From Transparency and Users' Control in Recommender System Interfaces to Guidelines for Visual XAI

Denis Parra
PUC Chile & CENIA, iHealth, IMFD

APEx-UI Workshop co-located at ACM IUI
Virtually hosted by U. Of Helsinki, Finland
March 21st, 2022

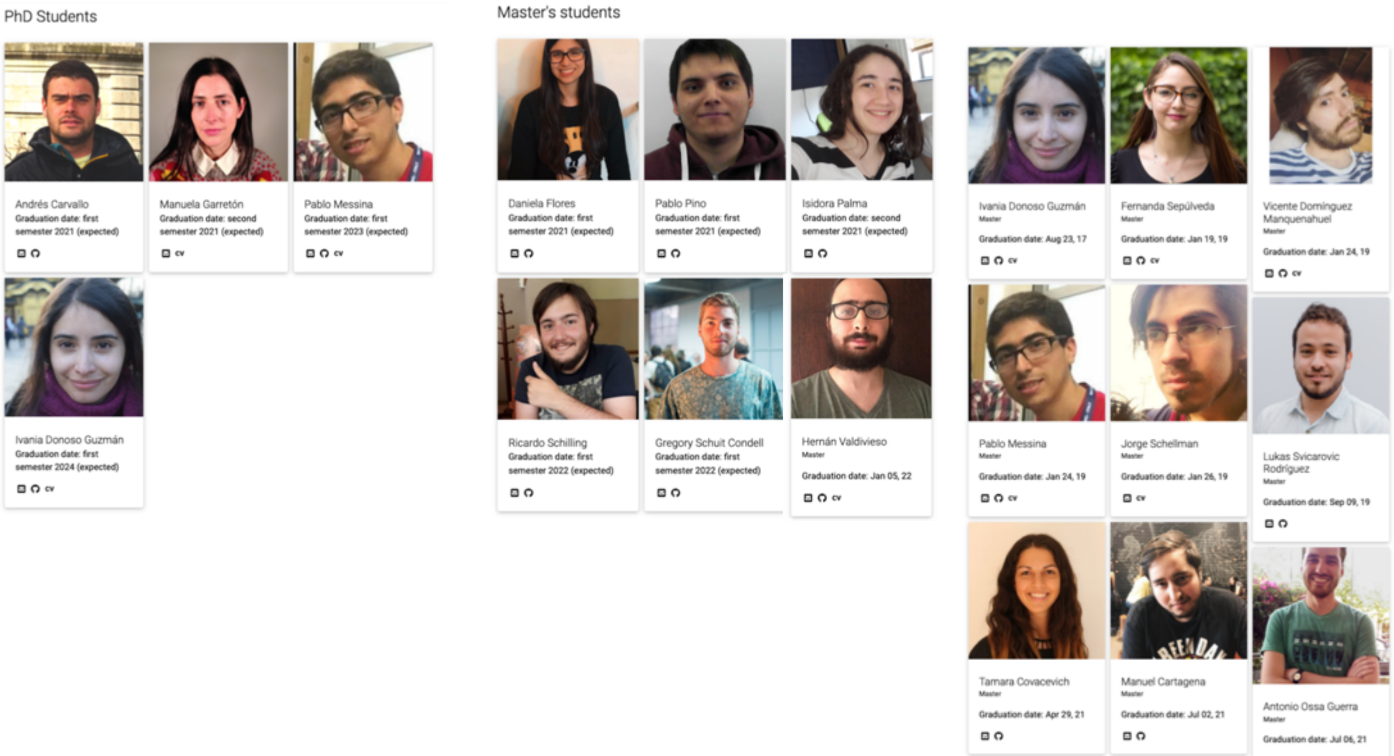
@denisparra

Presents ...


Denis Parra

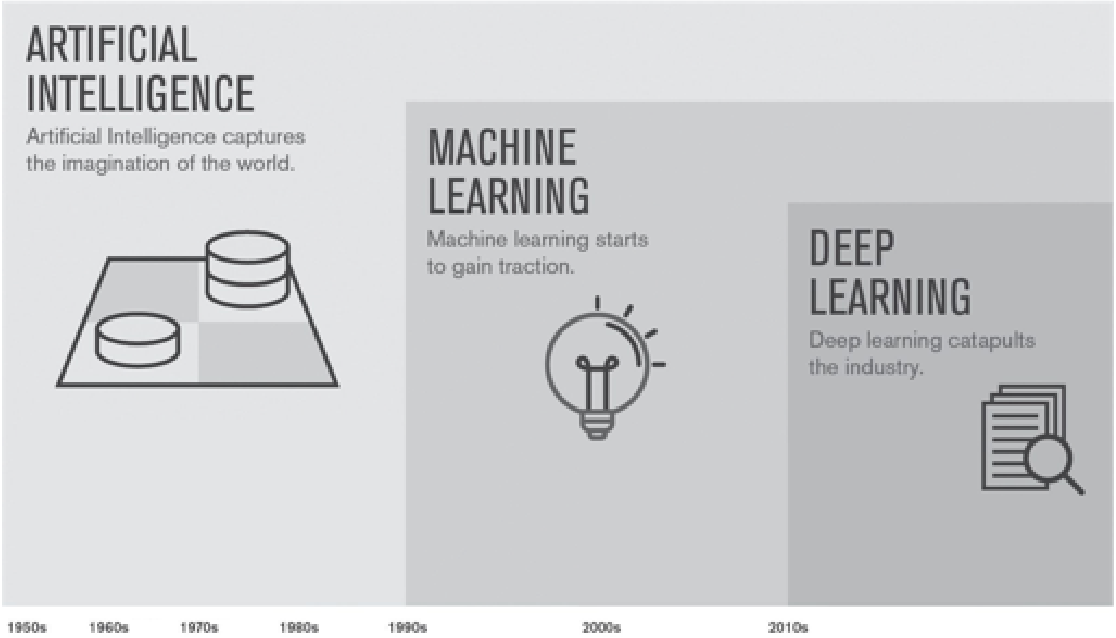
- Associate Professor, CS Department, PUC Chile
- Teaching: Recommender Systems, Information Visualization
- Research: HAiVis Lab ⊂ IA Lab
  - Machine learning applications (RecSys, Medical & Creative AI), intelligent user interfaces, information visualization
  - Students: 5 PhD, 5 Master, 1 Undergraduate
- Principal Researcher at:
  - Chile's National Center on Artificial Intelligence (CENIA, 2021 -)
  - Millenium Institute in Healthcare Engineering (iHealth, 2021 -)
  - Millenium Institute Foundational Research on Data (IMFD, 2018 -)

HAIVis UC: This research has been possible thanks to



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

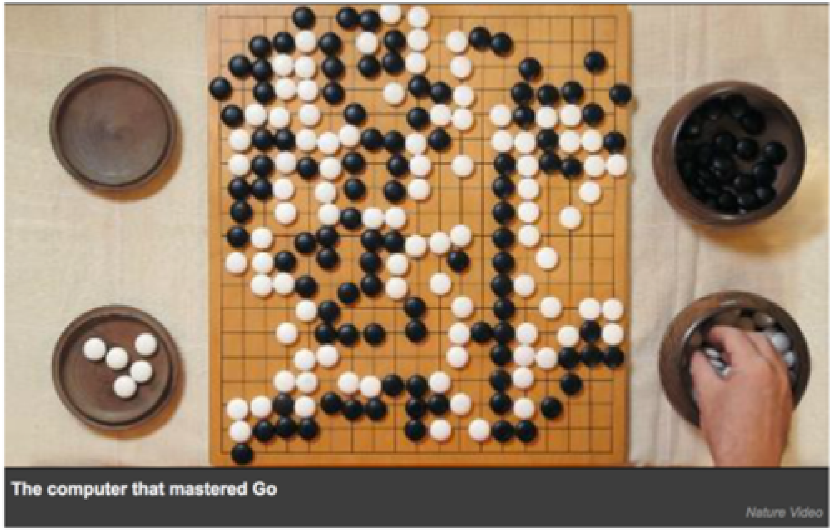

Google AI algorithm masters ancient game of Go

Deep-learning software defeats human professional for first time.

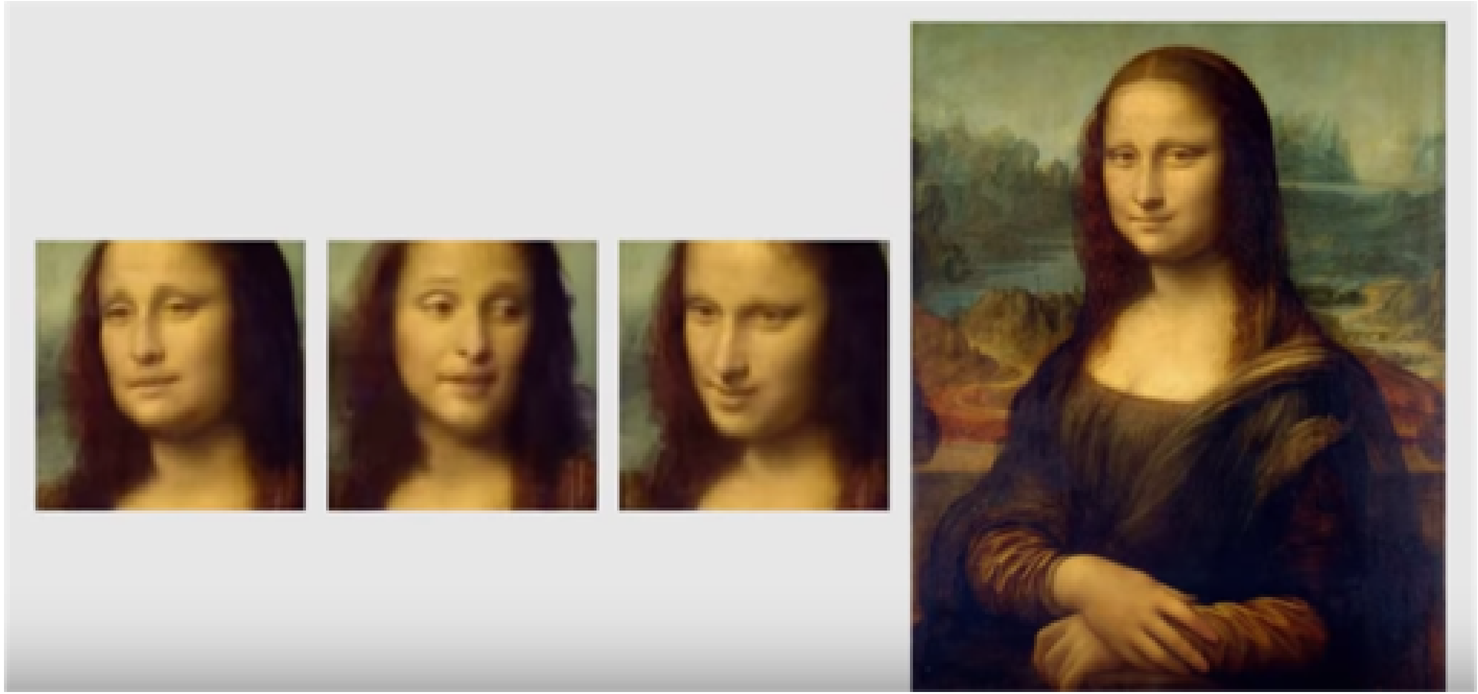
Elizabeth Gibney

27 January 2016

PDF Rights & Permissions



Zero Shot Learning: Picture to Movie



Zakharov, E., Shysheya, A., Burkov, E., & Lempitsky, V. (2019). Few-Shot Adversarial Learning of Realistic Neural Talking Head Models. *arXiv preprint arXiv:1905.08233*.

But there are some problems

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 23, 2016

ON 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.

- The COMPAS system is used in the USA to predict recidivism

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 23, 2016

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

But there are some problems

- 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

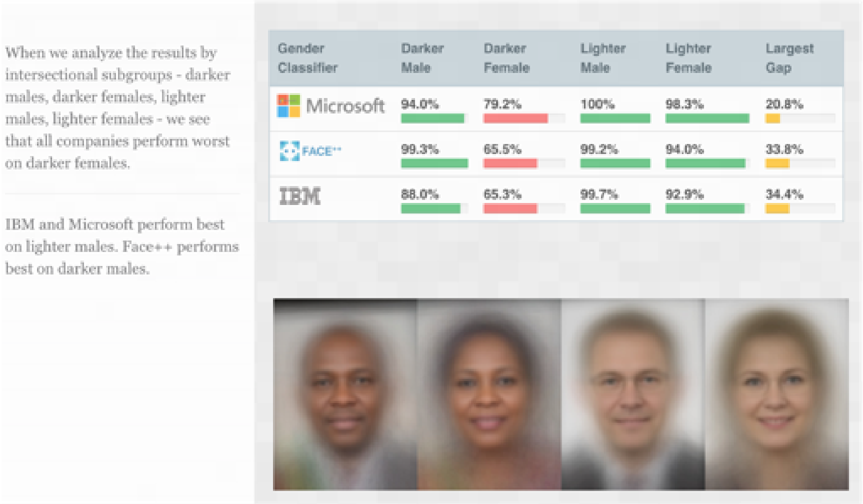
	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

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.


