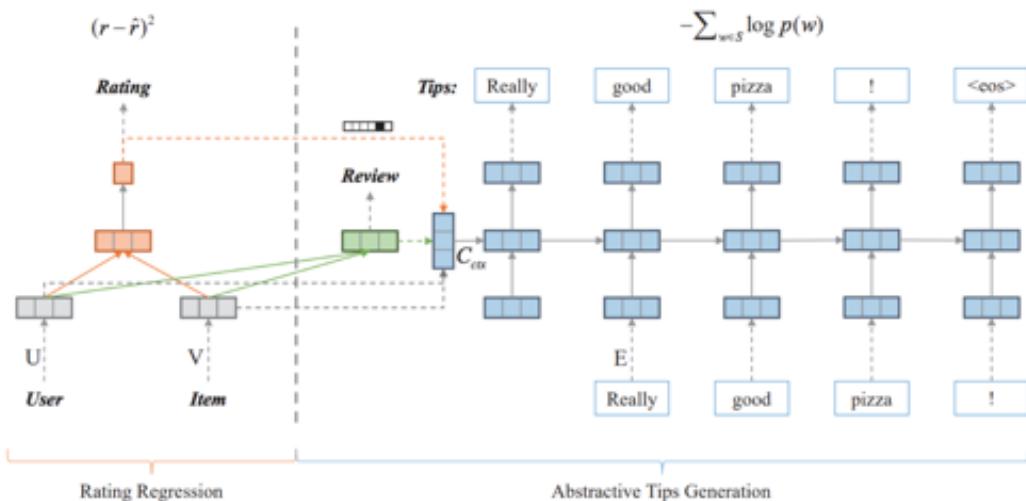


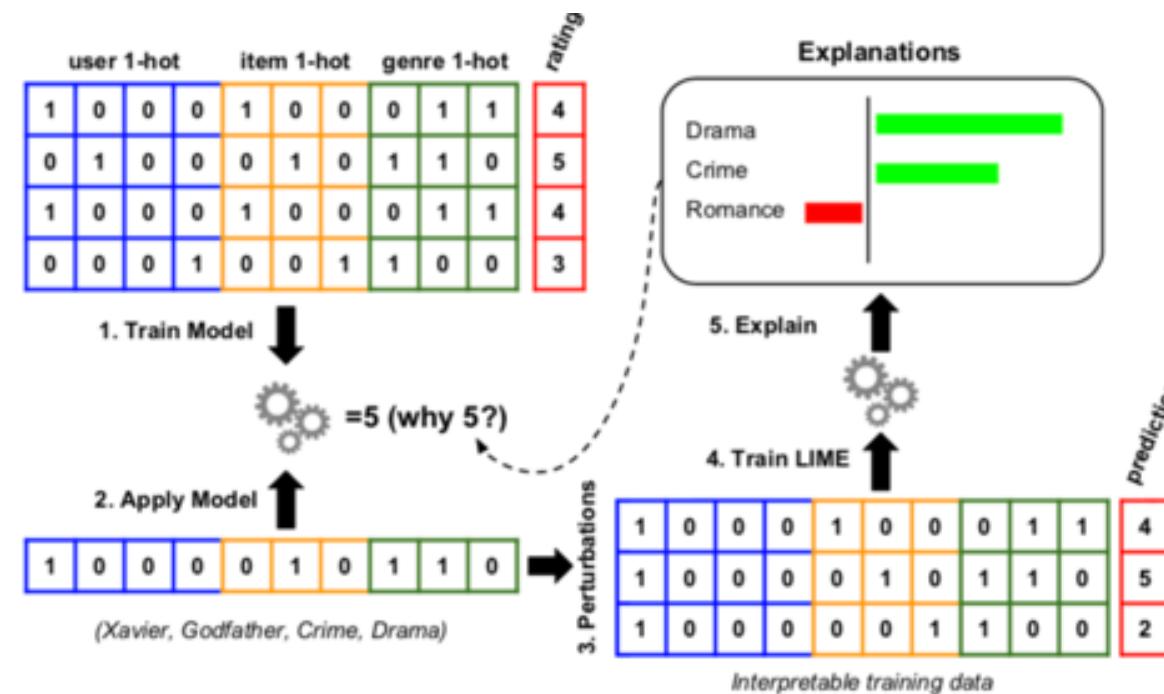
Figure 2: Our proposed framework NRT for rating regression and abstractive tips generation.

**Table 10:** Examples of the predicted ratings and the generated tips. The first line of each group shows the generated rating and tips. The second line shows the ground truth.

Rating	Tips
<b>4.64</b>	<b><i>This is a great product for a great price.</i></b>
5	Great product at a great price.
<b>4.87</b>	<b><i>I purchased this as a replacement and it is a perfect fit and the sound is excellent.</i></b>
5	Amazing sound.
<b>4.69</b>	<b><i>I have been using these for a couple of months.</i></b>
4	Plenty of wire gets signals and power to my amp just fine quality wise.
<b>4.87</b>	<b><i>One of my favorite movies.</i></b>
5	This is a movie that is not to be missed.
<b>4.07</b>	<b><i>Why do people hate this film.</i></b>
4	Universal why didnt your company release this edition in 1999.

# Adapting current XAI approaches to RecSys

- Adapting LIME to recommendation: Nóbrega, C., & Marinho, L. (2019). Towards explaining recommendations through local surrogate models. ACM/SIGAPP Symposium on Applied Computing.



# This survey is not exhaustive

- I strongly recommend visiting

<https://sites.google.com/view/ears-tutorial/>



[Yongfeng Zhang](#)

Assistant Professor

Rutgers University, New  
Brunswick, NJ, USA



[Jiaxin Mao](#)

Postdoc

Tsinghua University, Beijing  
China



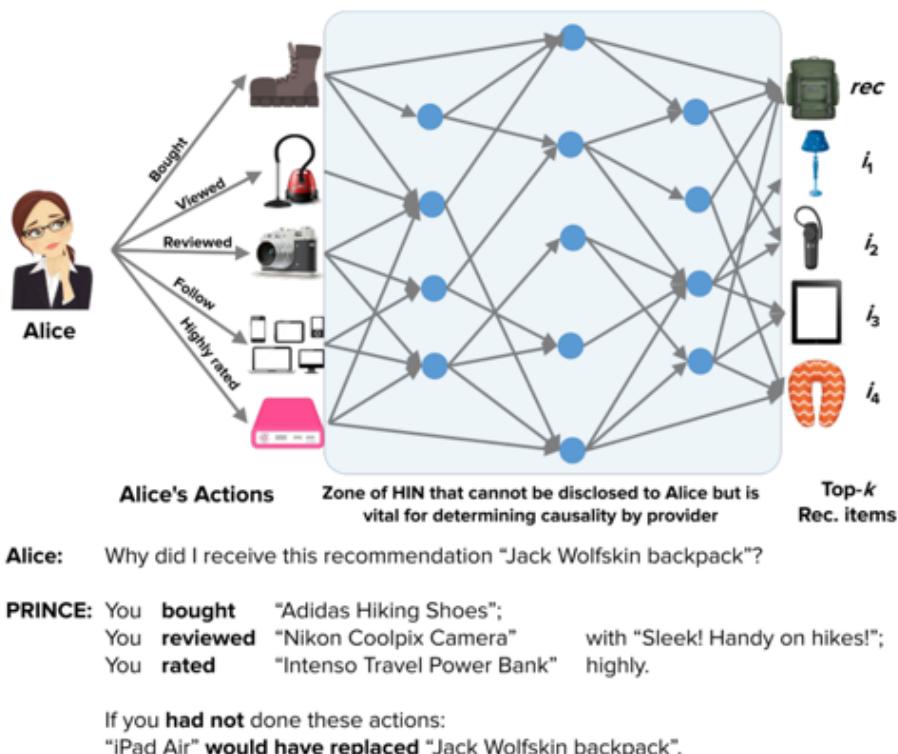
[Qingyao Ai](#)

Assistant Professor

University of Utah, Salt Lake  
City, UT, USA

# PRINCE – graph based and counterfactuals

- Ghazimatin, A., Balalau, O., Saha Roy, R., & Weikum, G. (2020). PRINCE: Provider-side Interpretability with Counterfactual Explanations in Recommender Systems. WSDM 2020.



# Araujo et al. – Attention on MOBA RecSys

- Villa, A., Araujo, V., Cattan, F. and Parra, D. "Interpretable Contextual Team-aware Item Recommendation: Application in Multiplayer Online Battle Arena Games." RecSys 2020.

