NDBI021, Lecture 5

User preferences, 2/1 ZK+Z,

Wed 12:20 – 13:50 S8

Wed 14:00 – 15:30 SW2 (odd weeks)

https://www.ksi.mff.cuni.cz/~peska/vyuka/ndbi021/2022/

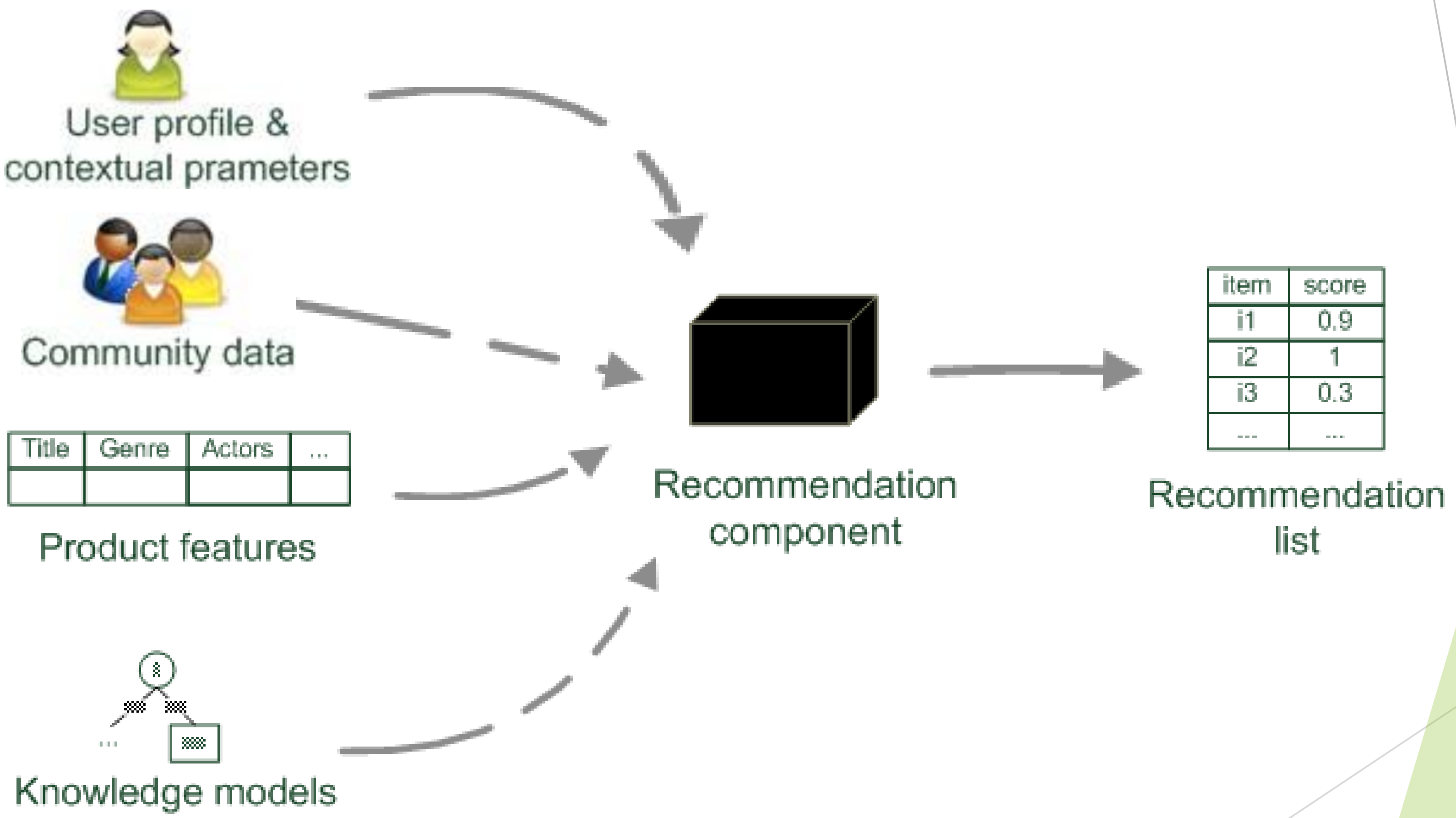
matfyz

siret research group

https://ksi.mff.cuni.cz

Recommender Systems: recap

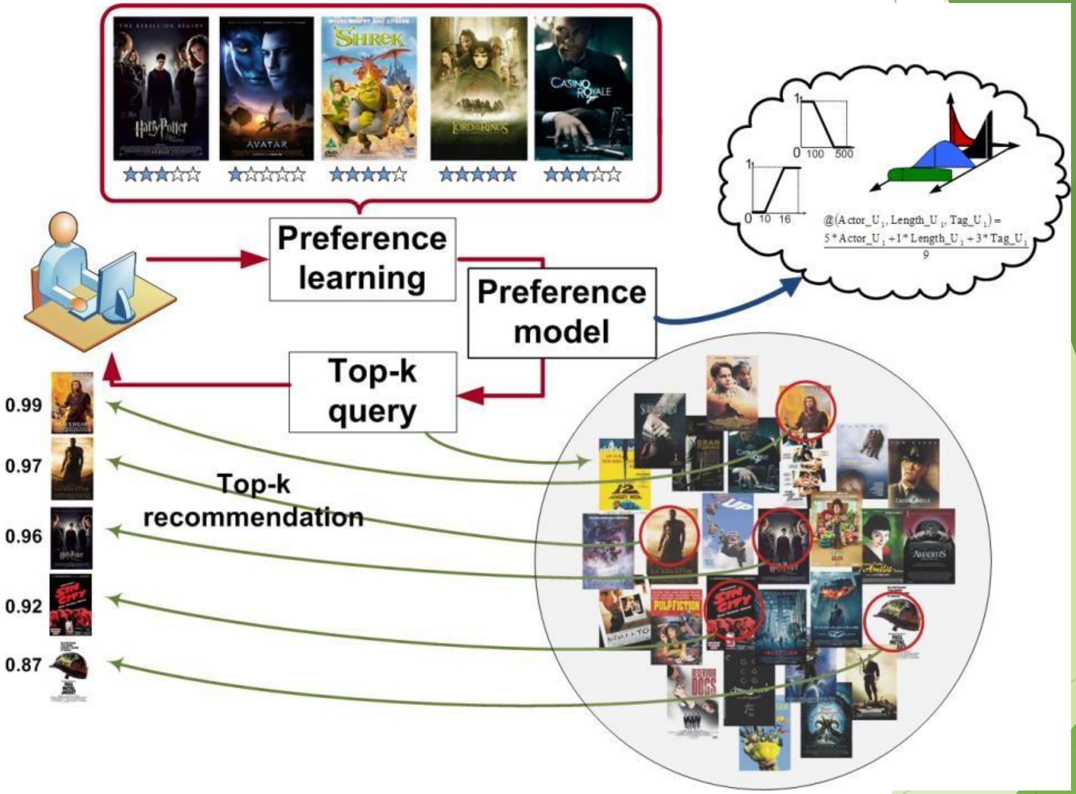
Paradigms of recommender systems



Lifecycle of Recommender Systems

1. **Get User Feedback**
2. **Learn internal model**
3. **Upon demand, recommend objects**

- **The process is asynchronous by nature**
- **Most recent usually most relevant**
- **Dynamic nature of the process seriously complicate things**
    - Partial re-train / model updates
    - Long-term vs short-term (context), preference drift
    - Repeated consumption & recommendation



Basic algorithms

- Non-personalized & item-based & session-based models
- KNN variants
- Matrix factorizations
  - Content-aware factorization methods
- Reinforced learning / multi-armed bandits

Basic evaluation

- On-line / off-line / lab studies
- Accuracy-based metrics (Recall, nDCG, MAP,…)
- „Beyond accuracy“ (diversity, novelty, coverage, fairness, popularity bias…)
- „Technical“ (time complexity, scalability, ability to predict for all…)

Fairness in Recommender Systems

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Tutorial on Fairness of Machine Learning in Recommender Systems

Yunqi Li
Rutgers University