When we analyze the results by intersectional subgroups - darker males, darker females, lighter males, lighter females - we see that all companies perform worst on darker females.

IBM and Microsoft perform best on lighter males. Face++ performs best on darker males.

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

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

http://gendershades.org/overview.html

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

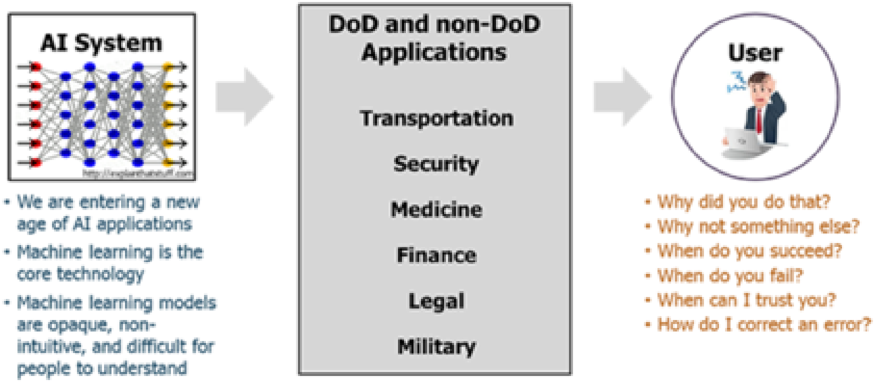
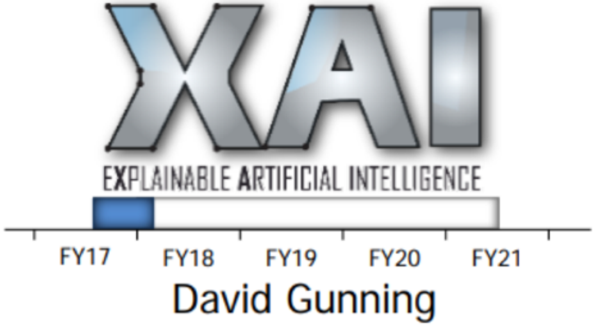


Figure 1. The Need for Explainable AI

Mr. David Gunning

Information Innovation Office (I2O)

Program Manager



What does XAI mean ?

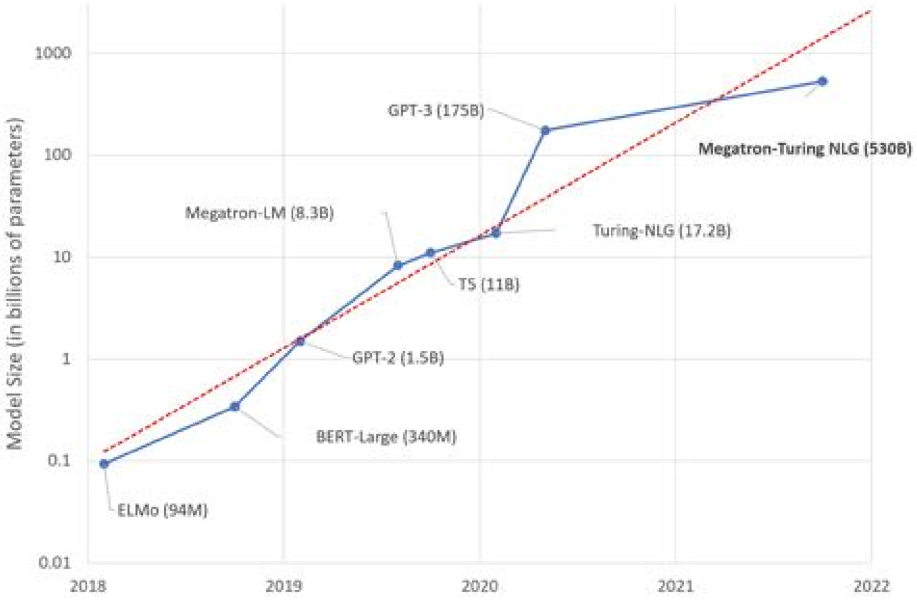
- In Gunning et al (2019), an X-AI system is defined by it ***Purpose***:
  - to make its behavior more intelligible to humans by providing explanations.

- 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.

Visual XAI

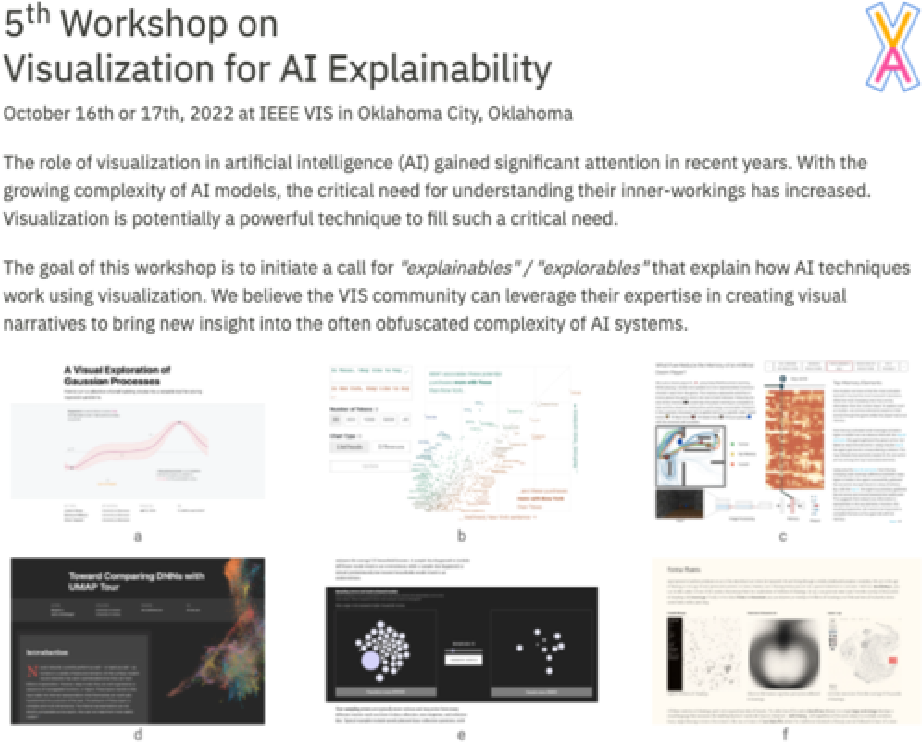
- With the growing complexity of AI models, the need for understanding their inner-workings has increased



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

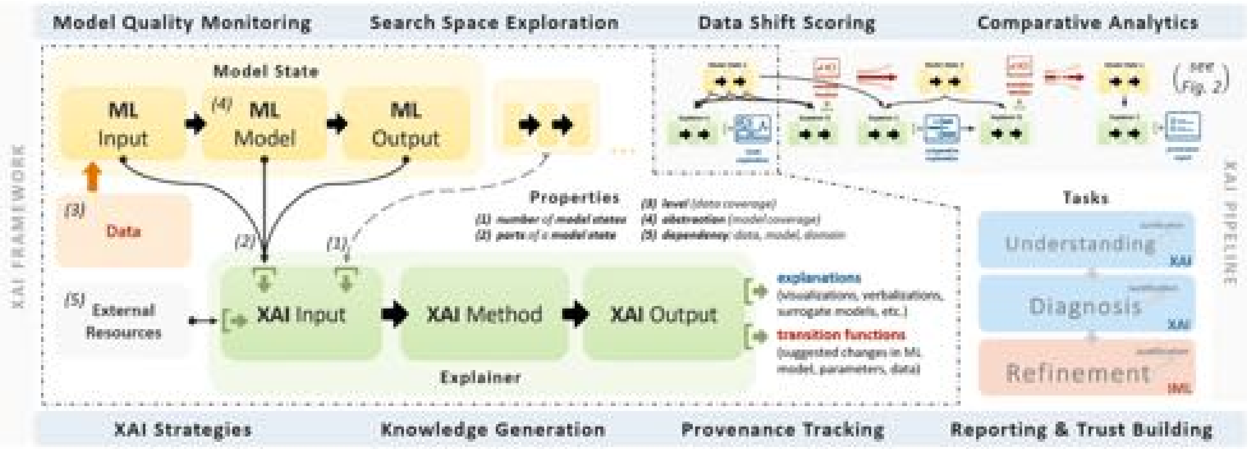
Visual XAI

- 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 – 5th VisXAI at IEEE Vis]



Visual XAI Frameworks

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

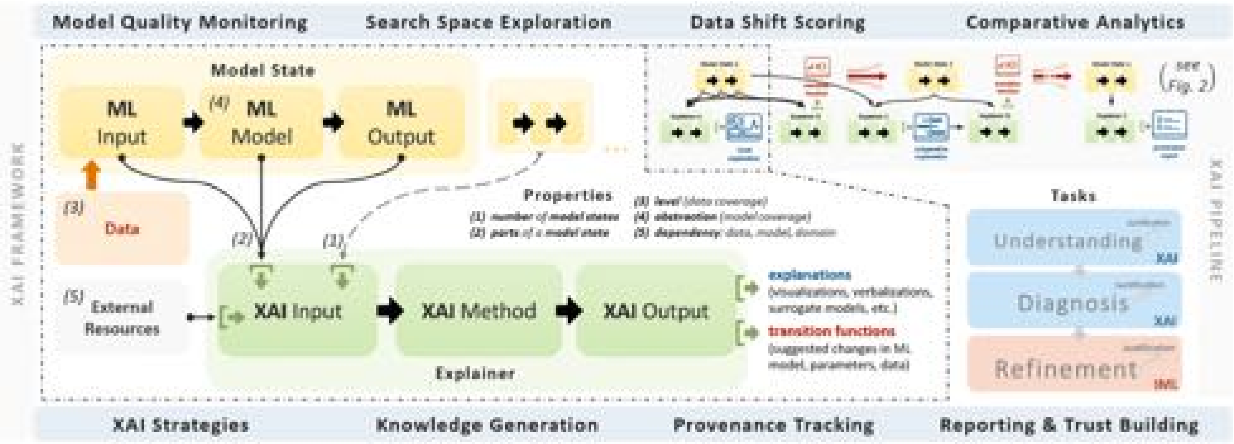


Spinner, T., et al (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE TVCG, 26(1), 1064-1074.

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 **<u>Not</u>** How to **<u>design</u>** and **<u>analyze</u>** a visualization for XAI



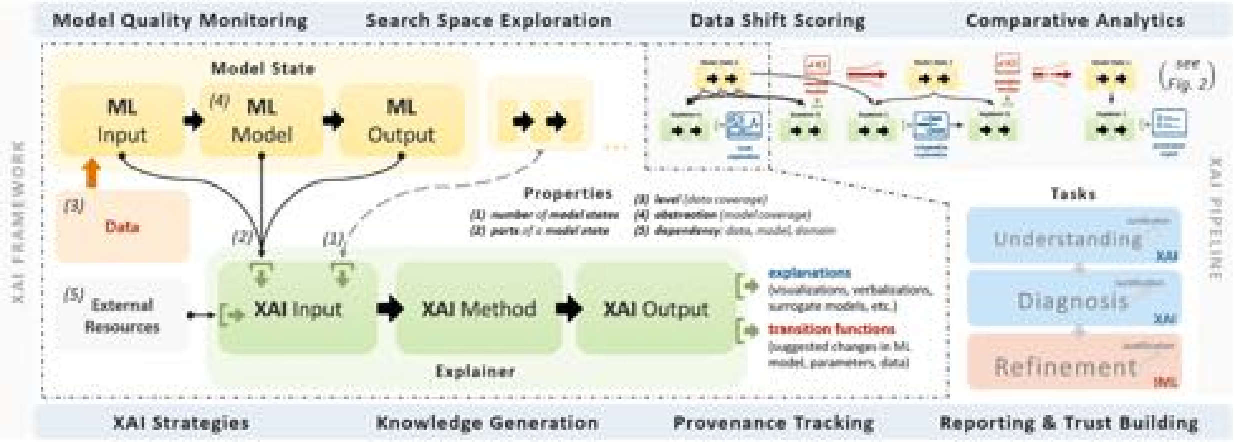
Spinner, T., et al (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE TVCG, 26(1), 1064-1074.