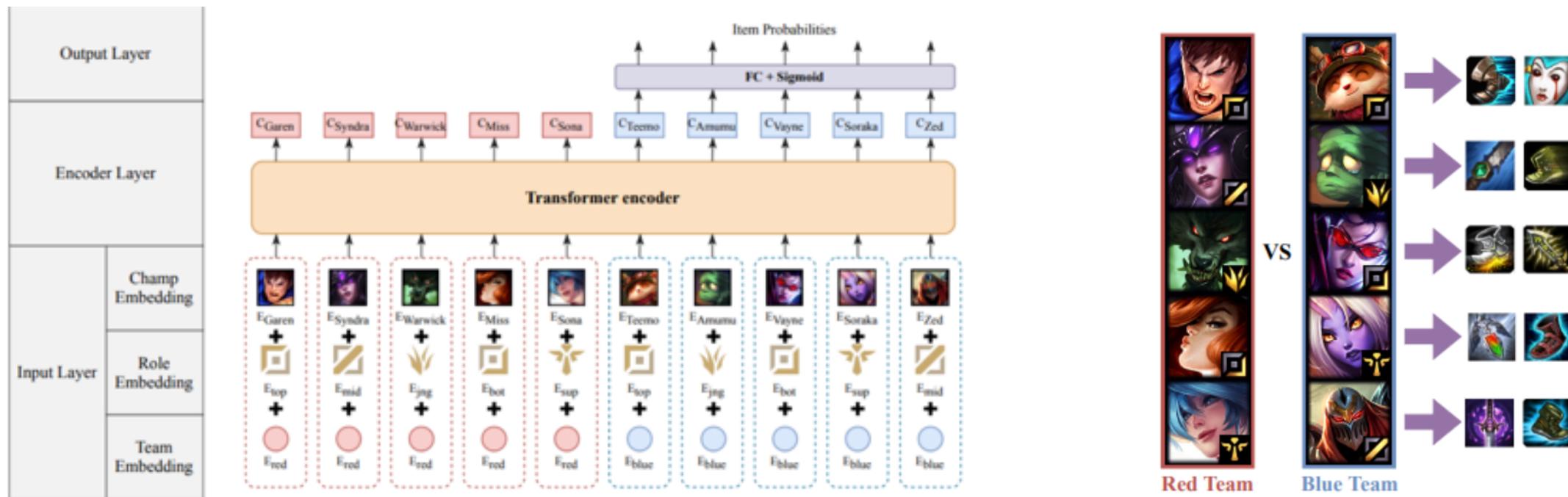
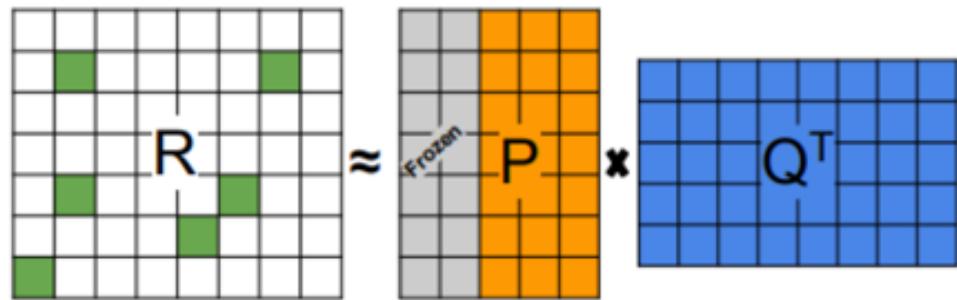


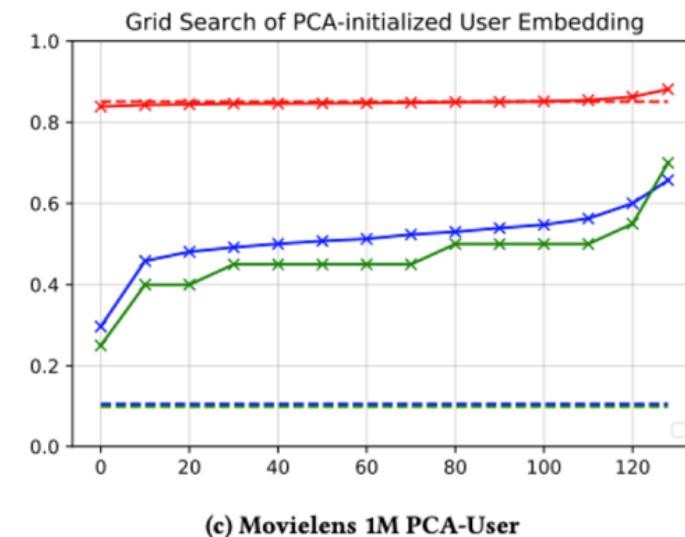
Fig. 2. Network architecture of TTIR.

# Similarity-based explanations

- Marinho, L. B., da Costa, J. B. G., Parra, D., & Santos, R. L. (2022). Similarity-Based Explanations meet Matrix Factorization via Structure-Preserving Embedding. IUI 2022.



**Figure 1: MF with the first two dimensions of  $P = PCA(R, d)$  frozen. By freezing the first factors of  $P$ , i.e., the factors that carry most of the data structure, we enable similarity-based explanations based on users that share similar preferences in the input data.**



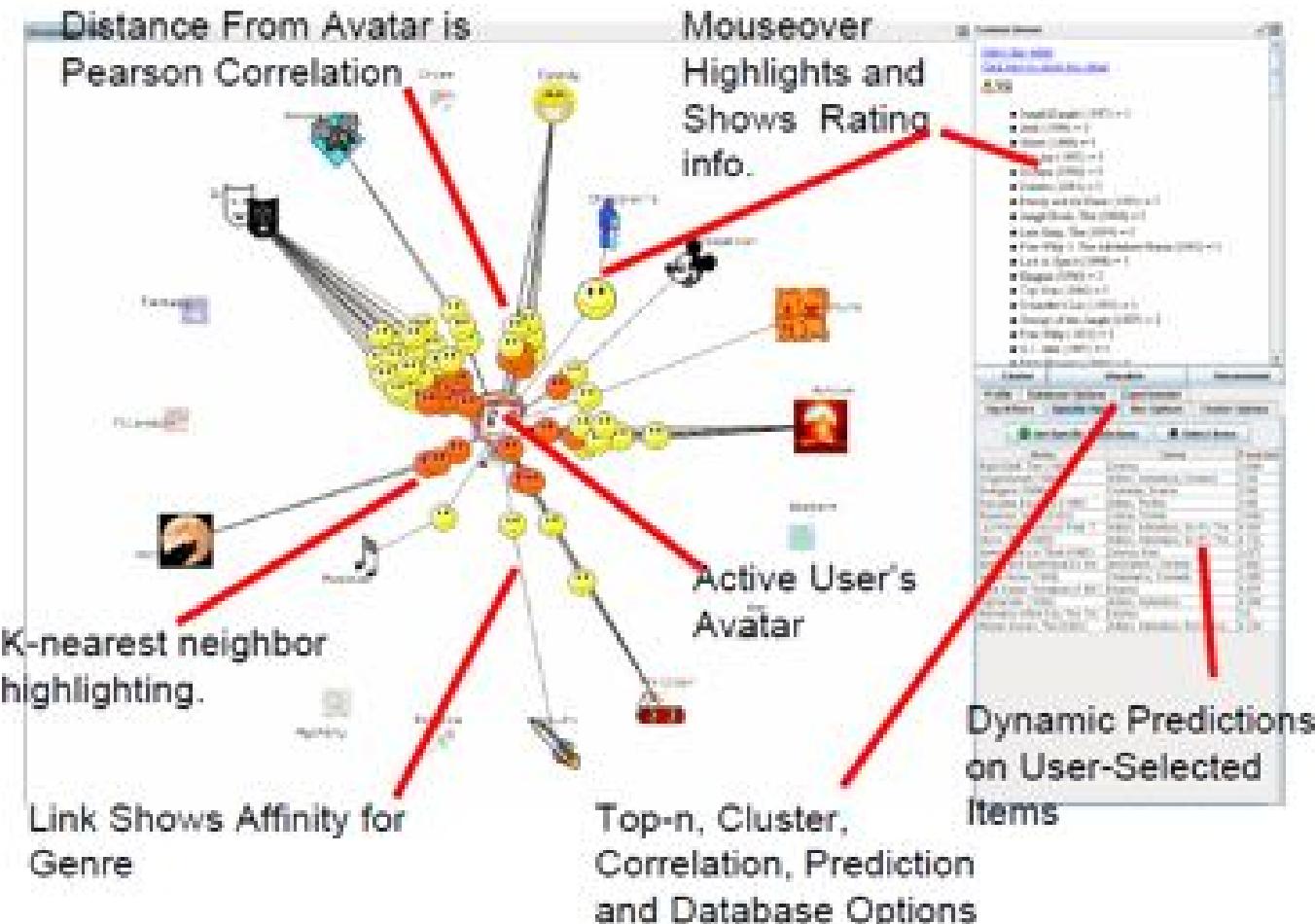
**Figure 3: Grid search for the best trade-off between explainability and RMSE. Horizontal dashed lines denote: RMSE (red), Mantel (blue), and neighborhood overlap (green) w.r.t. biased MF.**

# The Role of Interactive Visualization in RecSys XAI

- PeerChooser (O'Donovan et al, 2008)
- SmallWorlds (Gretarsson et al, 2010)
- To each his own (Knijnenburg et al, 2010)
- TasteWeights (Bostandjev et al. 2012, Knijnenburg et al. 2012)
- TalkExplorer/Aduna (Verbert et al. 2013)
- SetFusion (Parra et al., 2014)
- Moodplay (Andjelkovic et al., 2016)
- 3D Inspector (Loepp et al, 2017)
- Tuner+ (Tsai et al, 2019)
- Explain or Not (Millecamp et al, 2019)