yunqi.li@rutgers.edu

Yingqiang Ge
Rutgers University
yingqiang.ge@rutgers.edu

Yongfeng Zhang
Rutgers University
yongfeng.zhang@rutgers.edu

Tutorial from: https://fairness-tutorial.github.io/

Fairness issues in RecSys and IR

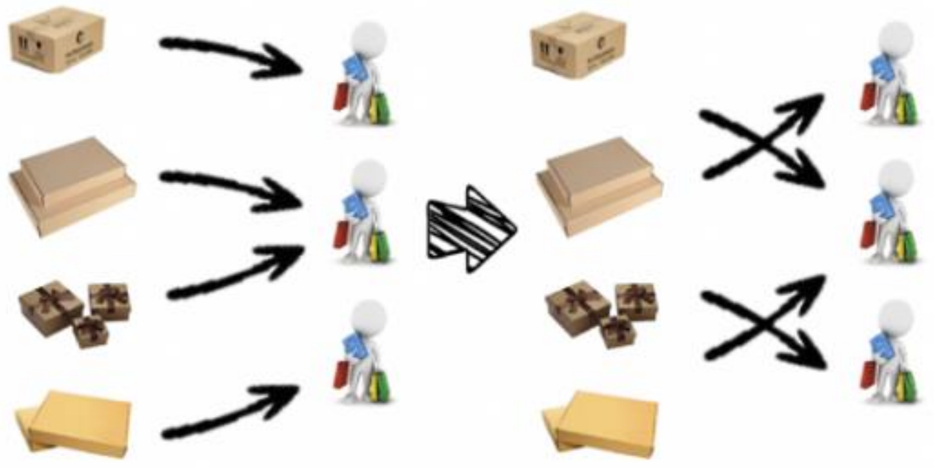
- News recommendation/social networks
  - Does the suggested articles close me into some opinion bubble?
    - Fairness of the presented opinions on controversary subjects
- Job matching & marketplaces
  - Am I omitted from the list of possible applicants just because [black/old/female…]
  - Is one content provider favored over others?
- Finance domain
  - Why am I not recommended for loan? Why is my credit score lower/higher?
- E-commerce
  - Is this product being recommended because it is the best for me… or because the provider earns the most from it?

What if these features are learned indirectly?

Social Impacts of Recommender Systems

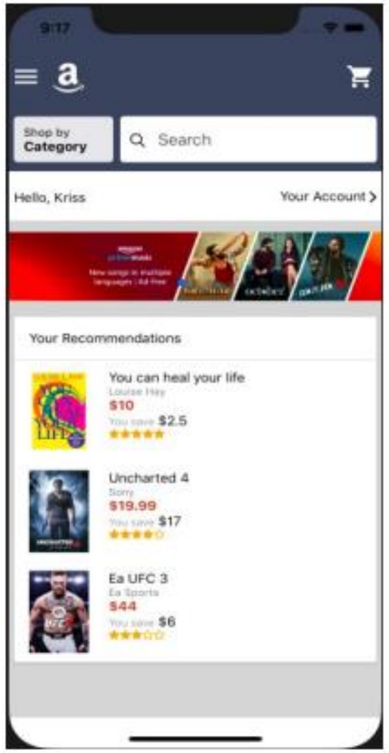
- Recommender Systems are far more than just information seeking tools
  - They control how resources are allocated among differnet parties
    - Resources can be exposure opportunies, products, jobs, information, etc.
    - Usually RS works in two-sided markets/environments [1]

The *Prosumer* Paradigm:
Consumers – *items* – *Producers*
Buyers – Goods – Sellers
Freelancer – Jobs – Employers
Borrowers – Money – Lenders
Passengers – Services – Drivers