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

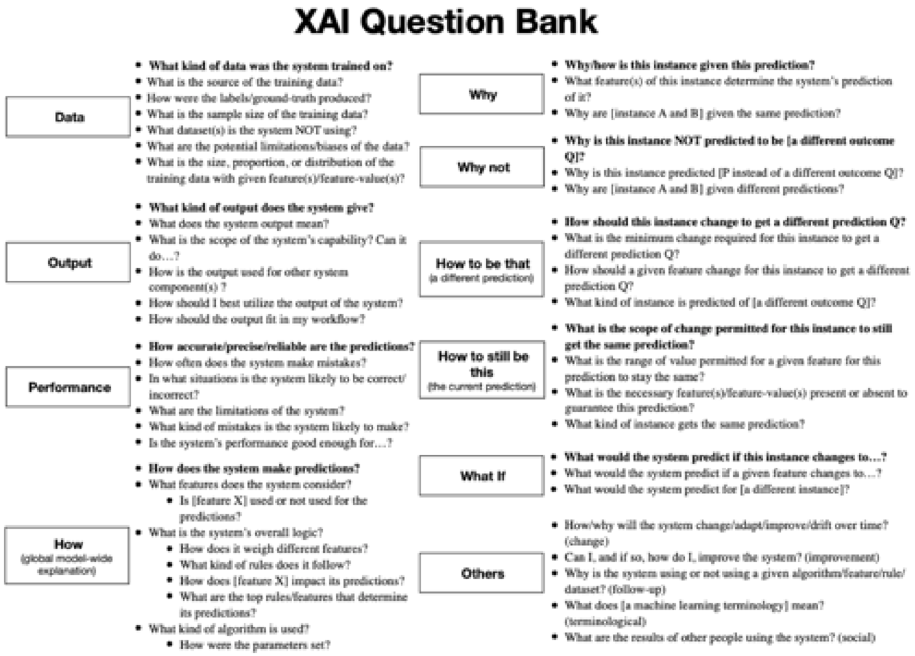


Spinner, T., et al (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE TVCG, 26(1), 1064-1074.

XAI question bank

- Vera 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

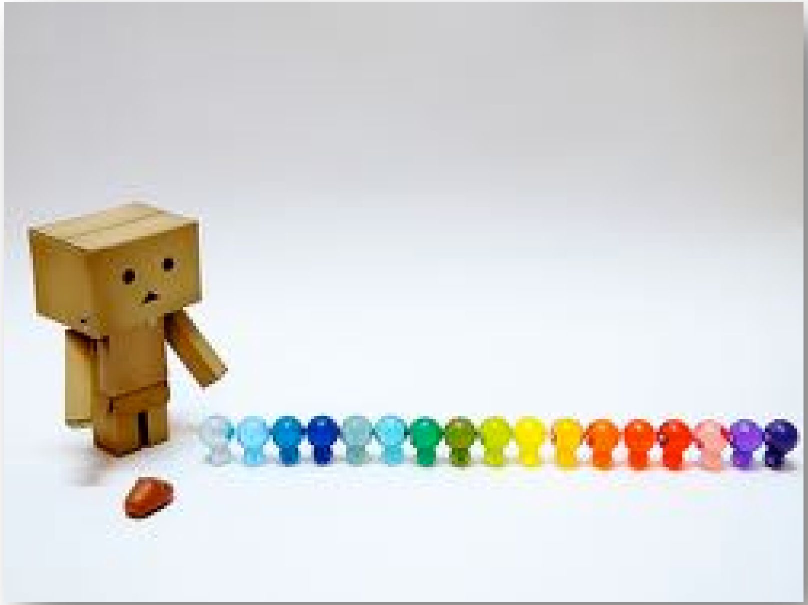
- An effective explanation will take the target user group of the system into account, **who might vary** in their **background knowledge** and **needs** for what should be explained.
  - How should we proceed under these circumstances ?

- 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.

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 ?



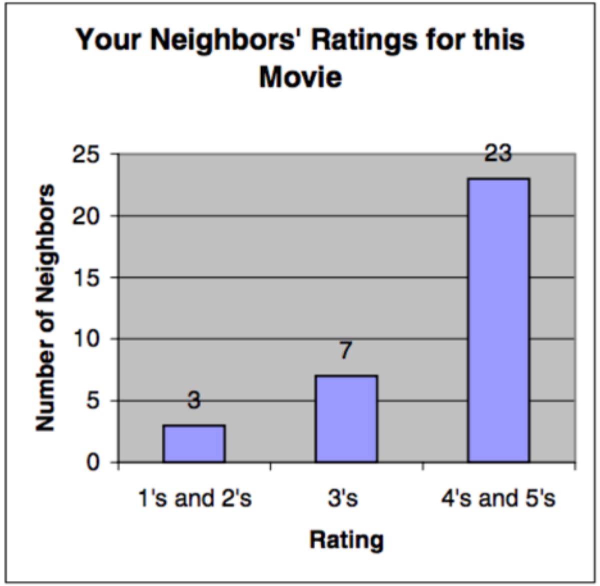
XAI in Recommender Systems: back to 2000s

- 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.



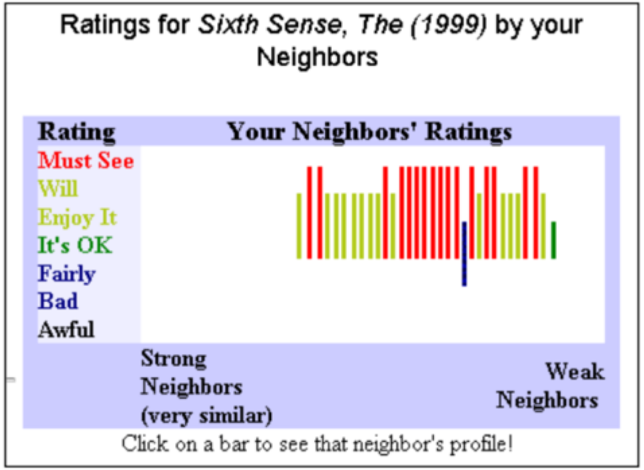
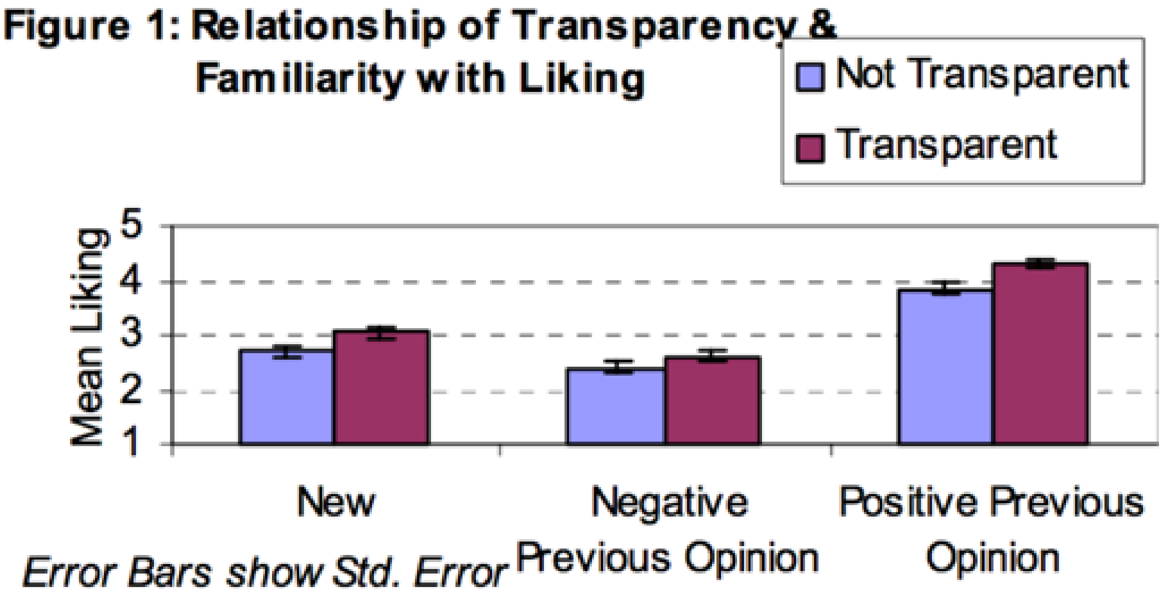


Figure 4. A screen explaining the recommendation for the movie "The Sixth Sense." Each bar represents a rating of a neighbor. Upwardly trending bars are positive ratings, while downward trending ones are negative. The x-axis represents similarity to the user.

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

2007

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

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.

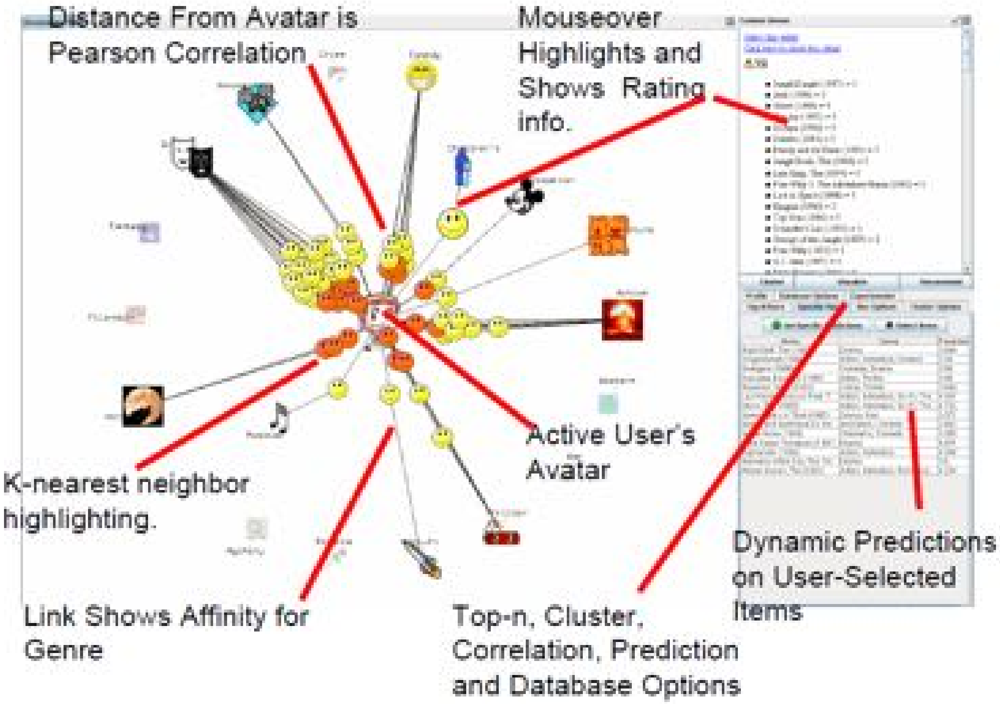
- In the coming slides, I will focus on research related **Visual User Interfaces** for recommendation Systems and how they contributed early to XAI and Visual XAI.

The Role of Interactive Visualization in RecSys XAI

- PeerChooser (O'Donovan et al, 2008)
- SmallWorlds (Gretarsson 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

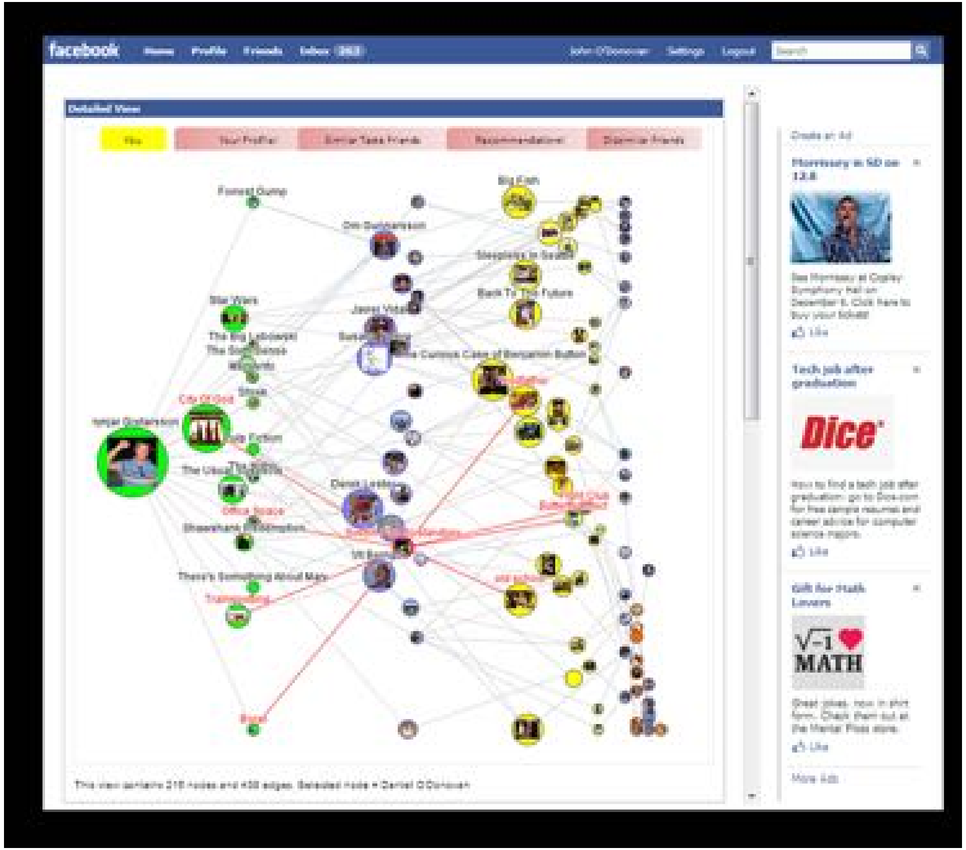
- PeerChooser (CHI 2008)
- John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, Tobias Hollerer



2: Annotated Screenshot of PeerChooser's Interactive Interface.