# Peerchooser

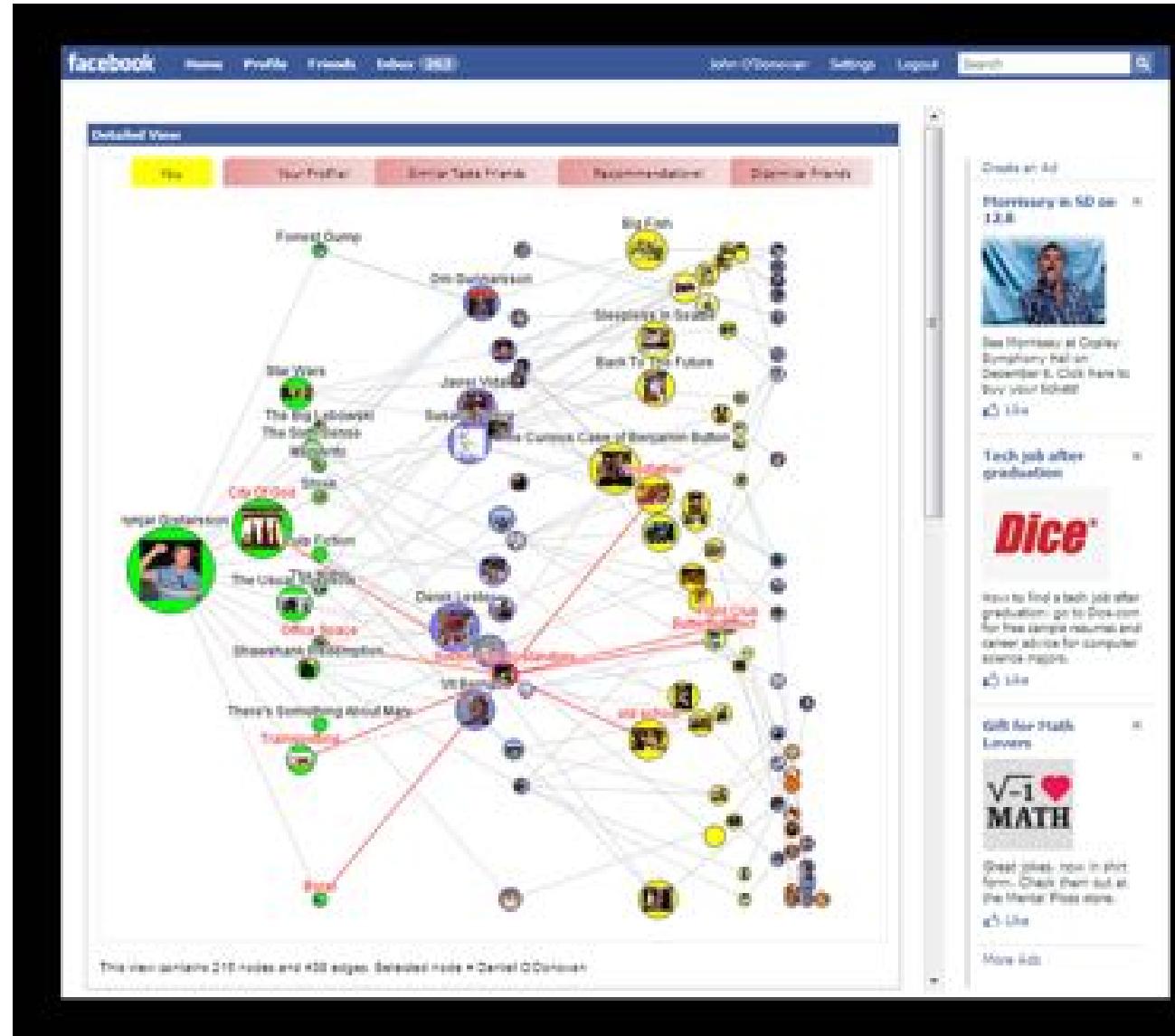
- PeerChooser  
(CHI 2008)
- John O'Donovan,  
Barry Smyth,  
Brynjar  
Gretarsson,  
Svetlin  
Bostandjiev,  
Tobias Hollerer



2: Annotated Screenshot of PeerChooser's Interactive Interface.

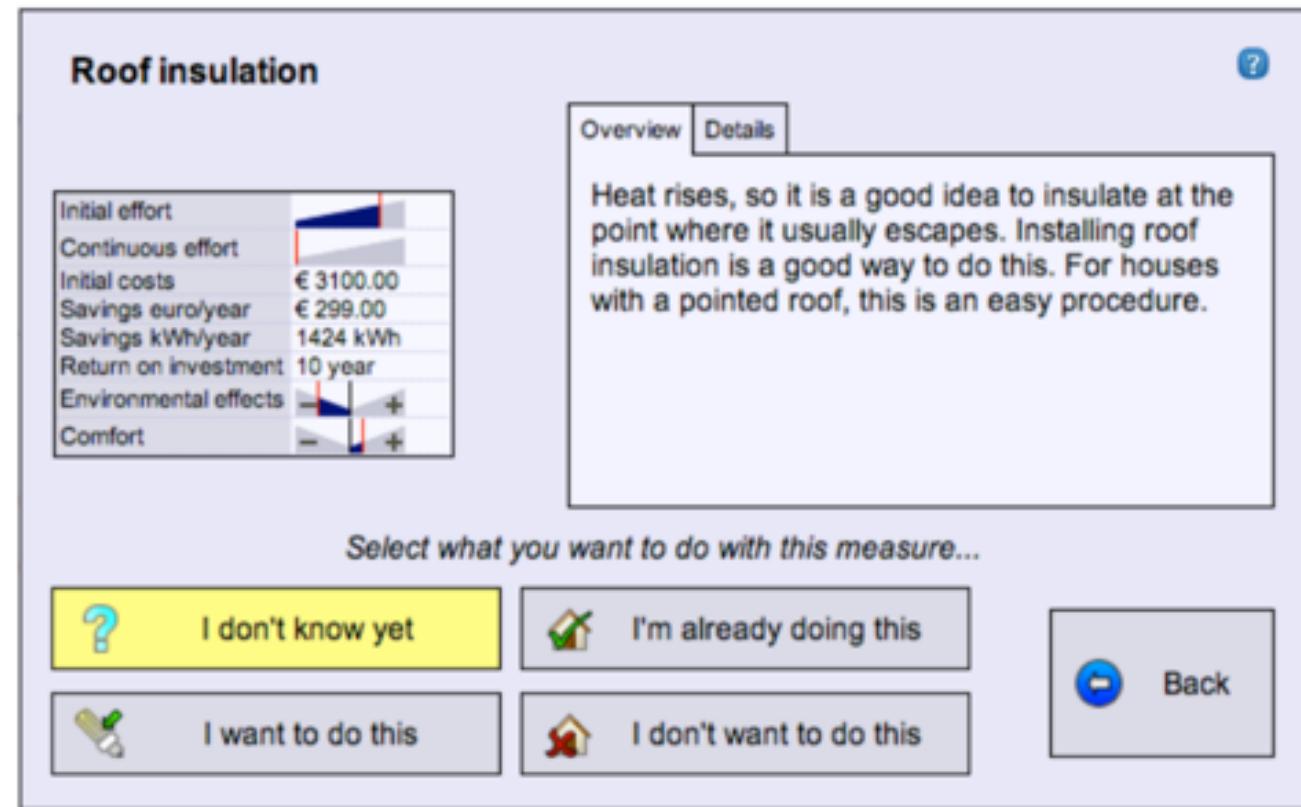
# Smallworlds

- SmallWorlds:  
Visualizing Social  
Recommendations  
(IEEE-TVCG 2010)
- Brynjar Gretarsson,  
John O'Donovan ,  
Svetlin Bostandjiev,  
Christopher Hall,  
Tobias Höllerer
- User study with 17  
users



# Each to his own

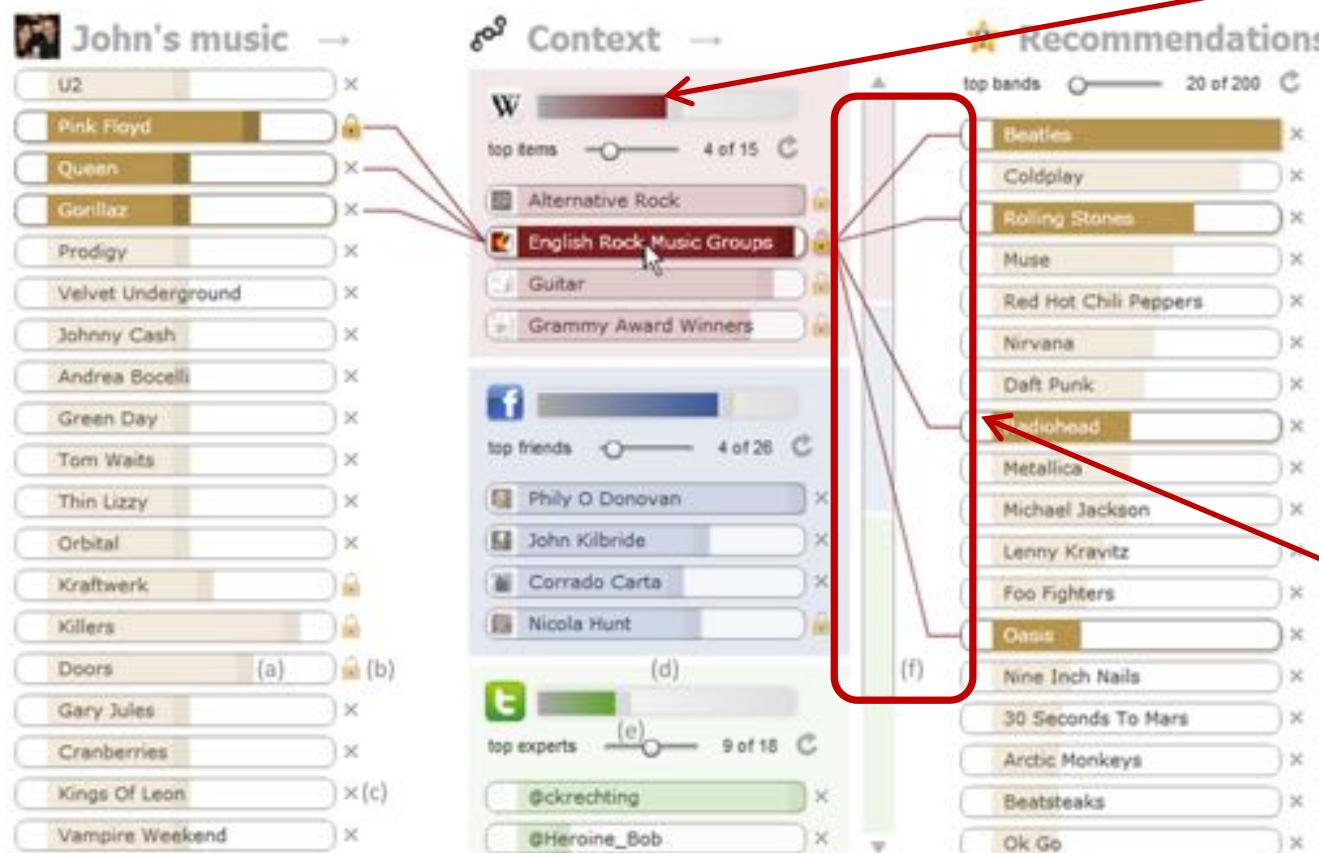
- Each to his own:  
how different users  
call for different  
interaction methods  
in recommender  
systems (RecSys  
2010)
- Knijnenburg, B. P.,  
Reijmer, N. J., &  
Willemsen, M. C.
- UX with Interface  
details depends on  
previous user  
knowledge of the  
topic



**Figure 2. Screen shown to users when they click on an item**

# TasteWeights

- TasteWeights: a visual interactive hybrid recommender system (RecSys 2012)
- Bostandjiev, S., O'Donovan, J., & Höllerer, T.