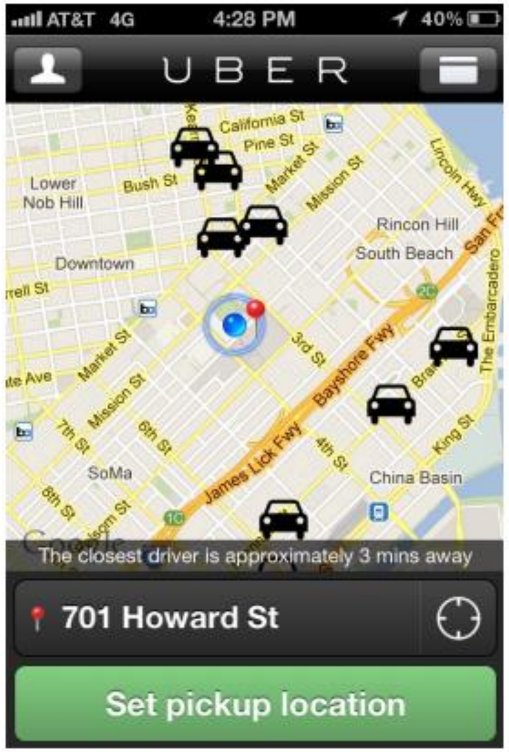
[1] Y. Zhang, Q. Zhao, Y. Zhang, D. Friedman, M. Zhang, Y. Liu, S. Ma. Economic Recommendation with Surplus Maximization. WWW 2016.

Why Fairness in RecSys? Resources Could be Limited

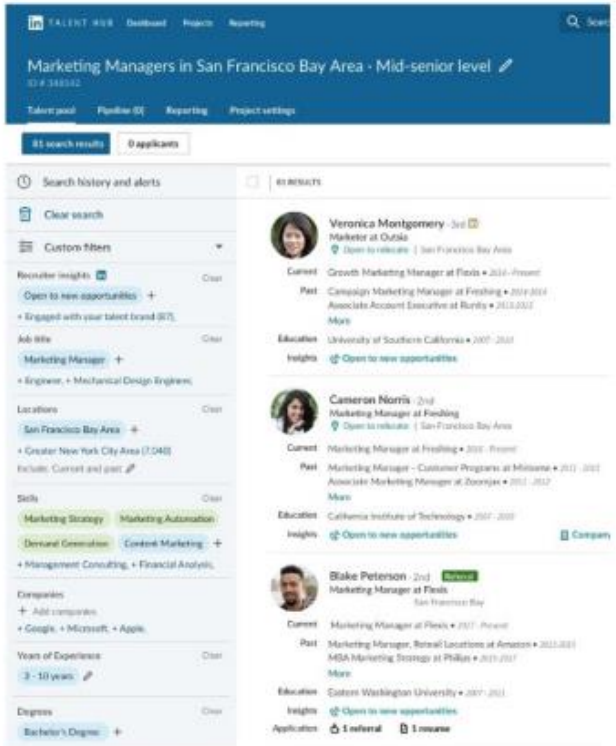


Recommendation slot positions are limited, which producers' items should be recommended and get the exposure opportunity to users?

User attention is a limited resource, whose twite should get exposure on the timeline?



Passengers are limited, which driver should get the task and make money?

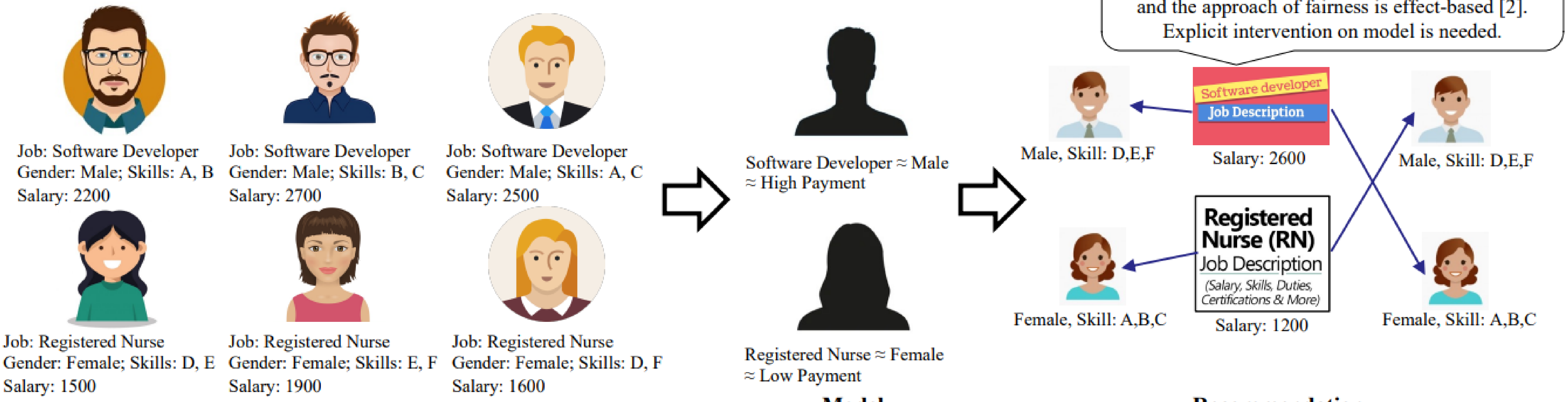


Interview opportunities are limited, which candidate(s) should get an interview opportunity?

Why Fairness in RecSys? Data Could be Biased

- Most RecSys models are ML models trained on some training data
  - Training data may encode social bias
  - Recommendation models may learn "shotcuts" for decision making
  - Model may echo or even reinforce the bias in training data

Just data debias is not enough because AI doesn't know which are sensitive features (e.g., gender) and the approach of fairness is effect-based [2]. Explicit intervention on model is needed.



Training Data
Strong correlationn between Job, Gender and Salary Level, while the skill feature shows less consistency among samples

Model
Model learns this strong correlation

Recommendation
Model echos/reinforces such correlation, the influence of skills is weakened by the strong data bias.