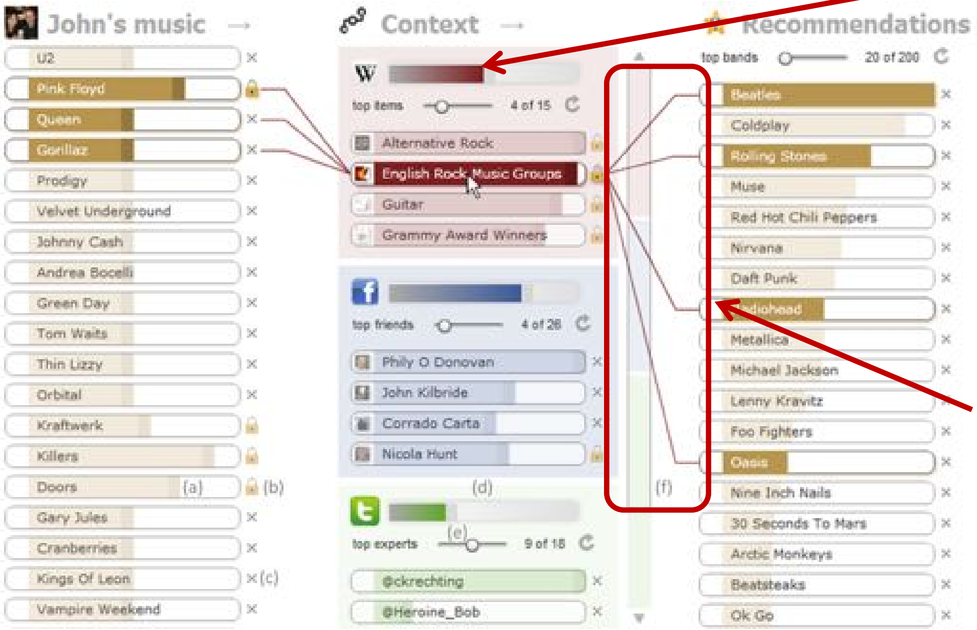
Smallworlds

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



TasteWeights

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



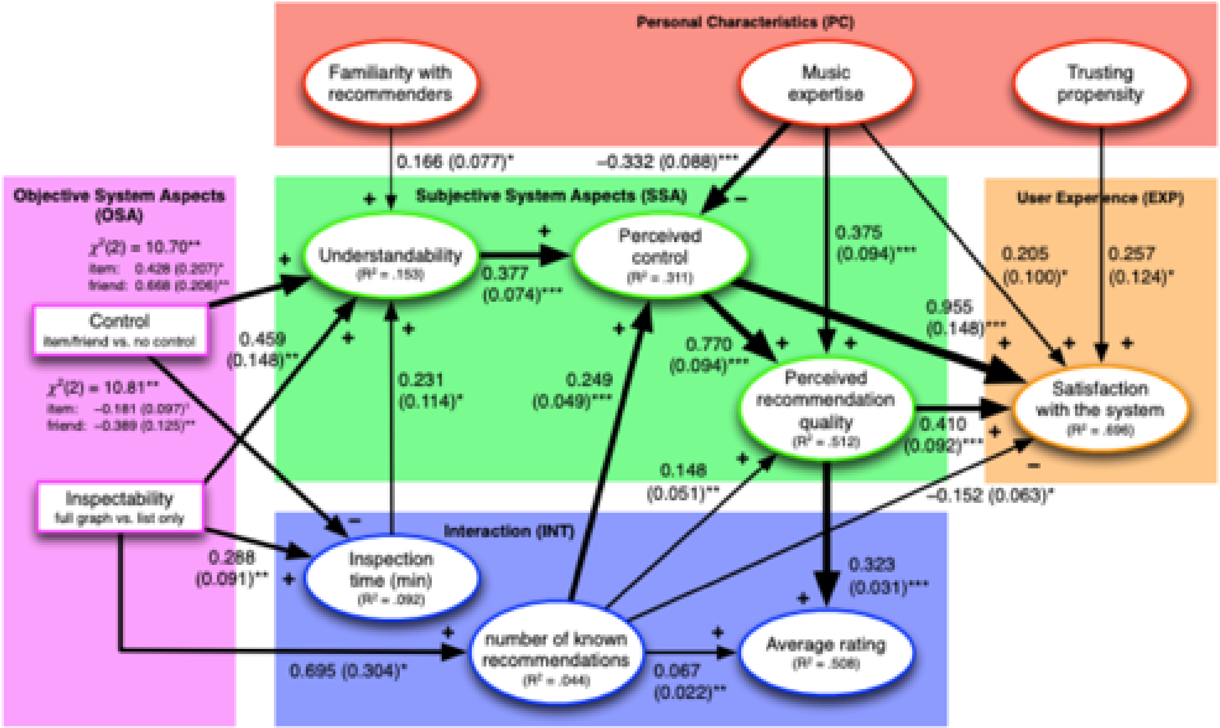
Controllability: Sliders that let users control the importance of preferences and contexts

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 (IUI 2013)
- Verbert, K., Parra, D., Brusilovsky, P., & Duval, E.

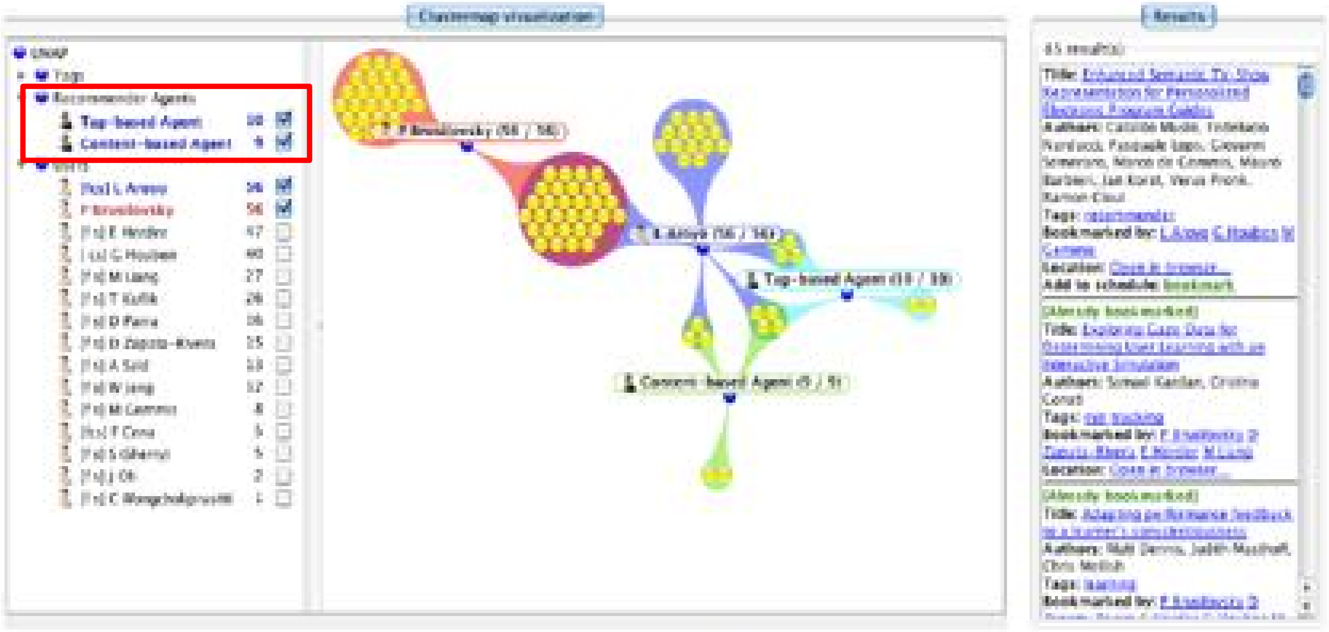
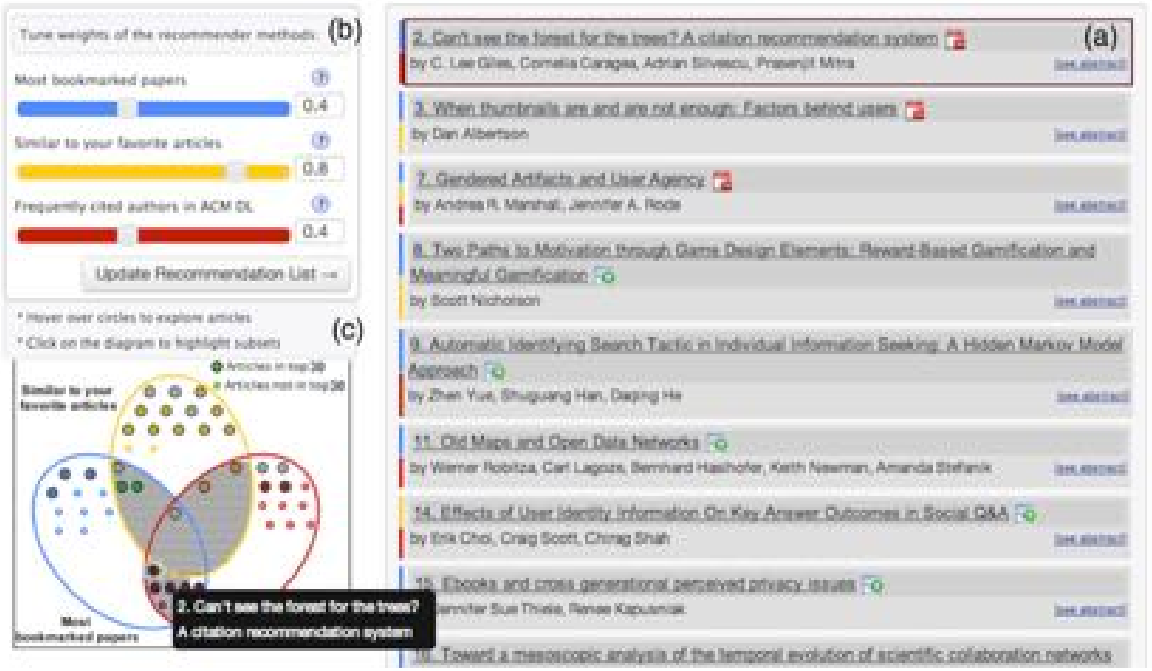