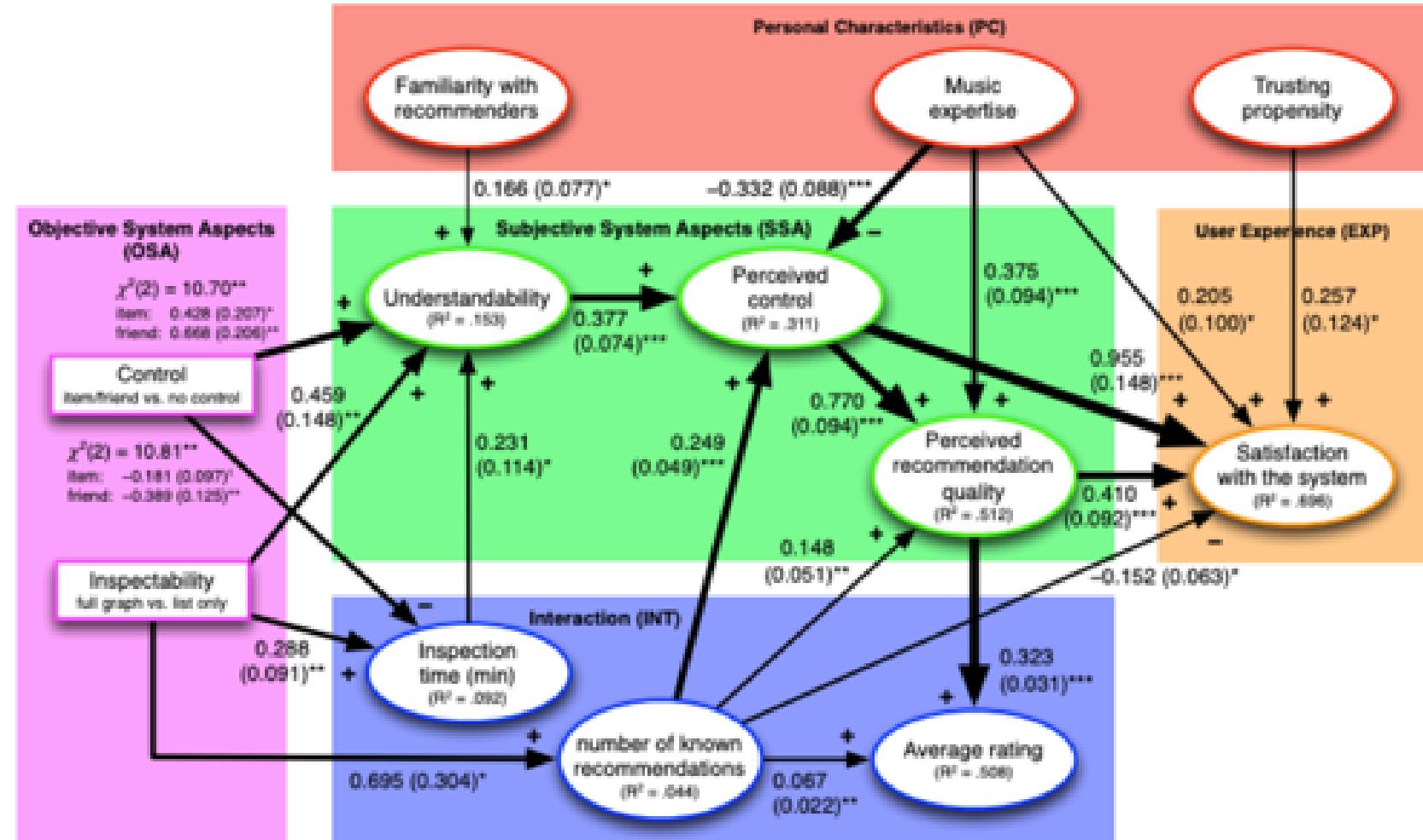
**Controllability:** Sliders that let users control the importance of preferences and contexts

**Inspectability:** lines that connect recommended items with contexts and user preferences

Also : Knijnenburg, B. P., Bostandjiev, S., O'Donovan, J., & Kobsa, A. (2012). Inspectability and control in social recommenders. In Proceedings of ACM RecSys.

# TasteWeights II

- TasteWeights: a visual interactive hybrid recommender system (RecSys 2012)
- Bostandjiev, S., O'Donovan, J., & Höllerer, T.



Also : Knijnenburg, B. P., Bostandjiev, S., O'Donovan, J., & Kobsa, A. (2012). Inspectability and control in social recommenders. In Proceedings of ACM RecSys.

# TalkExplorer

- Visualizing recommendations to support exploration, transparency and controllability
- Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. (IUI 2013)

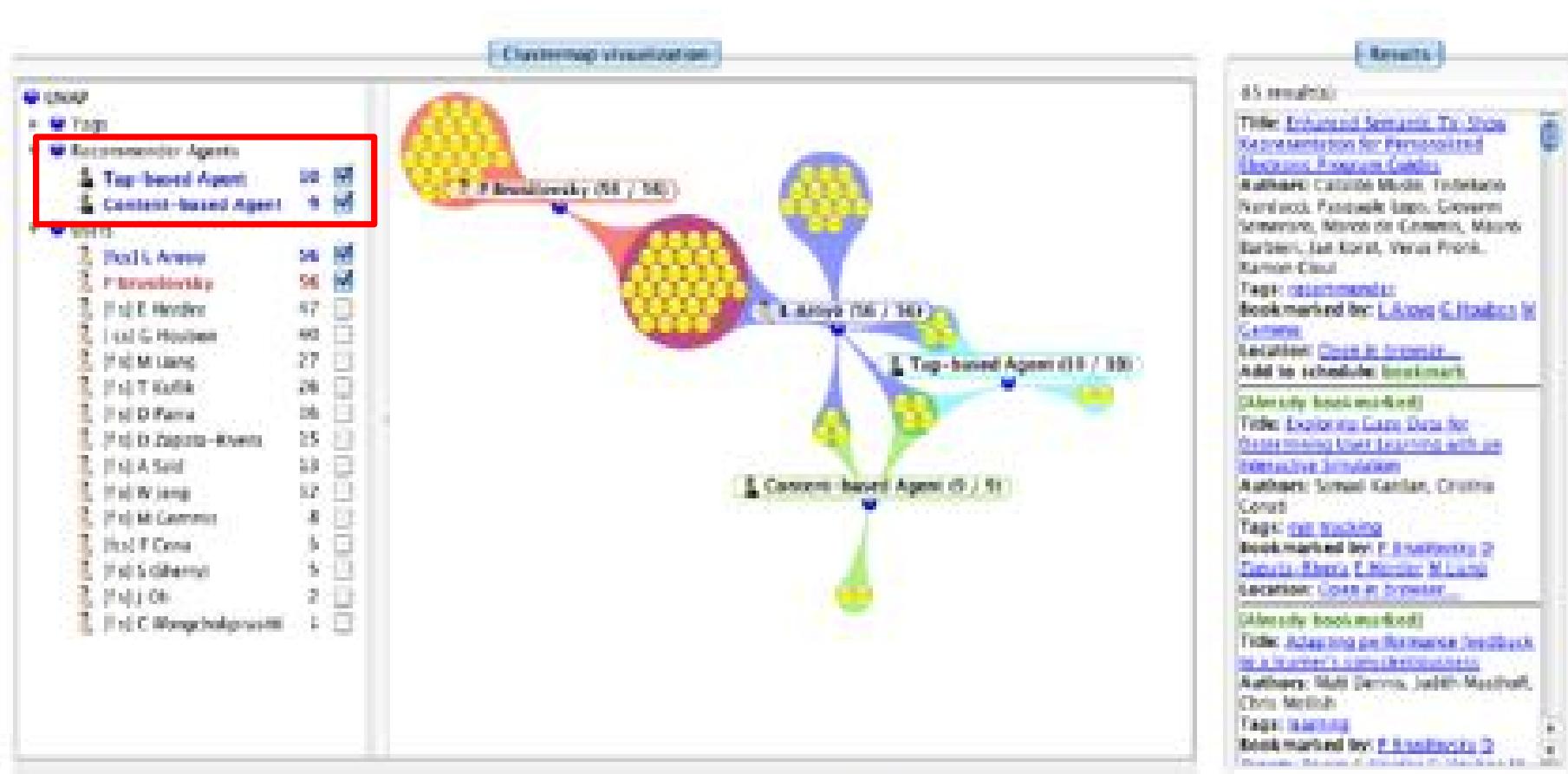


Figure 2: TalkExplorer

# SetFusion

- See what you want to see: visual user-driven approach for hybrid recommendation
- Denis Parra, Peter Brusilovsky, and Christoph Trattner (IUI 2014)