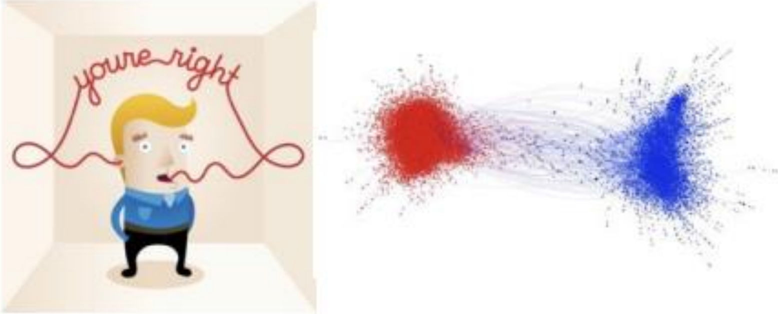
Potential Consequences of Unfairness in RecSys

Information Asymmetry

Knowing a piece of valuable information (e.g., a job opportunity) could change one's life

Matthew Effect

Advantaged users, items, or groups get further propagated by recommendations, sometimes not because of their good quality but because the recommendation model is dominated by their data

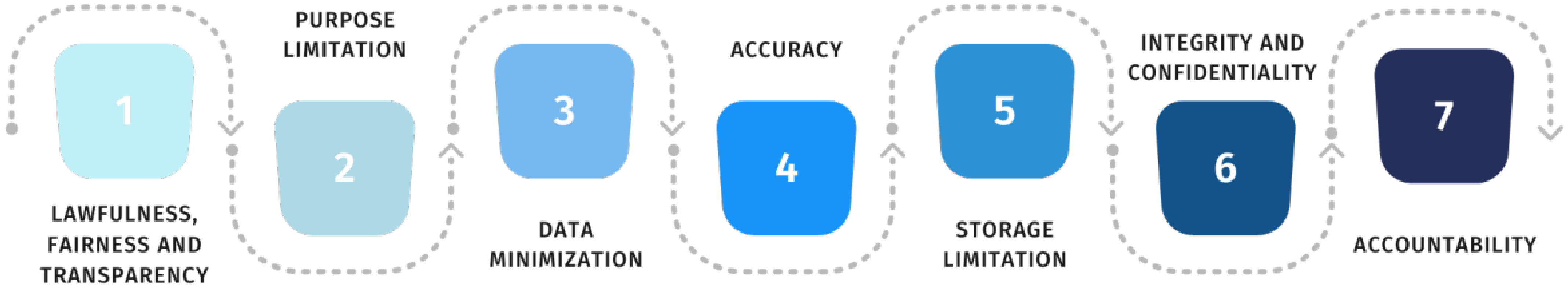


Echo Chambers

Unfair, undiversified exposure of news, messages, tweets, etc. may create echo chamber. Makes it difficult to explore new ideas and opinions different from one's own. Makes people feel like the whole world thinks the same way as they think. May even reinforce someone's extremist ideas

Fairness in RecSys: an AI Ethics Perspective

- Recommender systems as responsible AI
  - Provide fair decisions for users, item providers, and platform

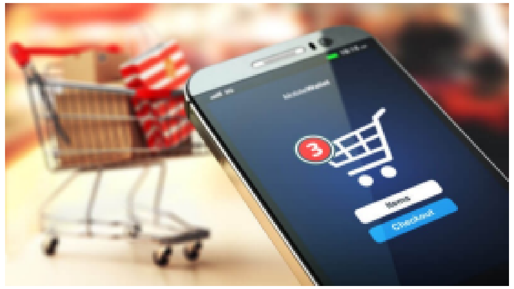


7 Principles of EU GDPR Regulation

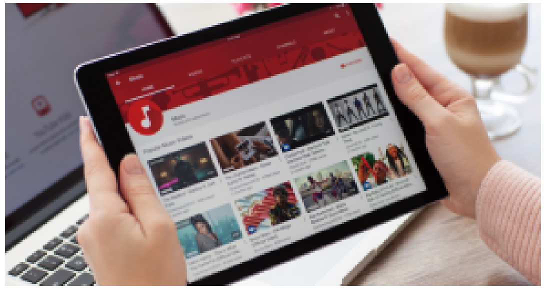
- Fairness often appears together with other responsible AI perspectives
  - e.g., transparency/explainability (honesty) of algorithmic decisions is the foundation of fairness

Fairness in RecSys: Beyond Ethics, a Utilitarian Perspective

- RecSys platforms should consider fairness for the sake of themselves
  - Not only for legal regulations, but for the sustainable/long-term development of the platform



An e-commerce example
Big retailors vs. Small retailors



A social network example
Star accounts vs. Grassroot accounts

If products from small retailors (e.g., family workshops) do not have fair exposure opportunity by e-commerce recommender system, they may eventually leave since they cannot survive in the platform, making the platform unsustainable.

Videos from famous accounts (e.g., a film star) usually get more attention, but if videos created by grassroot accounts do not have any exposure opportunity to users, they may leave the platform, making the platform's contents less diversified and even boring.

Fairness in RecSys

- User-wise fairness
  - Does the system work for me as good as for others?
- Fairness w.r.t. (sensitive) user groups
  - Are some groups being discriminated?
- Item-wise / content provider-wise fairness
- Multi-objective optimization
  - a.k.a. fairness for multiple metrics

RUTGERS

What exactly is Fairness in RecSys?

Many different perspectives:

- Group Fairness vs. Individual Fairness
- User Fairness vs. Item Fairness
- Associative Fairness vs. Causal Fairness
- Single-sided Fairness vs. Multi-sided Fairness
- Static Fairness vs. Dynamic Fairness
- Short-term Fairness vs. Long-term Fairness
- Populational Fairness vs. Personalized Fairness

Fairness in General

- Equality of opportunities
  - „You should not be disqualified /mistreated based on generic statistics that should not affect the outcome“
    - „You will not get the job because you are female“
    - What about already biased inputs?
- Equality of outcome
  - „Submission vs. acceptance ratio for male/female authors should not differ, if they differ, countermeasures should be taken“
    - Is this still fair?
    - Someone may be in „higher need“ of getting help vs. Someone had been mistreated in the past.
- Fairness vs. proportionality

More on biases in RS soon...