Figure 2: TalkExplorer

SetFusion

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

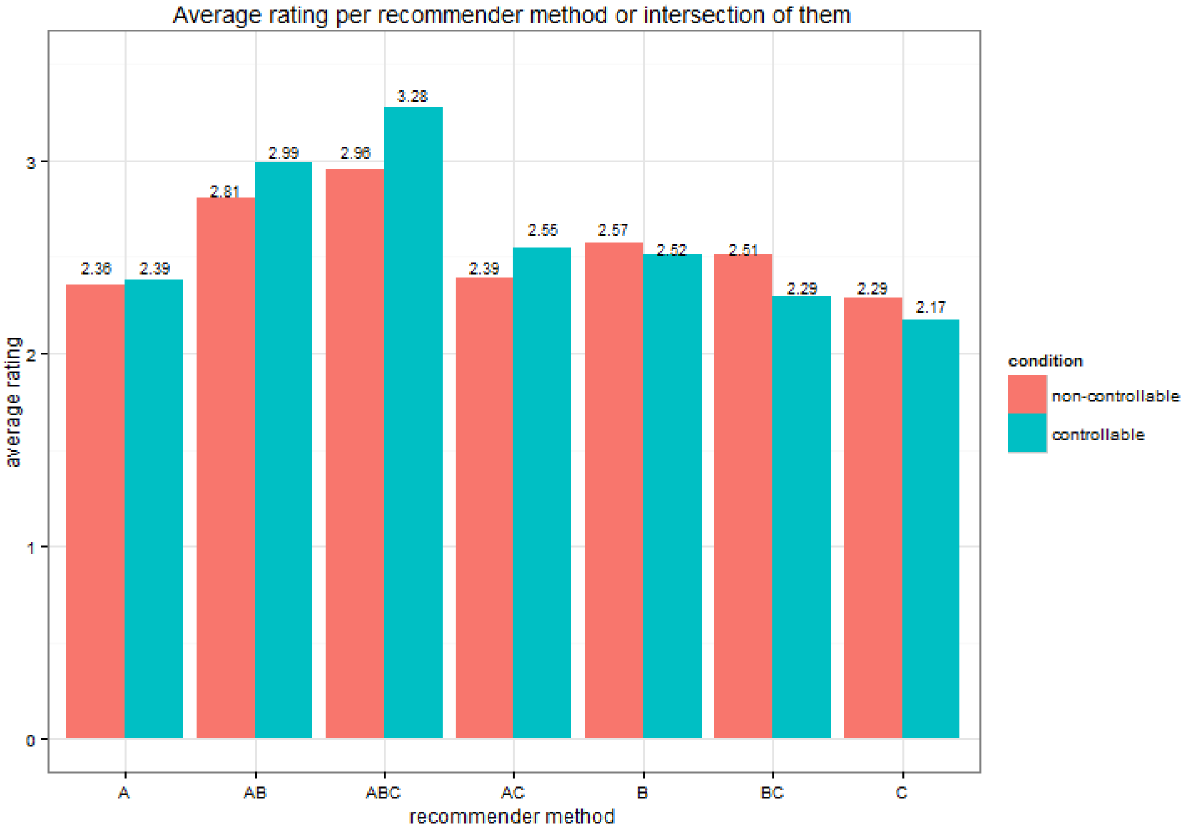


SetFusion: A Controllable Hybrid Recommender

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

IUI 2014, Haifa, Israel

SetFusion: User Control vs. non Control



Effect of Gender on Color Preference

Infographic Aesthetics: Designing for the First Impression

Lane Harrison
Tufts University
lane@cs.tufts.edu

Katharina Reinecke
University of Michigan
reinecke@umich.edu

Remco Chang
Tufts University
remco@cs.tufts.edu

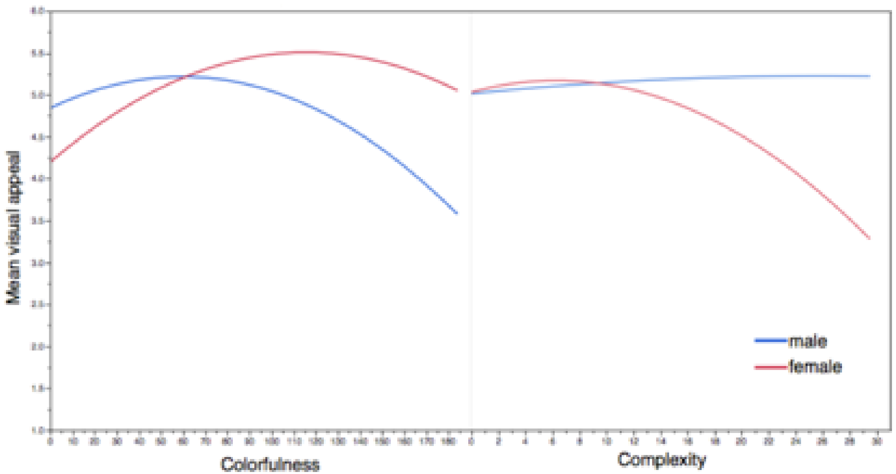
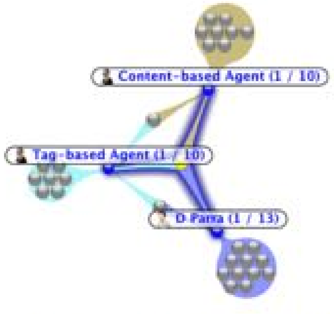
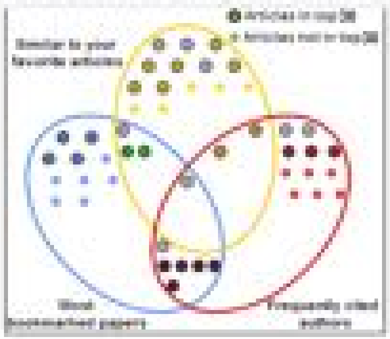


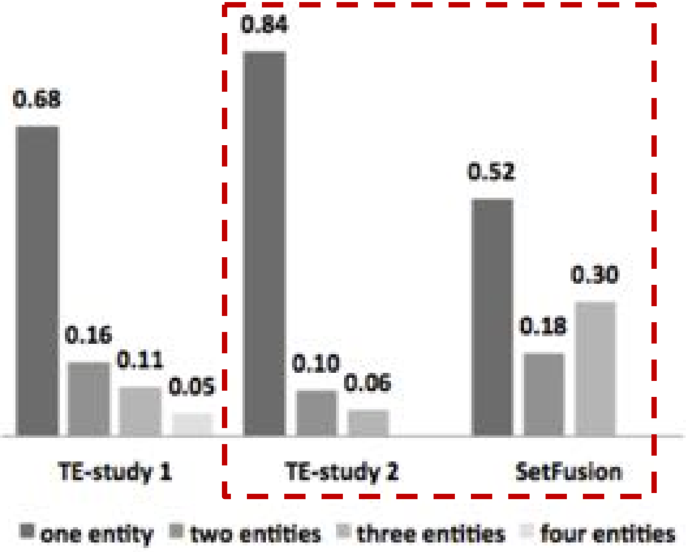
Figure 2: Average visual appeal ratings for different levels of colorfulness and complexity for females and males.

TalkExplorer vs. SetFusion

- How can we promote people's interaction?



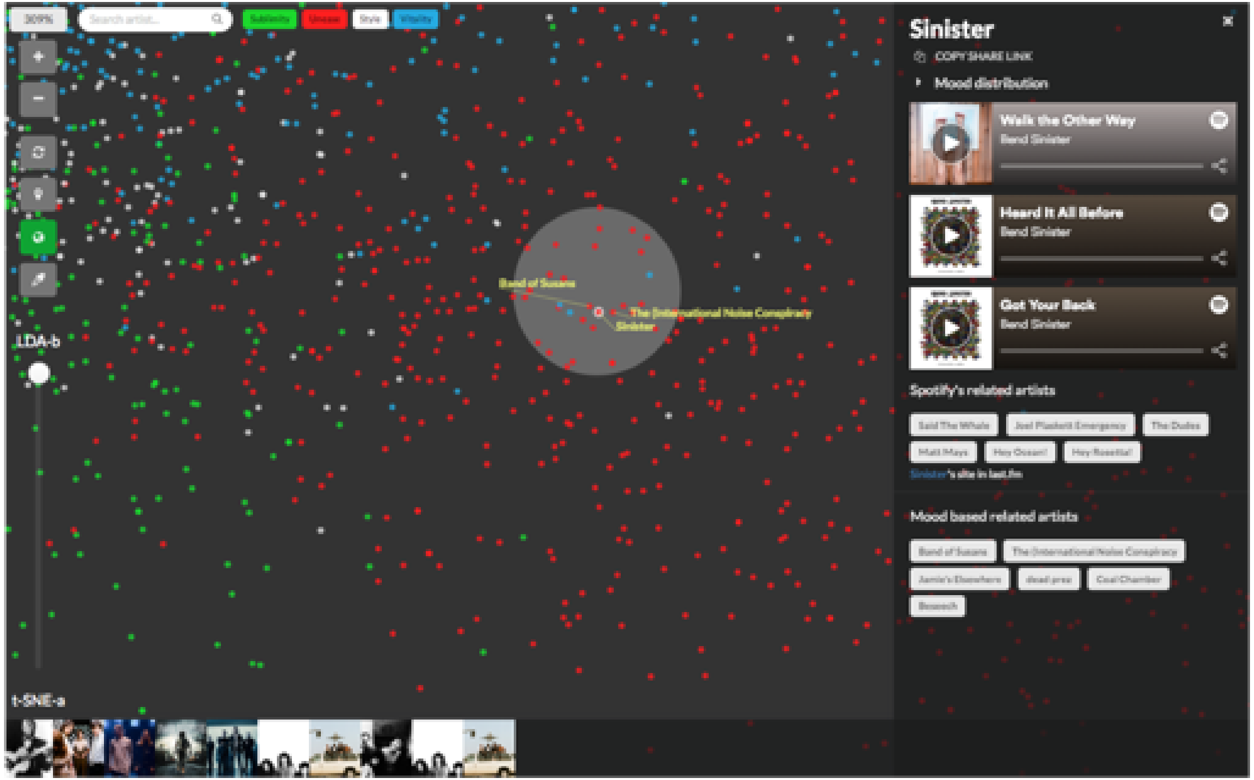




- In TalkExplorer, 84% of the explorations over intersections were performed over clusters connecting to 1 entity
- .
- In SetFusion, was only 52%, compared to 48% (18% + 30%) of multiple intersections.

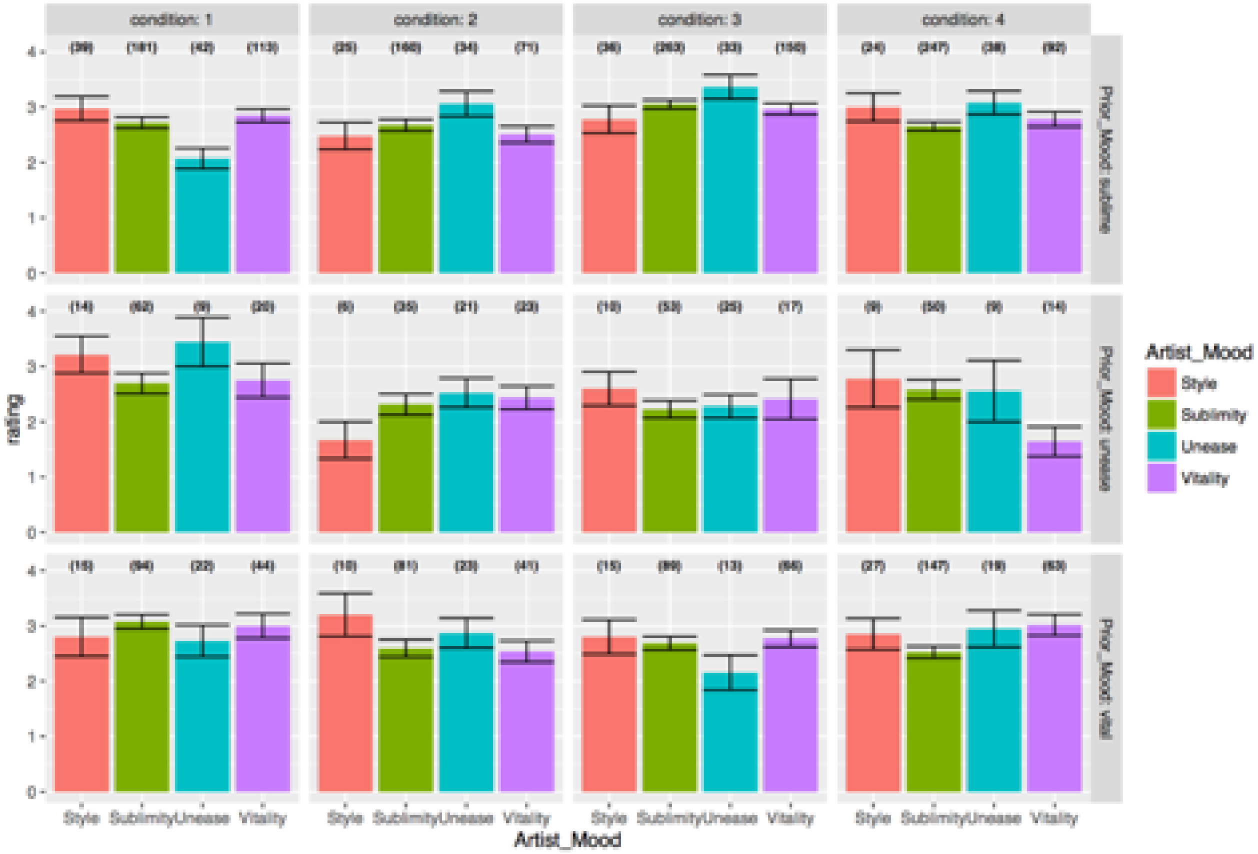
Moodplay

- Moodplay: Interactive Mood-based Music Discovery and Recommendation. (UMAP 2016)
- Ivana Andjelkovic, Denis Parra, and John O'Donovan.