## SetFusion: A Controllable Hybrid Recommender

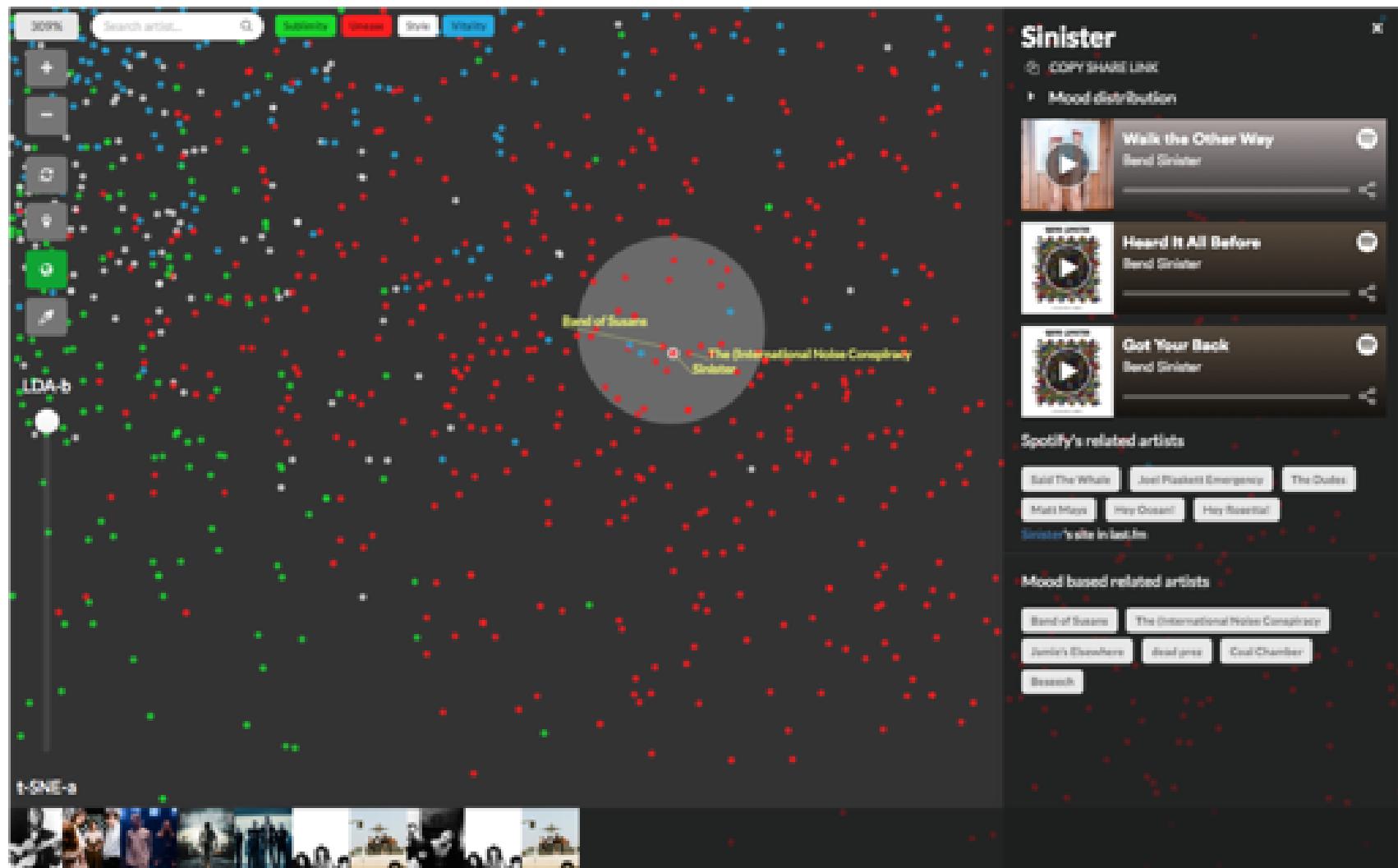
Parra, D., Brusilovsky, P., Trattner, C.

IUI 2014, Haifa, Israel

# Moodplay

- Moodplay:  
Interactive Mood-based Music Discovery and Recommendation.

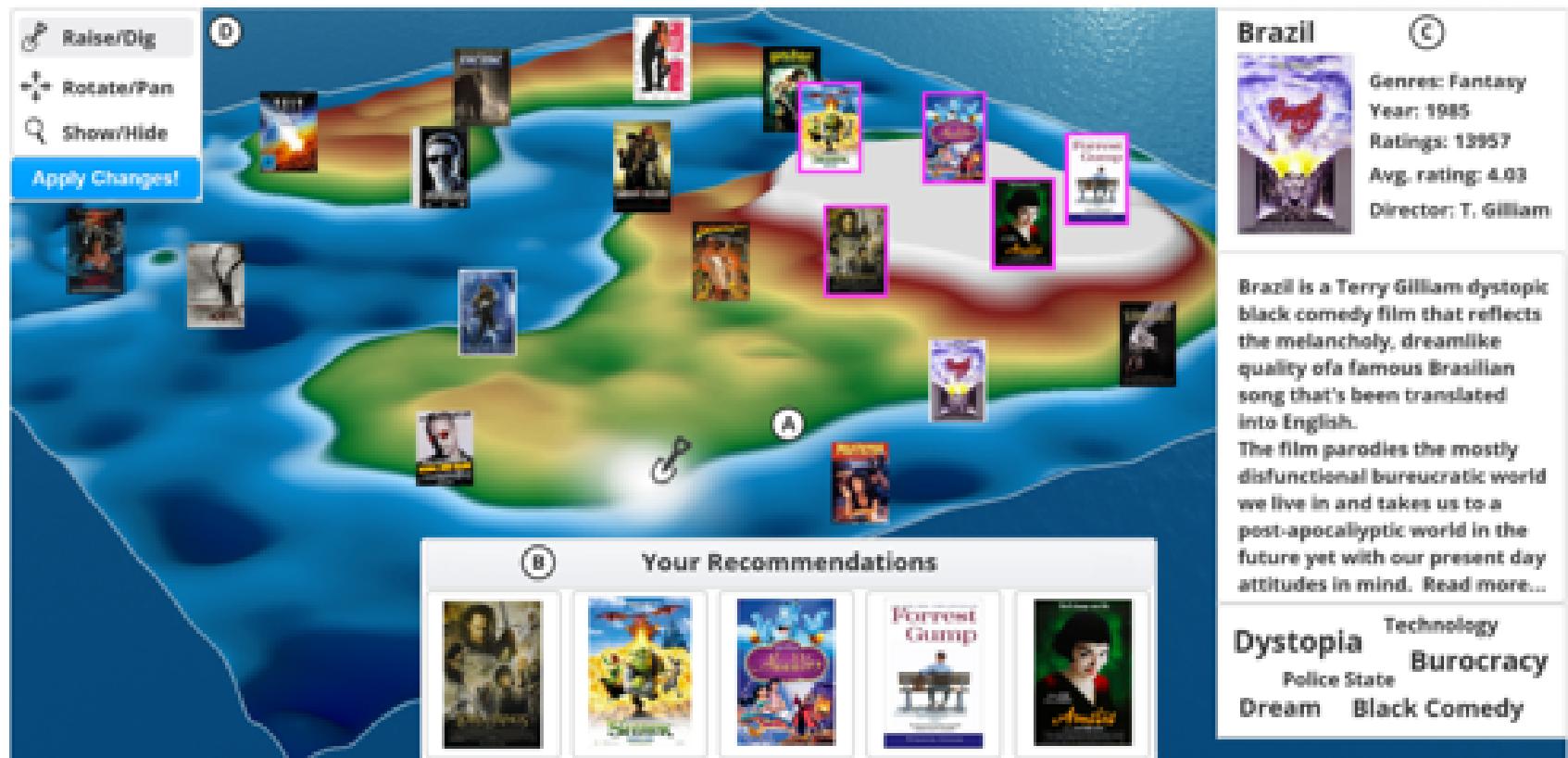
- Ivana Andjelkovic, Denis Parra, and John O'Donovan. (UMAP 2016; IJHCS 2019)