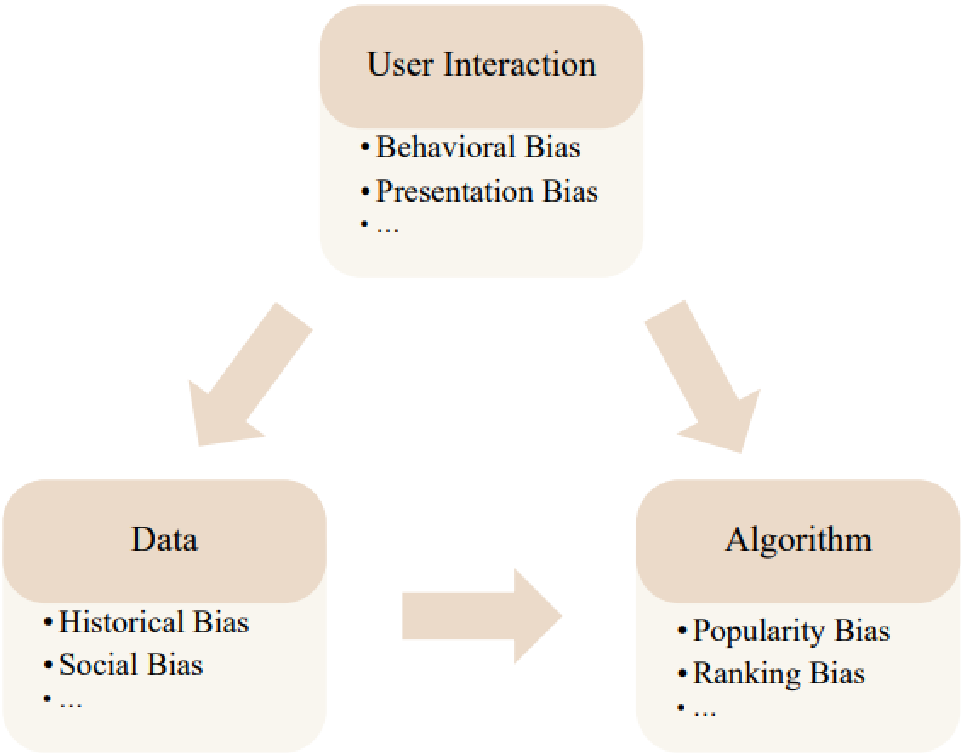
Fairness in Machine Learning — Causes

Data Bias

- Statistical Bias: non-random sample; record error
- Historical Bias: biased decision
- …

Algorithmic Bias

- Ranking Bias: exposure allocation
- Evaluation Bias: inappropriate benchmarks
- …



Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." *arXiv preprint arXiv:1908.09635* (2019).
Castelnovo, Alessandro, et al. "The zoo of Fairness metrics in Machine Learning." *arXiv preprint arXiv:2106.00467* (2021).

Fairness in Machine Learning — Methods

Pre-processing	In-processing	Post-processing
Try to transform the data so that the underlying discrimination is removed.	Try to modify the learning algorithms to remove discrimination during the model training process.	Perform after training by accessing a holdout set which was not involved during the training of the model.

Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." *arXiv preprint arXiv:1908.09635* (2019)

Fairness in Machine Learning — Evaluation

The evaluation usually depends on the requirement of fairness.

Statistical parity

- **Disparate Impact**: $P(\hat{y} = 1|z = 0) = P(\hat{y} = 1|z = 1)$
  - Evaluation: $DI = \left|P(\hat{y} = 1|z = 0) - P(\hat{y} = 1|z = 1)\right|$
- **False Positive Rate**: $P(\hat{y} \neq y|y = -1, z = 0) = P(\hat{y} \neq y|y = -1, z = 1)$
  - Evaluation: $DM_{FPR} = P(\hat{y} \neq y|z = 0, y = -1) - P(\hat{y} \neq y|z = 1, y = -1)$
- **False Negative Rate**: $P(\hat{y} \neq y|y = 1, z = 0) = P(\hat{y} \neq y|y = 1, z = 1)$
  - Evaluation: $DM_{FNR} = P(\hat{y} \neq y|z = 0, y = 1) - P(\hat{y} \neq y|z = 1, y = 1)$

Fairness in Machine Learning – Basic tasks

Fairness in Classification

Fairness in Ranking

Fairness in Classification — Introduction

Objective: Avoid unethical interference of protected attributes into the decision-making process.

Binary Classification: Fairness metrics can be expressed by **rate constraints** to regularize the classifier's positive or negative rates over different protected groups.

– Statistical parity:

$$P(\hat{Y} = 1 | Z = 0) = P(\hat{Y} = 1 | Z = 1)$$

– Equality of Opportunity:

$$P(\hat{Y} = 1 | Z = 0, Y = 1) = P(\hat{Y} = 1 | Z = 1, Y = 1)$$

…

Fairness in Classification – Method

Pre-processing: [3][4][5][6]…

Pros:

The transformed dataset can be used to train any downstream algorithm.

Cons:

Unpredictable loss in accuracy;

May not remove unfairness on the test data.

In-processing: [7][8][9][10]…

Pros:

Good performance;

May higher flexibility for the trade-off.

Cons:

A non-convex optimization problem and not guarantee optimality.

Post-processing: [11][12][13]…

Pros:

No need to modify classifier;

Relatively good performance especially fairness measures.

Cons:

Cannot be used in cases where sensitive feature information is unavailable.