http://moodplay.pythonanywhere.com/

Users' Prior Mood vs Artist Mood

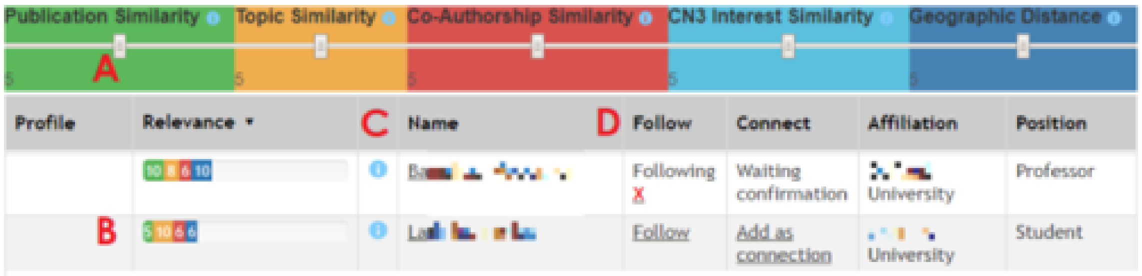


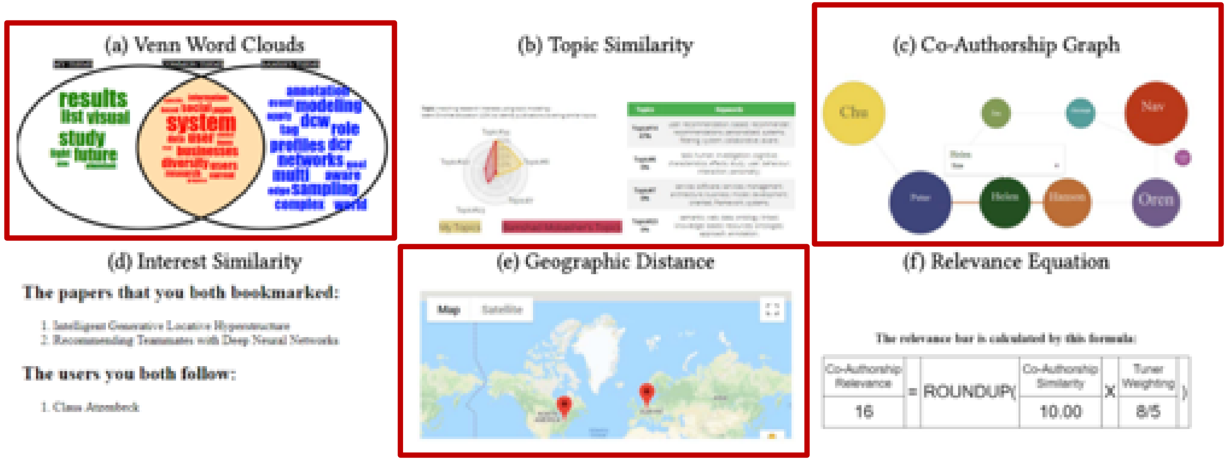
3D Landscape

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

Tuner+

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





To Explain or Not

- To explain or not to explain: the effects of personal characteristics when explaining music recommendations (IUI 2019)

- Millecamp, M., Htun, N. N., Conati, C., & Verbert, K.

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

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 see 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 see explanations or not
- Explanations should present different levels of details

If time allows ... Levels of details

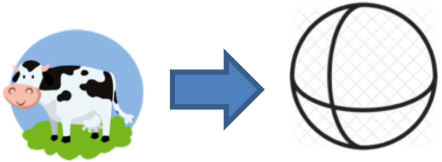
- A physicist, an engineer and a psychologist are called in as consultants to a dairy farm whose production has been below par.
  - **Engineer**: *Efficiency could be improved if the cows were more closely packed, with a net allotment of 275 cubic feet per cow*

If time allows …

- A physicist, an engineer and a psychologist are called in as consultants to a dairy farm whose production has been below par.
  - Engineer: *Efficiency could be improved if the cows were more closely packed, with a net allotment of 275 cubic feet per cow*
  - **Psychologist**: *The inside of the barn should be painted green. This is a more mellow color than brown and should help induce greater milk flow.*

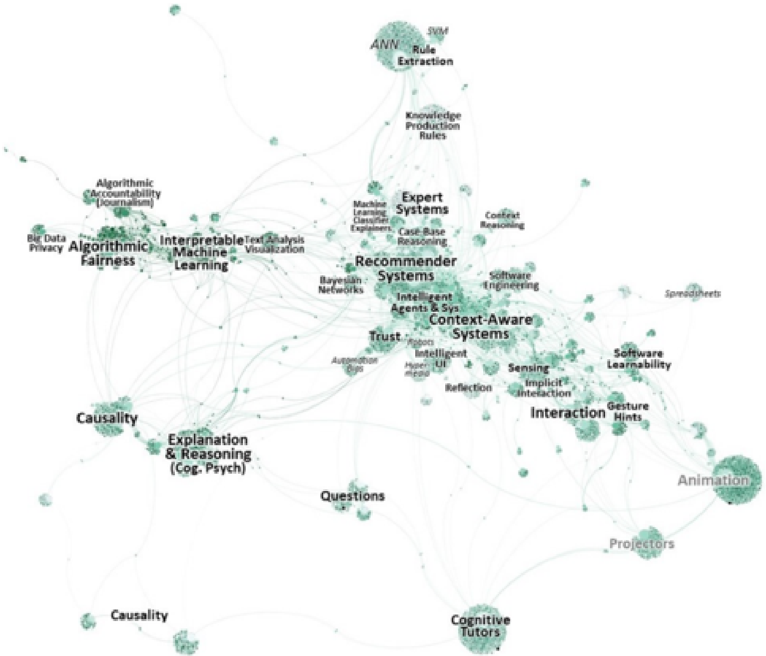
If time allows …

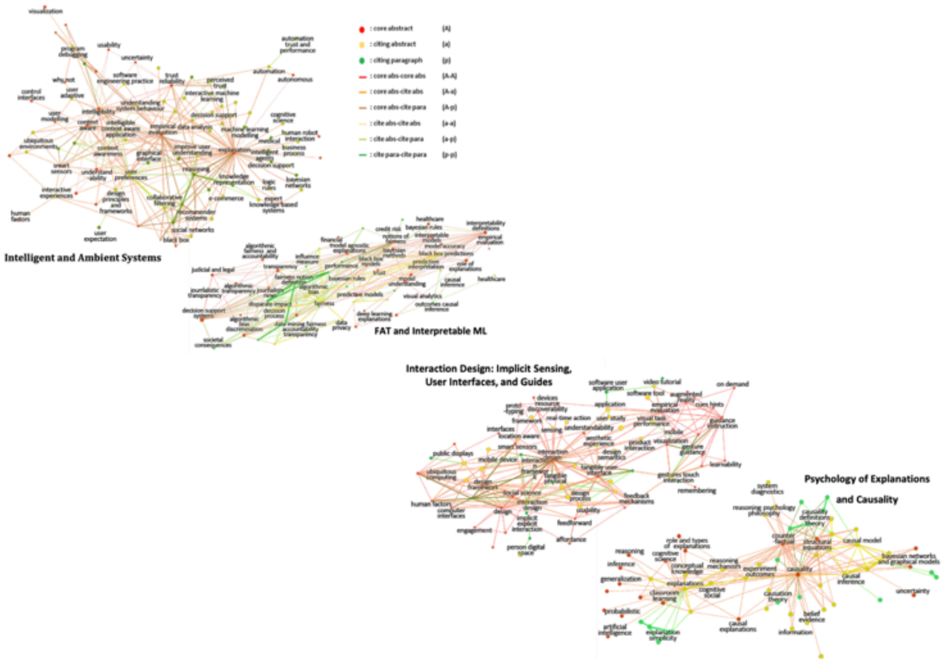
- A physicist, an engineer and a psychologist are called in as consultants to a dairy farm whose production has been below par.
  - Engineer: *Efficiency could be improved if the cows were more closely packed, with a net allotment of 275 cubic feet per cow*
  - *Psychologist: The inside of the barn should be painted green. This is a more mellow color than brown and should help induce greater milk flow.*
  - **Physicist***:* "Assume the cow is a sphere…."



Lawrence M. Krauss, Fear of Physics: A Guide for the Perplexed

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.

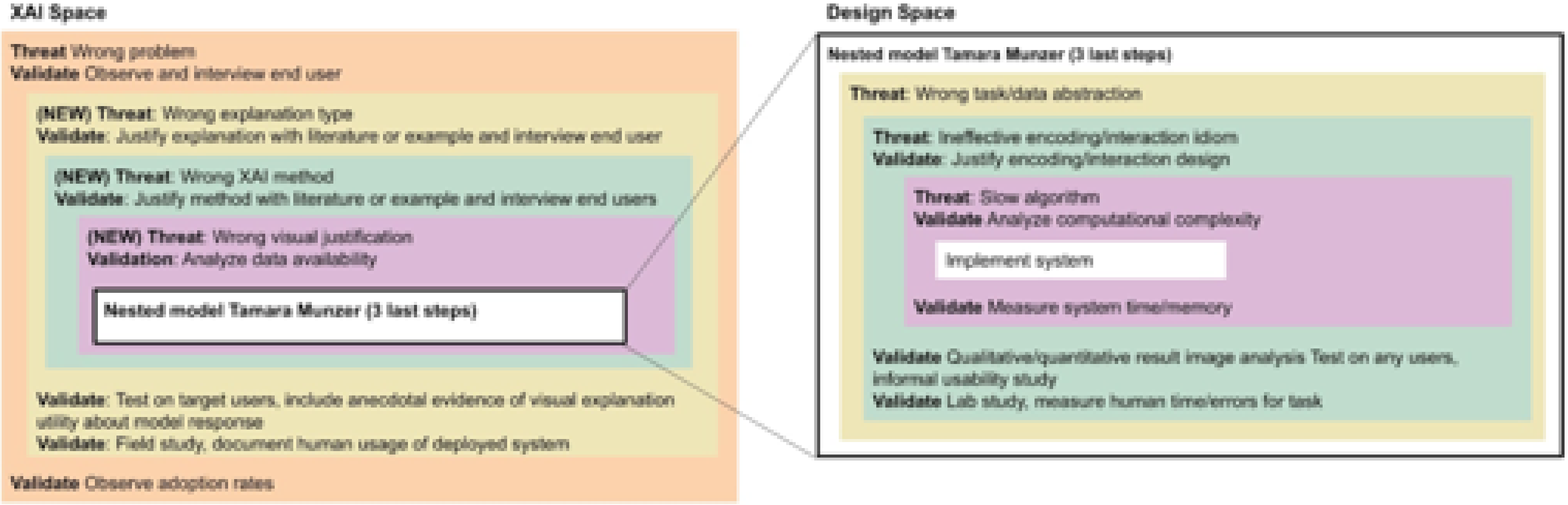
Back to Visual XAI

- How do we **analyze** and **design** visualizations for XAI?