<http://moodplay.pythonanywhere.com/>

# 3D Landscape

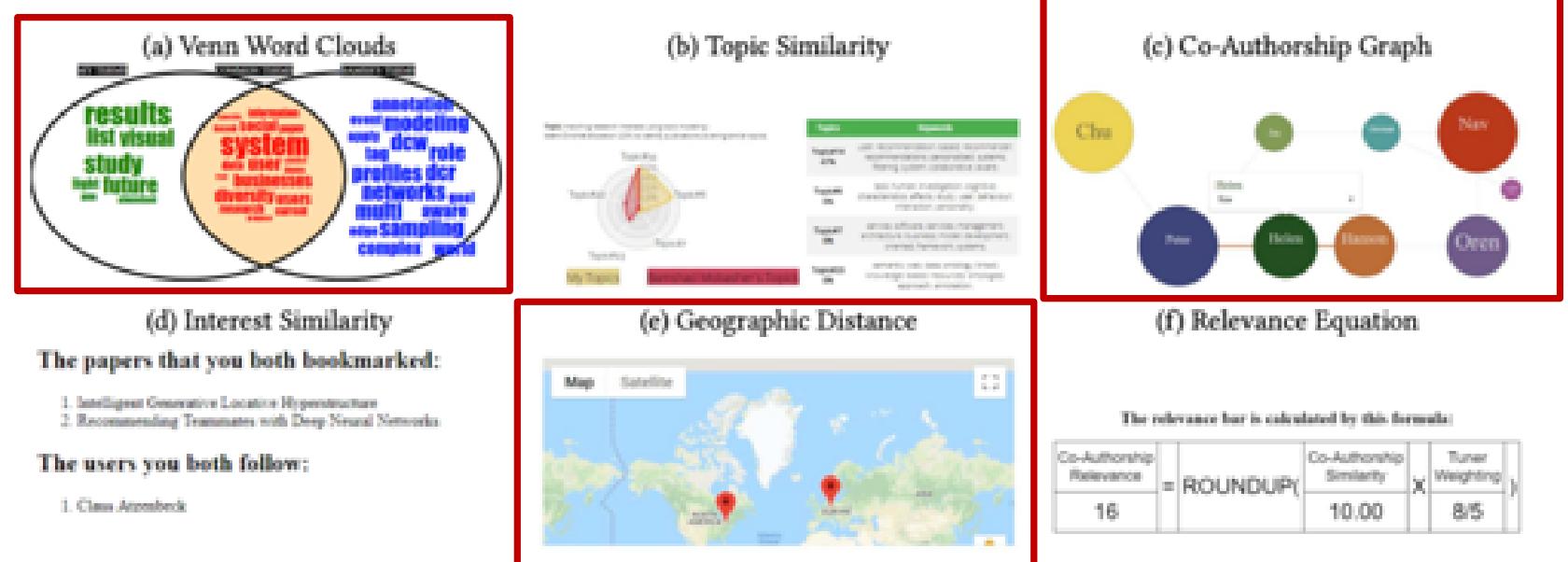
- A 3D item space visualization for presenting and manipulating user preferences in collaborative filtering
  - Kunkel, J., Loepp, B., & Ziegler, J. (IUI 2017)



# Tuner+

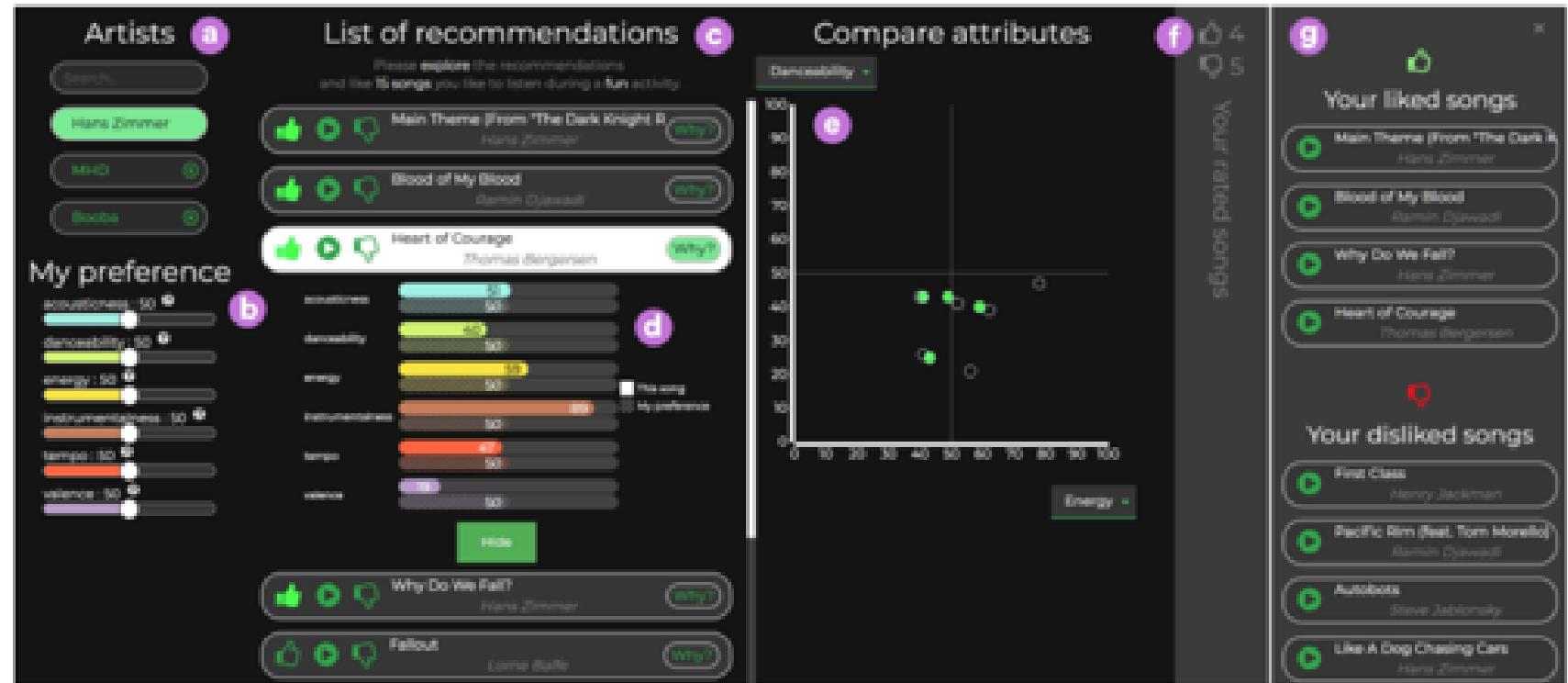
- Explaining recommendations in an interactive hybrid social recommender
- Tsai, C. H., & Brusilovsky, P. (IUI 2019)

Publication Similarity		Topic Similarity		Co-Authorship Similarity		CN3 Interest Similarity		Geographic Distance	
A									
Profile	Relevance	C	Name	D	Follow	Connect	Affiliation	Position	
	10 8 6 10	(i) Ben	Following	Waiting confirmation	University	Professor			
B	5 10 6 6	(i) Lou	Follow	Add as connection	University	Student			



# To Explain or Not

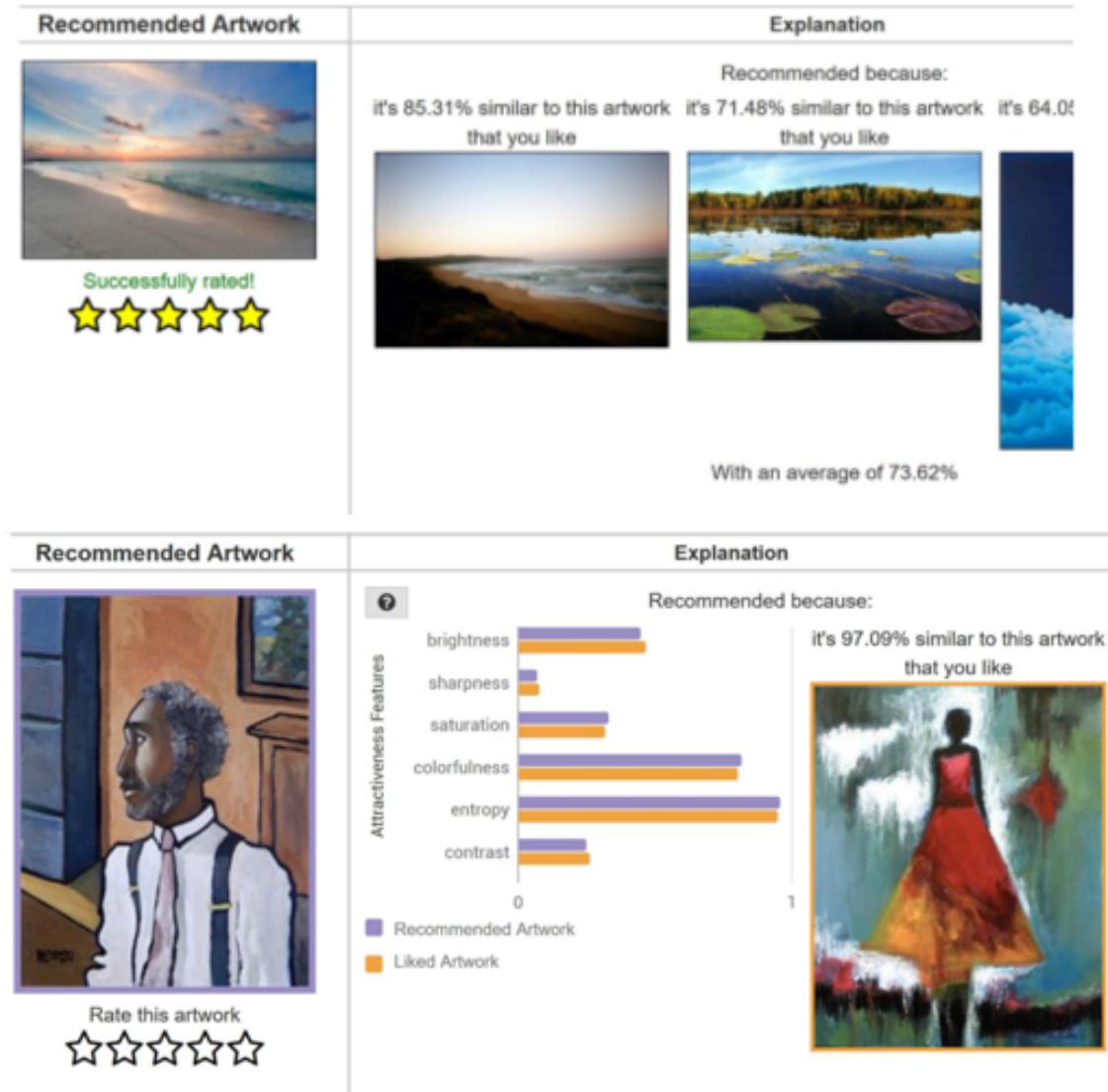
- To explain or not to explain: the effects of personal characteristics when explaining music recommendations
- Millecamp, M., Htun, N. N., Conati, C., & Verbert, K. (IUI 2019)



\* Locus of control, need for cognition, musical sophistication, visual literacy.