Fairness in Classification

- **Method:**

$$\begin{array}{ll} \text{minimize} & L(\boldsymbol{\theta}) \quad\quad\quad\quad\quad\quad\quad\quad \} \text{ Classifier loss function} \\ \text{subject to} & P(.|z=0) = P(.|z=1) \ \} \text{ Fairness constraints} \end{array}$$

- No disparate impact: $P(\hat{y}=1|z=0) = P(\hat{y}=1|z=1)$

$$\text{Cov}_{DI}(z, d_{\boldsymbol{\theta}}(\boldsymbol{x})) = \mathbb{E}[(z-\bar{z})d_{\boldsymbol{\theta}}(\boldsymbol{x})] - \mathbb{E}[(z-\bar{z})]\bar{d}_{\boldsymbol{\theta}}(\boldsymbol{x}) \approx \frac{1}{N}\sum_{(\boldsymbol{x},z)\in\mathcal{D}}(z-\bar{z})\,d_{\boldsymbol{\theta}}(\boldsymbol{x})$$

- Objective function for no disparate impact:

$$\begin{array}{ll} \text{minimize} & L(\boldsymbol{\theta}) \\ \text{subject to} & \frac{1}{N}\sum_{(\boldsymbol{x},z)\in\mathcal{D}}(z-\bar{z})\,d_{\boldsymbol{\theta}}(\boldsymbol{x}) \le c \\ & \frac{1}{N}\sum_{(\boldsymbol{x},z)\in\mathcal{D}}(z-\bar{z})\,d_{\boldsymbol{\theta}}(\boldsymbol{x}) \ge -c \end{array}$$

“Zafar, Muhammad Bilal, et al. "Fairness Constraints: A Flexible Approach for Fair Classification." *J. Mach. Learn. Res.* 20.75 (2019): 1-42

Fairness in Ranking – Introduction

List-wise definitions for fairness: depend on the entire list of results for a given query

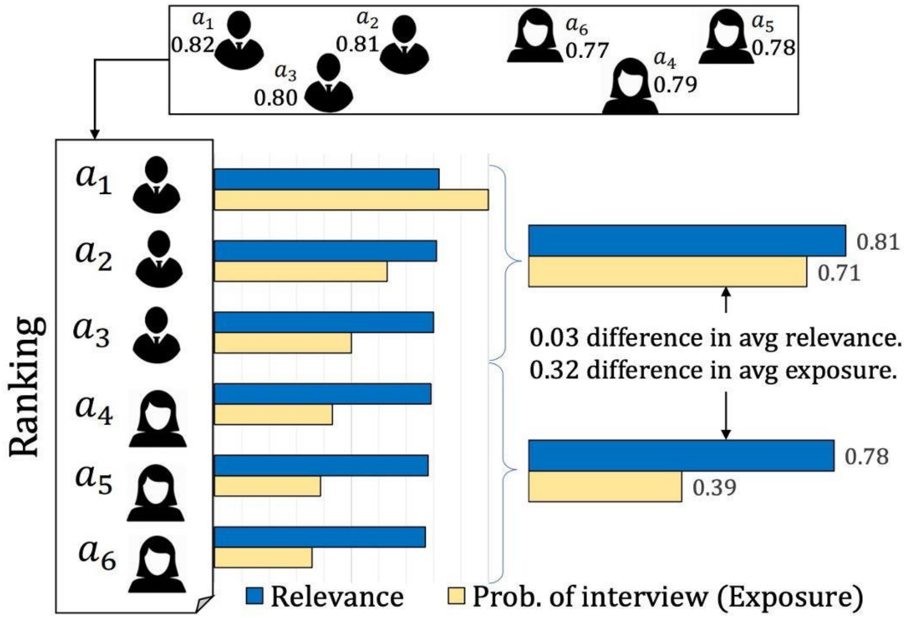
Unsupervised criteria: the average **exposure** near the top of the ranked list to be **equal for different groups** [71][72][75]

Supervised criteria: the average **exposure** for a group to be proportional to the average **relevance** of that group's results to the query [65][67]

Fairness in Ranking

- **Fairness Concerns**: A conceptual and computational framework that allows the formulation of fairness constraints on rankings in terms of **exposure allocation.**

- Job seeker example: a small difference in **relevance** can lead to a large difference in **exposure** (an opportunity) for the group of females.



Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD*'2018.

Fairness in Ranking

- **Method:** $r = \text{argmax}_r\, \text{U}(r|q)$ s.t. r is fair

- **Exposure** for a document d_i under a probabilistic ranking P as:

$$\text{Exposure}(d_i|\mathbf{P}) = \sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j \qquad \text{Exposure}(G_k|\mathbf{P}) = \frac{1}{|G_k|}\sum_{d_i \in G_k} \text{Exposure}(d_i|\mathbf{P})$$

- **Demographic Parity Constraints**:

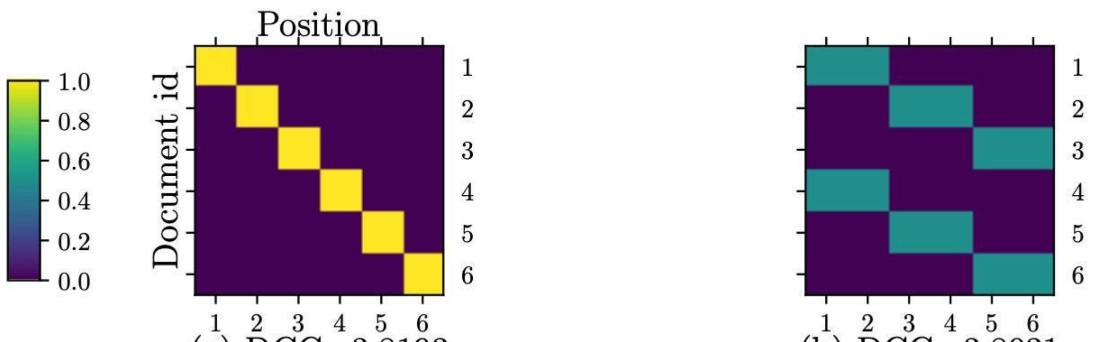
$$\text{Exposure}(G_0|\mathbf{P}) = \text{Exposure}(G_1|\mathbf{P}) \Leftrightarrow \mathbf{f}^T P\mathbf{v} = 0$$

$$\left(\text{with } \mathbf{f}_i = \frac{\mathbb{1}_{d_i \in G_0}}{|G_0|} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|}\right)$$

Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD*'2018

Fairness in Ranking

- Figure (a) is optimal unfair ranking that maximizes DCG.
- Figure (b) is optimal fair ranking under demographic parity.
- Compared to the DCG of the unfair ranking, the optimal fair ranking has slightly **lower utility** with a DCG.