- There are frameworks and guidelines for designing visualizations (e.g. Munzer VAD framework)

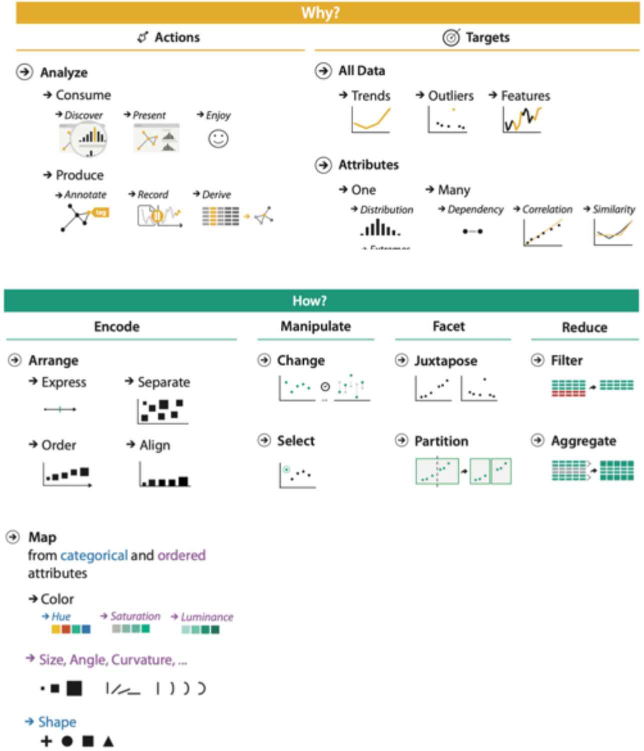
- … but the there is no framework yet for designing Visual XAI

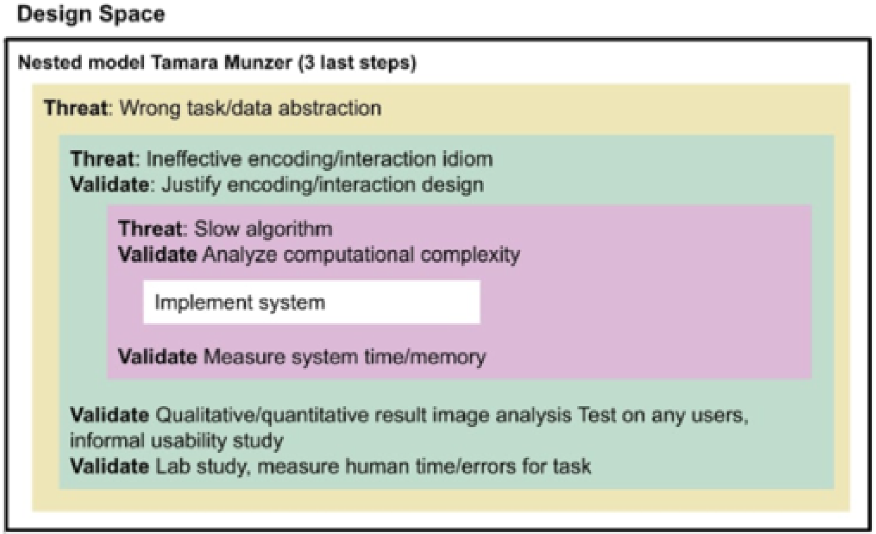
Work in Progress: VD4XAI



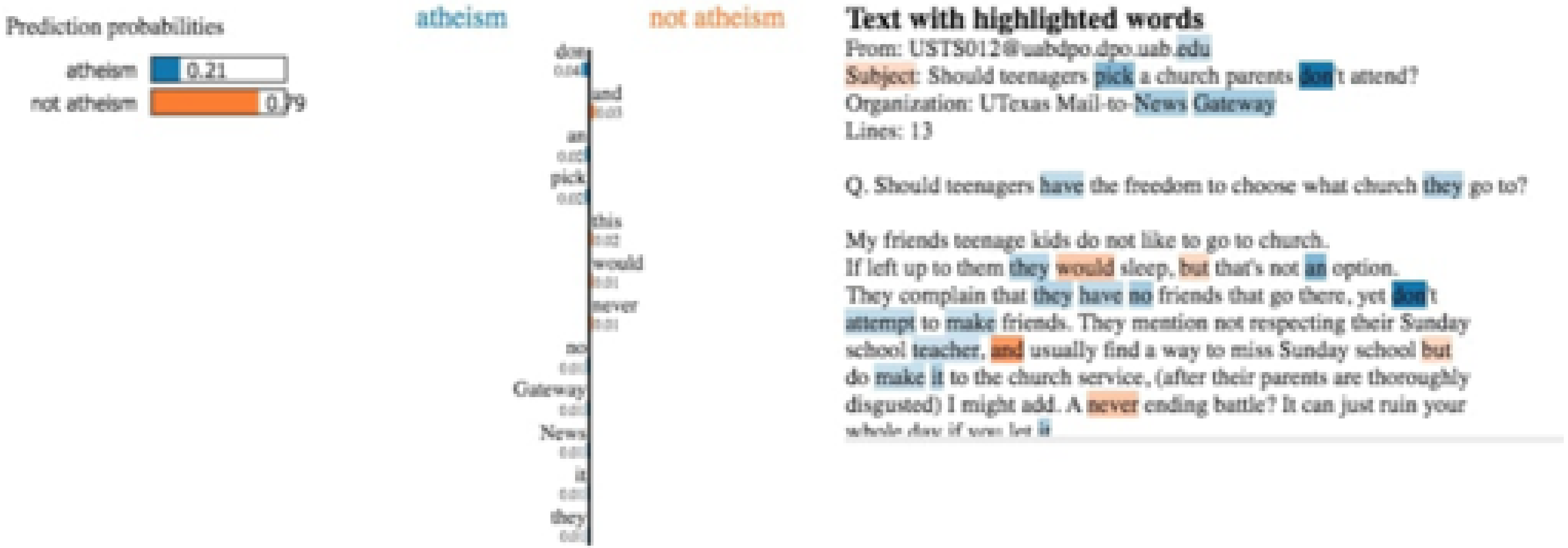
H Valdivieso, D Parra, I Donoso, T Schreck, C Wang and K Verbert.
VD4XAI: Connecting the Visualization Design Space with the Explainable AI Task Space. Work in progress

Munzner VAD model

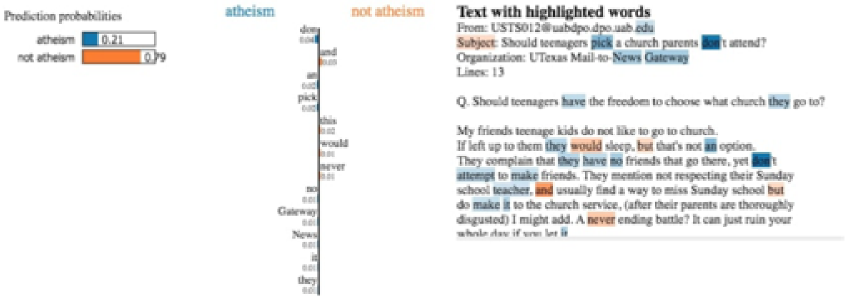


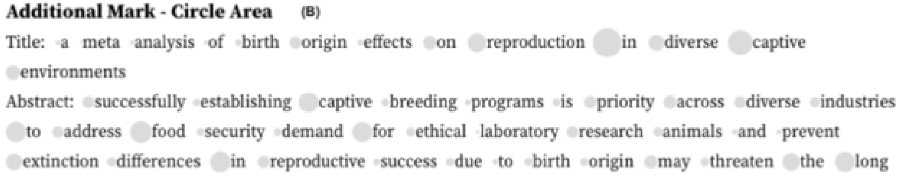


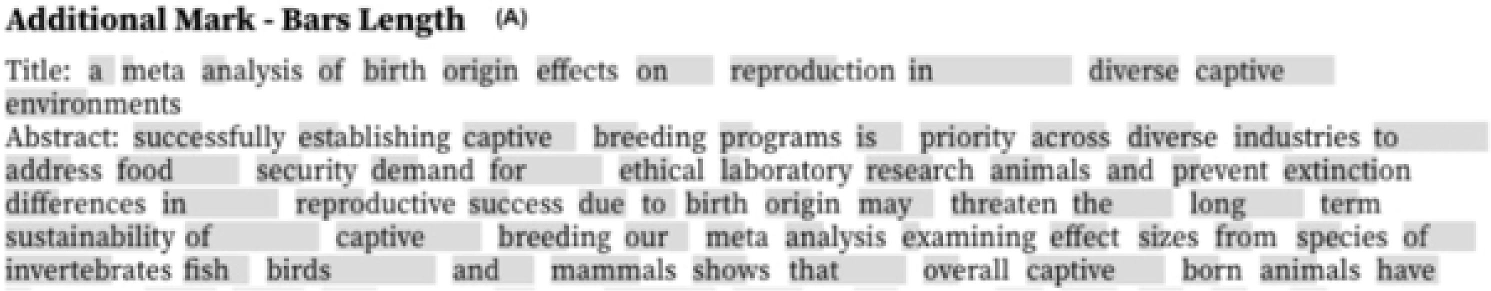
Analyze LIME



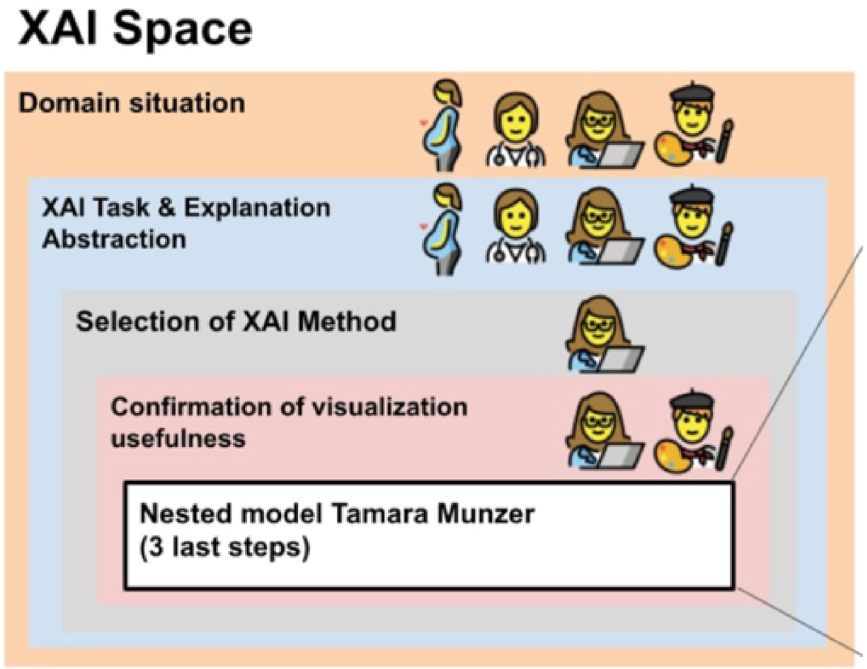
Analyze LIME: Marks, channels and Interaction





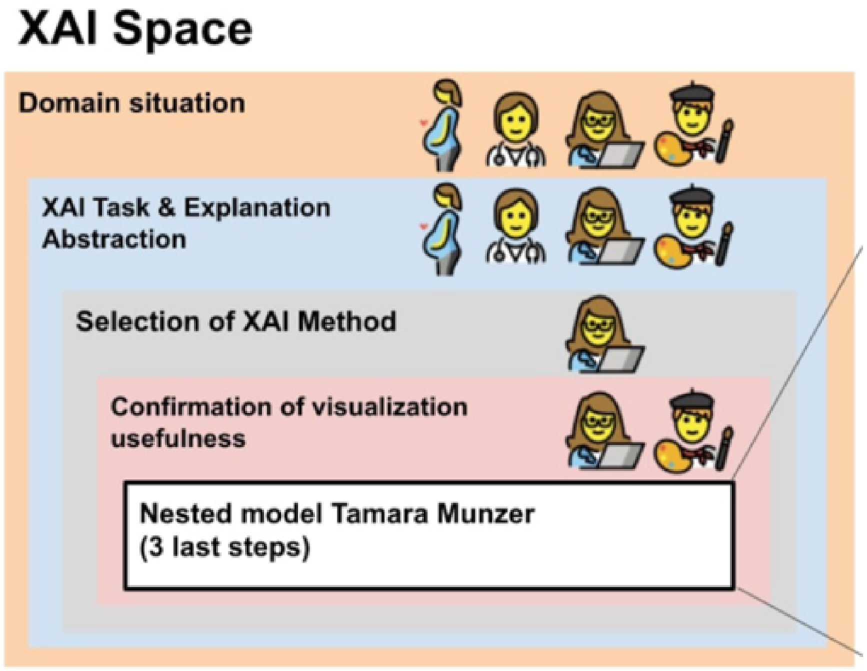


XAI Space



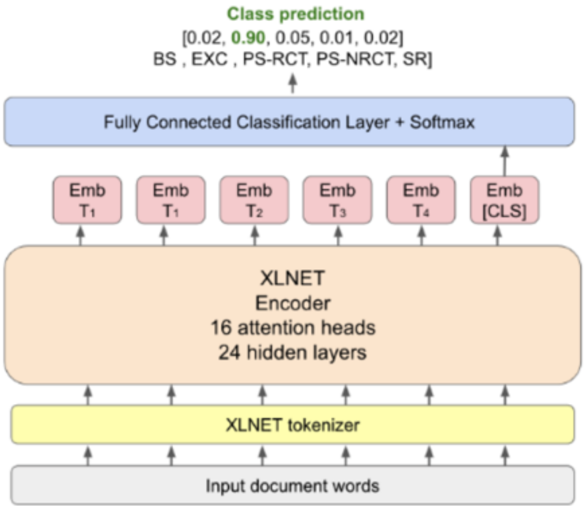
- Data type: tabular, text, image, graph, etc.
- **XAI tasks** (Gunning et al., 2019):
  - Understand why does the given AI/ML model give its prediction,
  - Understand when does the AI/ML model fail,
  - Understand when does the AI/ML model succeed,
  - Understand why the AI/ML model does not predict something else