# “Explanations” only work with an accurate RecSys

- The Effect of Explanations and Algorithmic Accuracy on Visual Recommender Systems of Artistic Images (IUI2019)
- Dominguez et al.



# RecSys 2021- TTIR LoL item RecSys

- Interpretable Contextual Team-aware Item Recommendation: Application in Multiplayer Online Battle Arena Games (RecSys2020)
- Villa, Araujo, Cattan, Parra
- Explanation Usefulness: mostly for newer players



Question	Global M±SD (N=16)	Subjects by year of first play		
		2009-11 (N=5)	2012-14 (N=5)	2015-2017 (N=6)
Q1. How good were the recommendations for the <i>Blue team</i> ?	$7.98 \pm 1.22$	$7.7 \pm 1.24$	$7.7 \pm 1.16$	$8.46 \pm 1.3$
Q2. Is it understandable the influence of every team member upon each champion being recommended ?	$7.44 \pm 1.72$	$7.4 \pm 1.55$	$7.1 \pm 0.8$	$7.75 \pm 2.49$
Q3. Is it useful the information provided by the visualization in order to understand the item recommendations made ?	$6.9 \pm 2.15$	$6.7 \pm 1.98$	$6.6 \pm 1.65$	$7.33 \pm 2.87$

# What have we learned from this research?

# What have we learned from this research?

- **Perception of control** is key, but how much depends on knowledge/literacy level of the users on the domain

# What have we learned from this research?

- Perception of control is key, but how much depends on knowledge/literacy level of the users on the domain
- We should let users choose whether seeing explanations or not

# What have we learned from this research?

- Perception of control is key, but how much depends on knowledge/literacy level of the users on the domain
- We should let users choose whether seeing explanations or not
- Explanations should present different levels of details, we should give users' control to explore them

*Overview first, details on demand*  
– Schneiderman mantra

# What have we learned from this research? II

- People **might prefer a less transparent recommender** if they are able to predict what the system is going to suggest

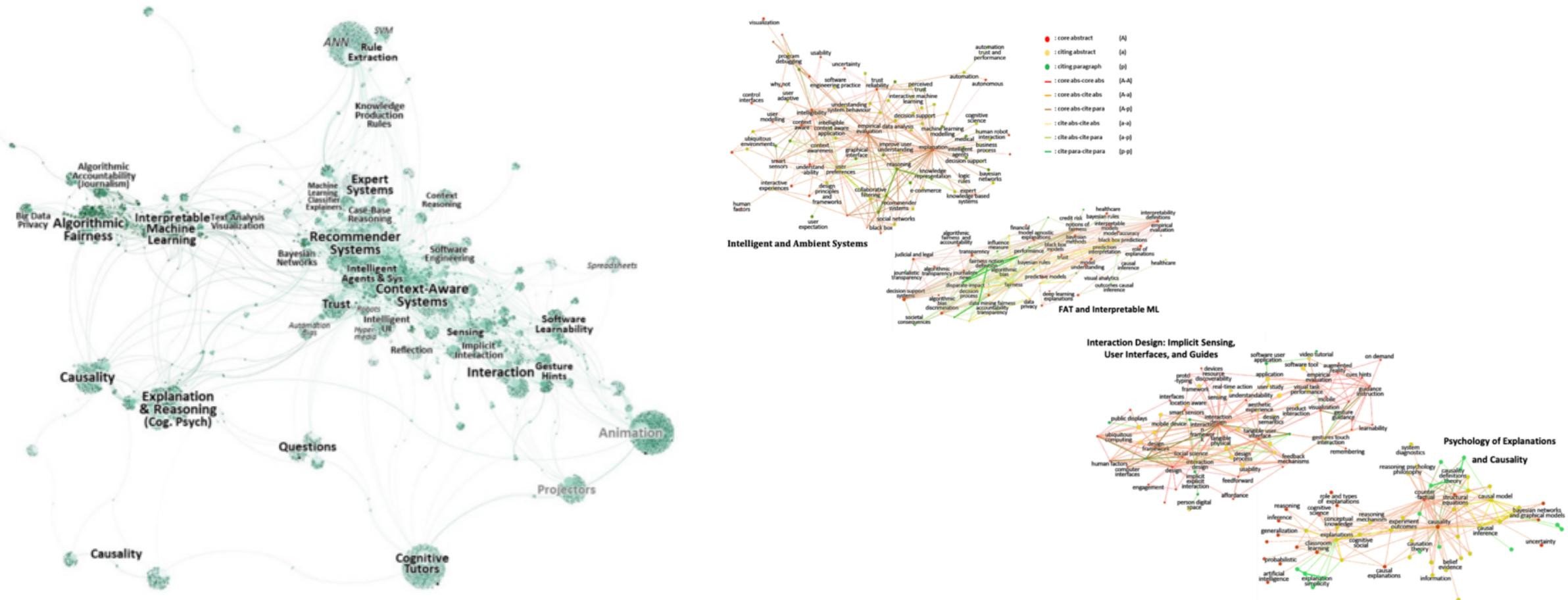
# What have we learned from this research? II

- People might prefer a less transparent recommender if they are able to predict what the system is going to suggest
- The **explanation** must provide something **beyond the obvious** (we need further research on this)

# What have we learned from this research? II

- People might prefer a less transparent recommender if they are able to predict what the system is going to suggest
- The explanation must provide something beyond the obvious! (we need further research on this)
- The **right visual encoding** can promote curiosity and exploratory behavior

# Need for further interaction between AI/HCI communities



Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). **Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda.** CHI 2018.

# Back to Visual XAI

- How do we analyze and design visualizations for XAI?
- There are frameworks and guidelines for designing visualizations (e.g. Munzer VAD framework)
- ... but there is no framework yet for designing Visual XAI

# Work in Progress: VD4XAI

## XAI Space

Threat: Wrong problem

Validate: Observe and interview end user

(NEW) Threat: Wrong explanation type

Validate: Justify explanation with literature or example and interview end user

(NEW) Threat: Wrong XAI method

Validate: Justify method with literature or example and interview end users

(NEW) Threat: Wrong visual justification

Validation: Analyze data availability

Nested model Tamara Munzer (3 last steps)

## Design Space

Nested model Tamara Munzer (3 last steps)

Threat: Wrong task/data abstraction

Threat: Ineffective encoding/interaction idiom

Validate: Justify encoding/interaction design

Threat: Slow algorithm

Validate: Analyze computational complexity

Implement system

Validate: Measure system time/memory

Validate: Qualitative/quantitative result image analysis. Test on any users, informal usability study

Validate: Lab study, measure human time/errrors for task

Validate: Test on target users, include anecdotal evidence of visual explanation utility about model response

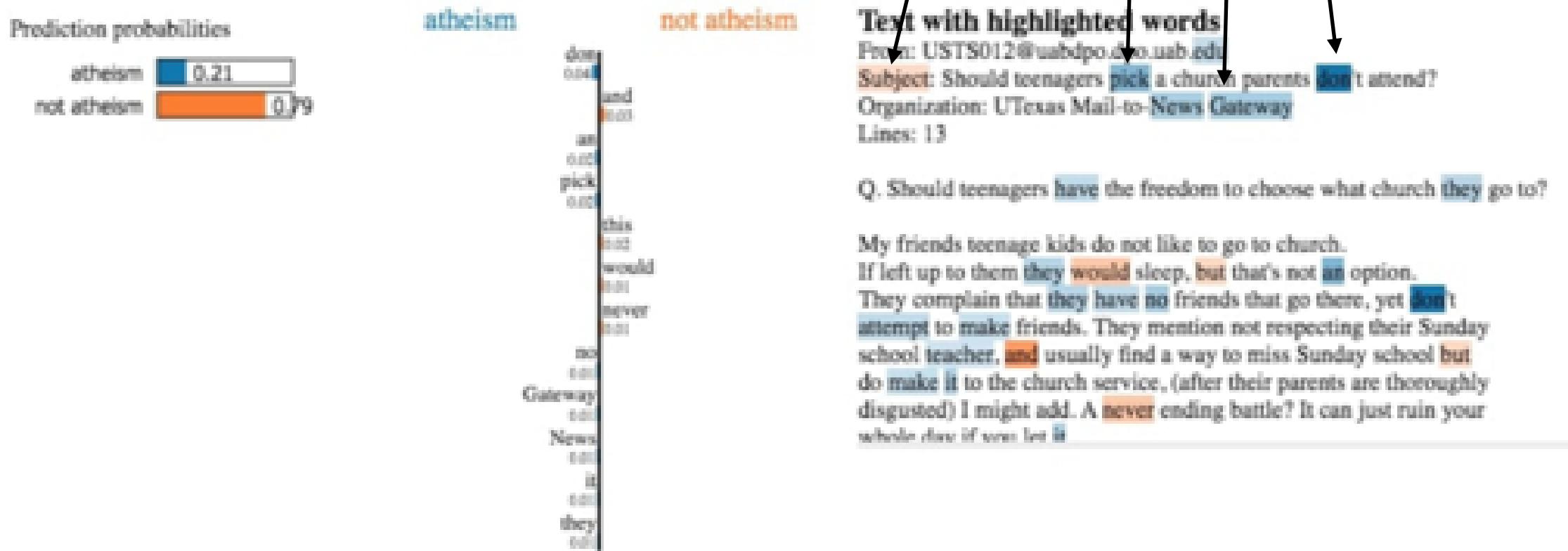
Validate: Field study, document human usage of deployed system

Validate: Observe adoption rates

H Valdivieso, D Parra, I Donoso, K Verbert, et al.

**VD4XAI: Connecting the Visualization Design Space with the Explainable AI Task Space.** Work in progress

# Analyze LIME: why colored background ?



# Analyze LIME: Marks, channels and Interaction

## Text with highlighted words

From: USTS012@uabdpo.dpo.uab.edu

Subject: Should teenagers pick a church parents don't attend?