(a) DCG=3.8193

(b) DCG=3.8031

Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD*'2018

Fairness in Recommendation – Challenges

More Perspectives

Multiple Models And Goals

Extreme Data Sparsity

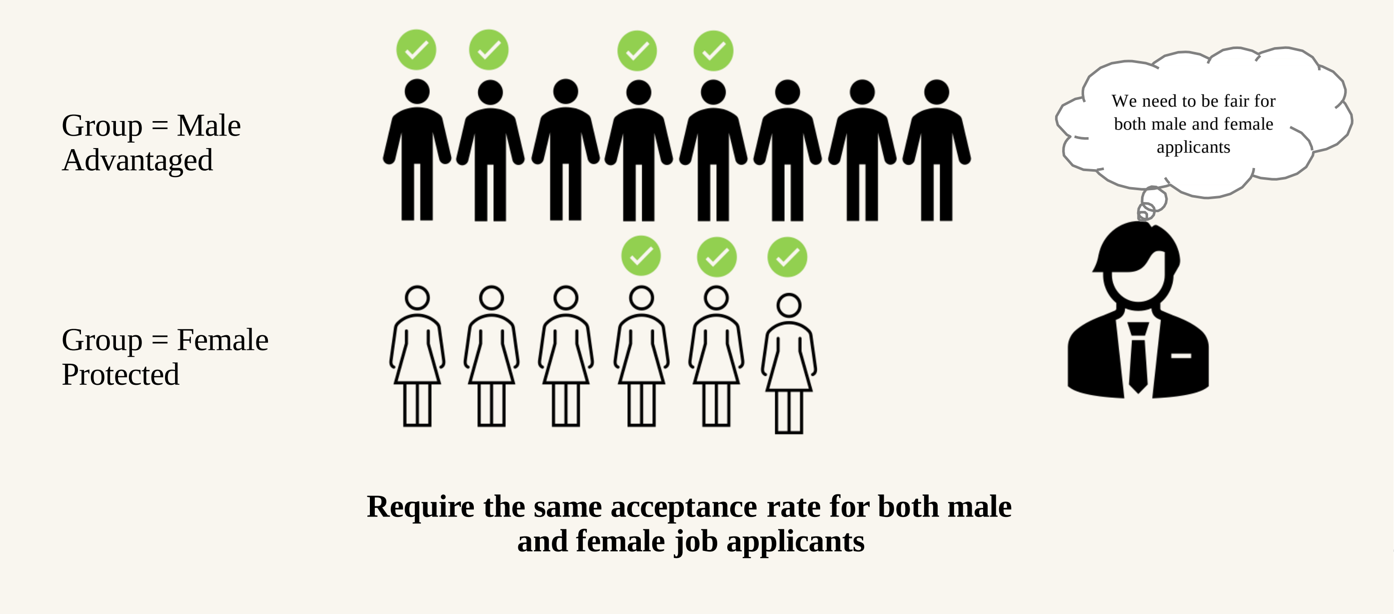
Dynamics

Taxonomies

- Group vs. Individual
- User vs. Item
- Association vs. Causality
- Single-sided vs. Multi-sided
- Static vs. Dynamic

Group Fairness vs. Individual Fairness

Group fairness requires that the protected groups should be treated similarly to the advantaged group.



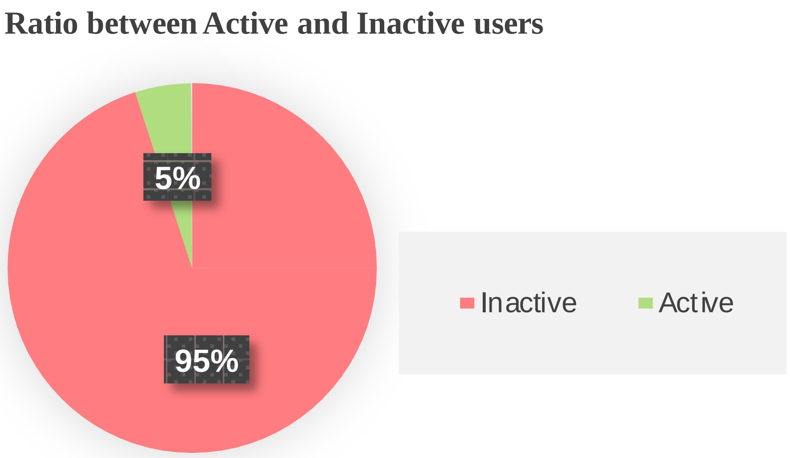
Group Fairness vs. **Individual Fairness**

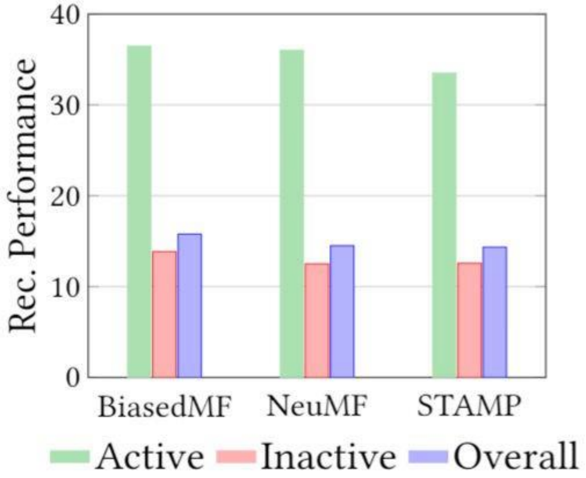
- Individual fairness requires that the similar individual should be treated similarly.