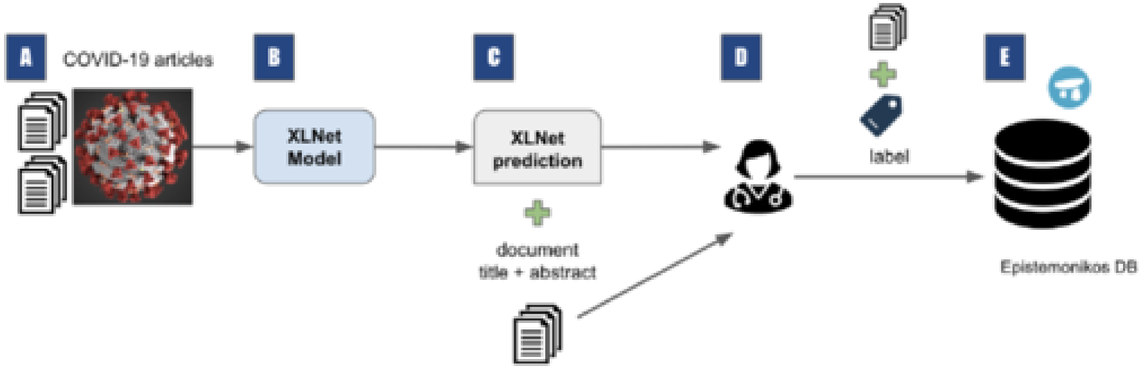
XAI Space



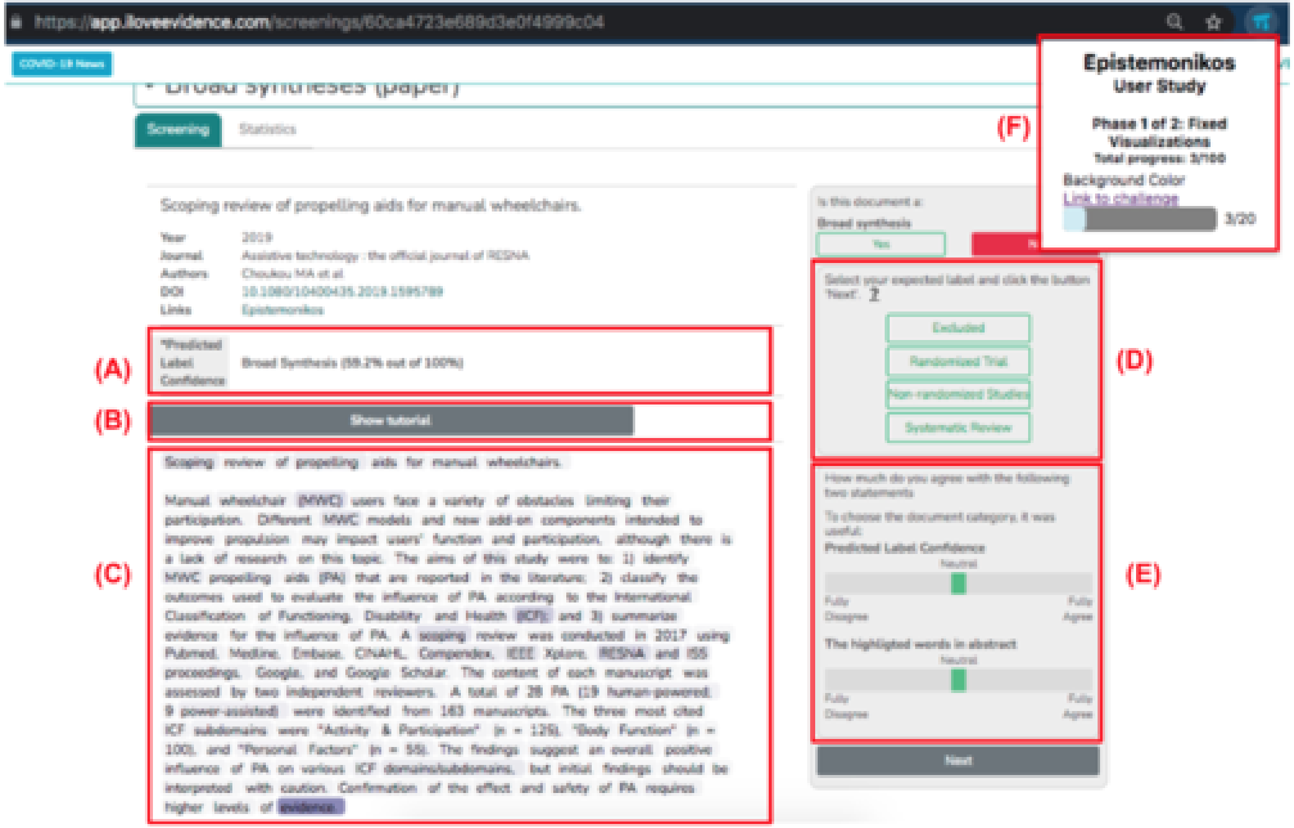
- **<u>XAI Strategy</u>**:
  - Feature importance,
  - Analogies (similar to),
  - Counterfactuals,
  - Rules
- **<u>XAI method</u>**:
  - Post-hoc: LIME, SHAP, etc.
  - Method-wise: DT, LR
  - Others
- AI model (neural network, SVM, DT, etc.)

Recent study: Biomedical Document Classification

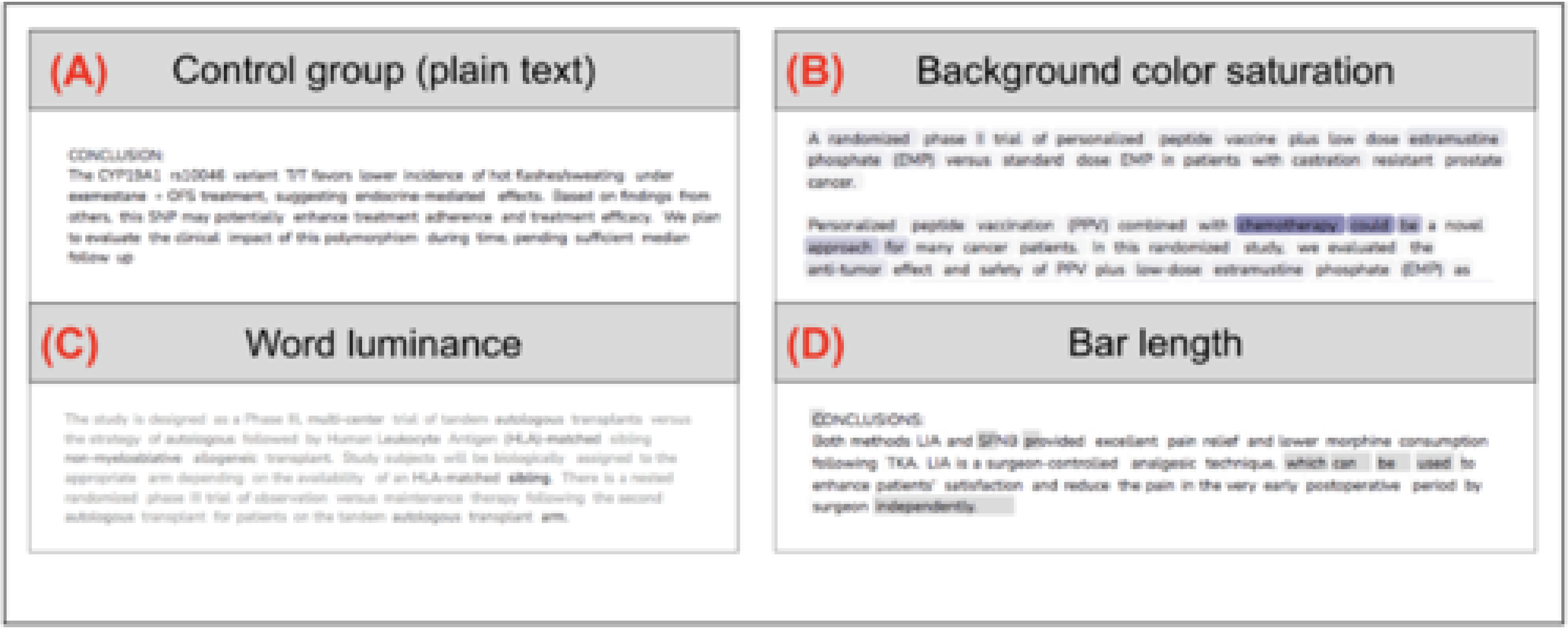




Recent study: Biomedical Document Classification

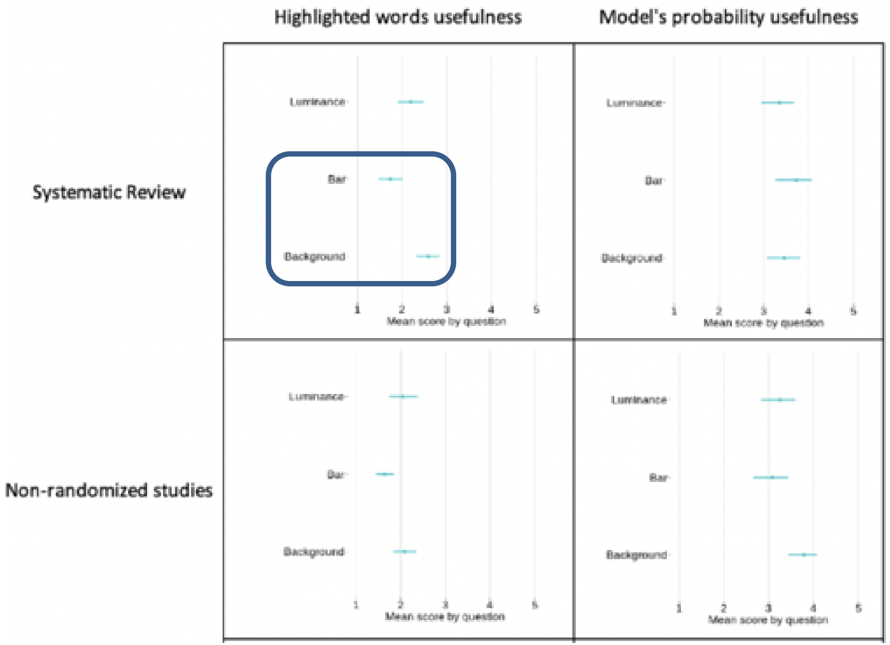


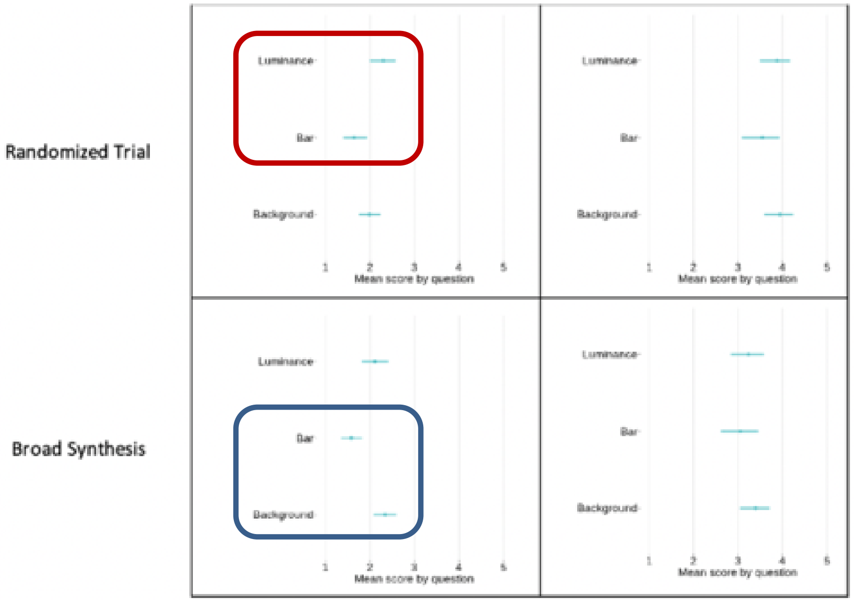
Designs to test: marks and channels



Recent study: Document Classification

* Perception of usefulness of visual XAI encoding for attention-based explanations vary **depending 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!

FUNDING: ANID FONDECYT 1191791