Organization: UTexas Mail-to-News Gateway

Lines: 13

Q. Should teenagers have the freedom to choose what church they go to?

My friends teenage kids do not like to go to church.

If left up to them they would sleep, but that's not an option.

They complain that they have no friends that go there, yet don't attempt to make friends. They mention not respecting their Sunday school teacher, and usually find a way to miss Sunday school but do make it to the church service, (after their parents are thoroughly disgusted) I might add. A never ending battle? It can just ruin your whole day if you let it

## Font property - Luminance

Title: a meta analysis of birth origin effects on reproduction in diverse captive environments

Abstract: successfully establishing captive breeding programs is priority across diverse industries to address food security demand for ethical laboratory research animals and prevent extinction differences in reproductive success due to birth origin may threaten the long term sustainability of captive breeding our meta analysis examining effect sizes from species of invertebrates fish birds and mammals shows that overall captive born animals have decreased odds of reproductive success in captivity compared to their wild born counterparts the largest effects are seen in commercial aquaculture relative to conservation or laboratory settings and offspring survival and offspring quality were the most sensitive traits although somewhat weaker trend reproductive success in conservation and laboratory research breeding programs is also in negative direction for captive born animals our study provides the foundation for future investigation of non genetic and genetic drivers of change

## Additional Mark - Bars Length (A)

Title: a meta analysis of birth origin effects on reproduction in diverse captive environments

Abstract: successfully establishing captive breeding programs is priority across diverse industries to address food security demand for ethical laboratory research animals and prevent extinction differences in reproductive success due to birth origin may threaten the long term sustainability of captive breeding our meta analysis examining effect sizes from species of invertebrates fish birds and mammals shows that overall captive born animals have

# Recent study: Biomedical Document Classification

The screenshot shows a web-based application for biomedical document classification. At the top, the URL is <https://app.loverevidence.com/screening/60ca02ece689d3e0147900c3>. The interface includes a search bar and navigation buttons for 'Screening' and 'Statistics'.

**(A)** Predicted Label Confidence: Randomized Trial (43.4% out of 100%).

**(B)** Show tutorial.

**(C)** Document abstract: Fluctuations and the rate-limiting step of peptide-induced membrane leakage. The text discusses peptide-induced vesicle leakage kinetics and proposes a stochastic model for the interaction between peptides and membranes.

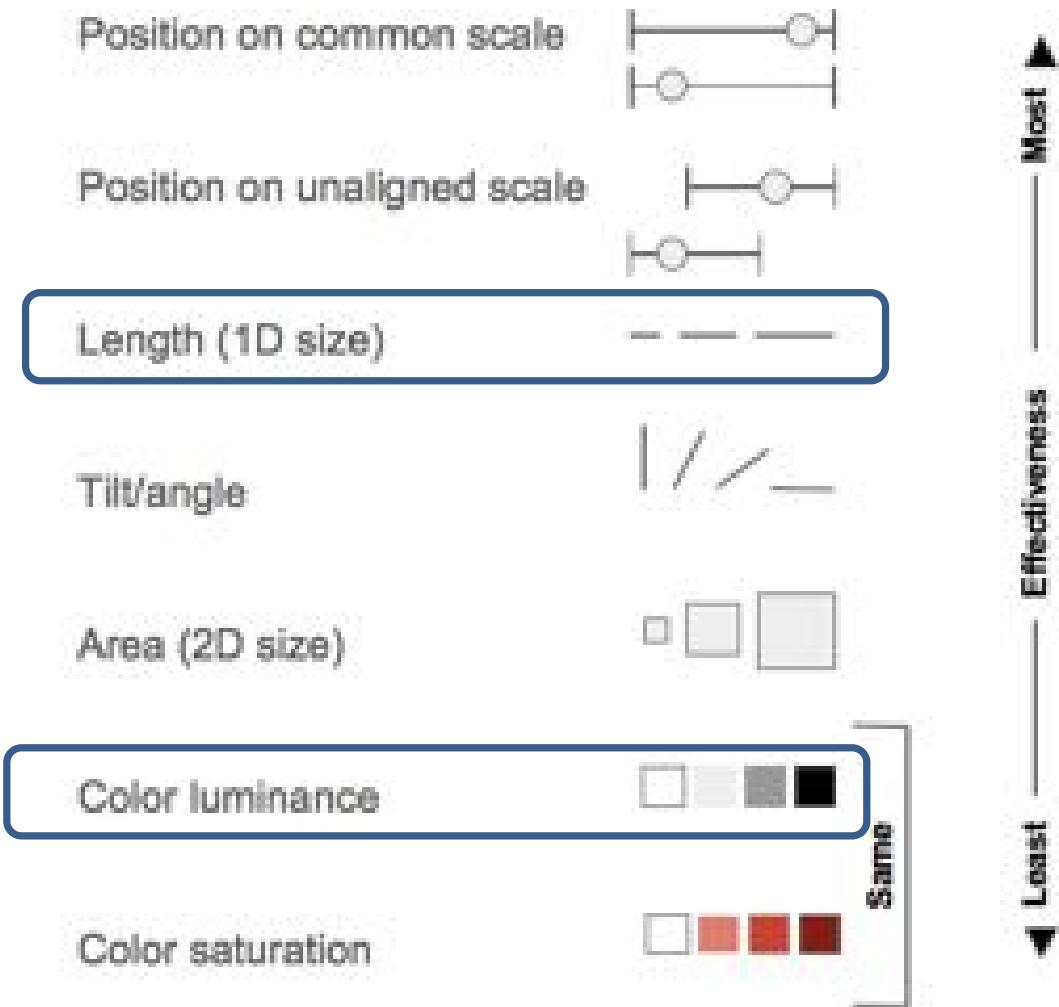
**(D)** Document classification options: Broad Synthesis, Excluded, Non-randomized Studies, Systematic Review.

**(E)** User study section: 'Is this document a randomized trial (outcome)?' with a 'Yes' button. Below it, questions about agreement with statements and predicted label confidence (Neutral) are shown. A blue box highlights the 'The highlighted words in abstract' section, which also shows a 'Neutral' rating.

**(F)** Epistemmonikos User Study Phase 1 of 2: Fixed Visualizations. Total progress: 7/40. Brightness Link to challenge.

... and tell us how useful were the highlighted words

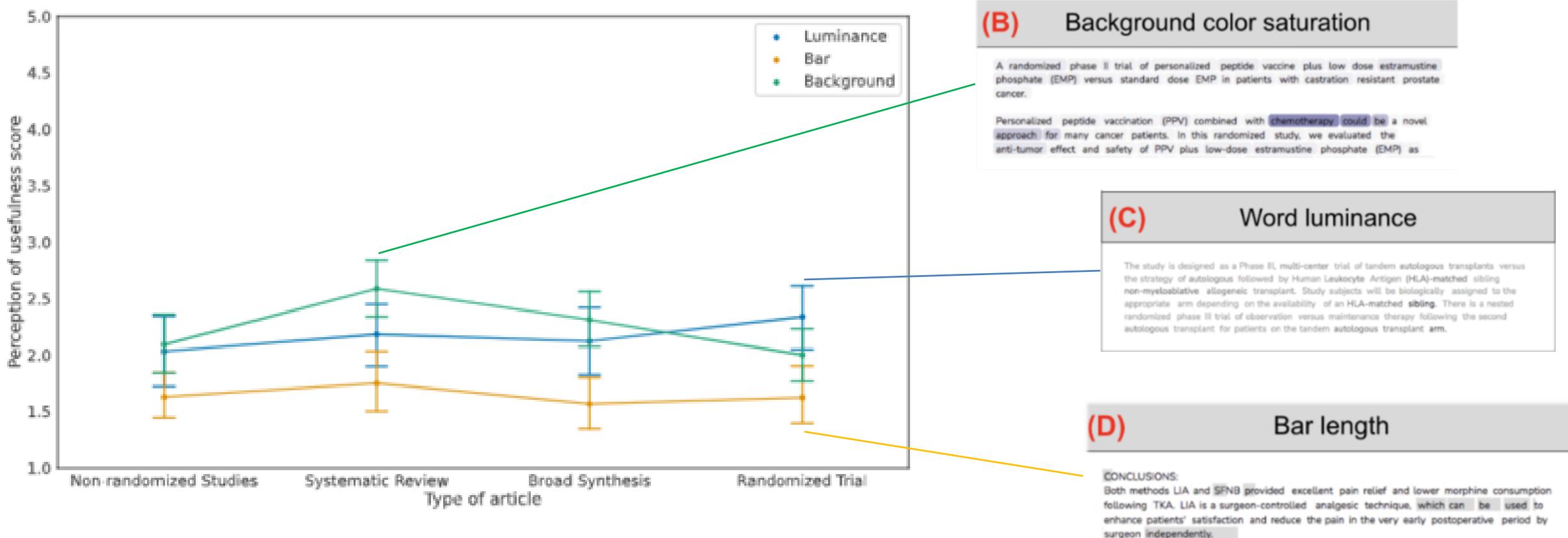
# Designs: marks and channels (effectiveness principle)



Do people prefer  
more effective channels  
for neural attention-based  
explanations?

# Recent study: Document Classification

- \* In this task, a more visually effective channel is not the most useful
- \* Perception of usefulness depends on document type



# Conclusion

- XAI and Visual XAI are active topics of critical research in AI
- Despite XAI being coined around 2017, the RecSys-IUI community has been researching explainability and transparency since early 2000s, and there are important lessons to contribute to XAI
- There's still a need to connect different disciplines and areas to address challenges in XAI and Visual XAI

dparra@ing.puc.cl

**THANKS!**