Image source: https://mitibmwatsonailab.mit.edu/wp-content/uploads/2020/04/Leadspace-GettyImages-598952582.jpg

Group Fairness in Recommendation

- **Fairness concerns**: The **unfair recommendation quality** between **user groups** with different activity levels, e.g., number of interactions.
- Unfairness of current recommender systems:
  - Active users only account for a **small** proportion of users.
  - The average recommendation quality on the small group (*active*) is **significantly better** than that on the remaining majority of users (*inactive*) for all baselines.





Li.Y et al. "User-oriented Fairness in Recommendation" WWW'21.

Group Fairness in Recommendation

Fairness-aware Algorithm: A re-ranking method with user-oriented group fairness constrained on the recommendation lists generated from any base recommender algorithm.

$$\max_{\mathbf{W}_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{N} \mathbf{W}_{ij} \boxed{S_{i,j}}$$

Preference of user i in terms of item j

$$\text{s.t.} \quad \boxed{\mathrm{UGF}(Z_1, Z_2, \mathbf{W}) < \varepsilon}$$

Fairness constraint

$$\boxed{\sum_{j=1}^{N} \mathbf{W}_{ij} = K, \mathbf{W}_{ij} \in \{0, 1\}}$$

Top-K list

Experiment Results: Improve fairness; Improve recommendation quality of overall and disadvantaged users. However, the performance of advantaged users is reduced to satisfy our fairness requirement.

			Beauty			
			Overall	Adv.	Disadv.	UGF
BiasedMF	F1	Orig.	14.27	30.68	12.77	17.91
		Fair	**15.06**	19.18	**14.68**	**4.50**
	NDCG	Orig.	43.25	67.79	41.00	26.79
		Fair	**43.97**	52.51	**43.19**	**9.32**

Improvement of overall accuracy

Disadv. ⇑

Adv. ⇓

Improvement of fairness

Li.Y et al. "User-oriented Fairness in Recommendation" WWW'21.

Individual Fairness in Recommendation

- **Fairness concerns**: the position bias which leads to disproportionately less attention being paid to low-ranked subjects.

- No single ranking can achieve individual attention fairness.

- **Equity of Amortized Attention**: A sequence of rankings $\{1,2, \ldots m\}$ offer equity of amortized attention if each subject u receives cumulative attention proportional to her cumulative relevance:

$$\frac{\underbrace{\sum_{l=1}^{m} a_{i1}^{l}}_{\text{attention}}}{\underbrace{\sum_{l=1}^{m} r_{i1}^{l}}_{\text{relevance}}} = \frac{\sum_{l=1}^{m} a_{i2}^{l}}{\sum_{l=1}^{m} r_{i2}^{l}}, \forall u_{i1}, u_{i2}$$

Biega, A. J. et al. "Equity of Attention: Amortizing Individual Fairness in Rankings" SIGIR'18.

Individual Fairness in Recommendation

- **Method (Offline optimization):**

$$\text{minimize} \quad \boxed{\sum_i |A_i - R_i|}$$
$$\text{subject to} \quad \boxed{\mathit{NDCG\text{-}quality@k}(\rho^j, \rho^{j*}) \geq \theta, \ j = 1, \ldots, m.}$$

- Fairness
- Ranking quality

- **Experiment Results:**
  - **Improving equity of attention is crucial**: the discrepancy between the attention received and the deserved attention can be substantial.
  - Improving equity of attention can often be done **without sacrificing much quality** in the rankings.

Biega, A. J. et al. "Equity of Attention: Amortizing Individual Fairness in Rankings" SIGIR'18.

Associative Fairness vs. Causal Fairness

Find the **discrepancy of statistical metrics** between individuals or sub-populations.

In **binary classification**, fairness metrics can be represented by regularizing the classifier's positive or negative rates over different protected groups.

Associative Fairness vs. **Causal Fairness**

- Fairness cannot be well assessed only based on association notions [46-49].

- Difference:
  - Reason about the **causal relations** between the protected features and the model outcomes.
  - Leverage **prior knowledge** about the world structure in the form of causal models, help to understand the propagation of variable changes in the system.

Causal Fairness

- **Disparate Impact**:
  - **Total Effect**: $TE_{a_1,a_0}(y) = P(y_{a_1}) - P(y_{a_0})$
  - **Effect of Treatment on the Treated**: $ETT_{a_1,a_0}(y) = P(y_{a_1} \mid a_0) - P(y \mid a_0)$
  - …
- **Disparate Treatment**:
  - **Direct Effect**: the causal effect along the causal path from the sensitive feature to the final decision
  - **Indirect Effect**: the causal effect along the causal path through proxy features
  - **Path-Specific Effect**: the causal effect over specific paths.

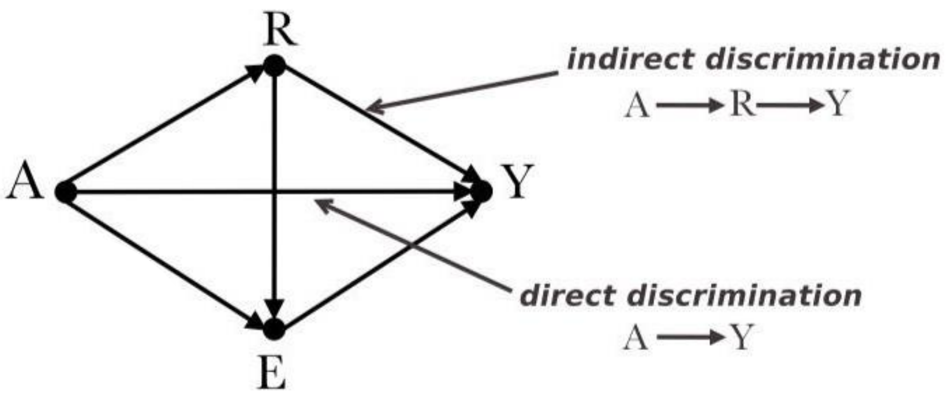
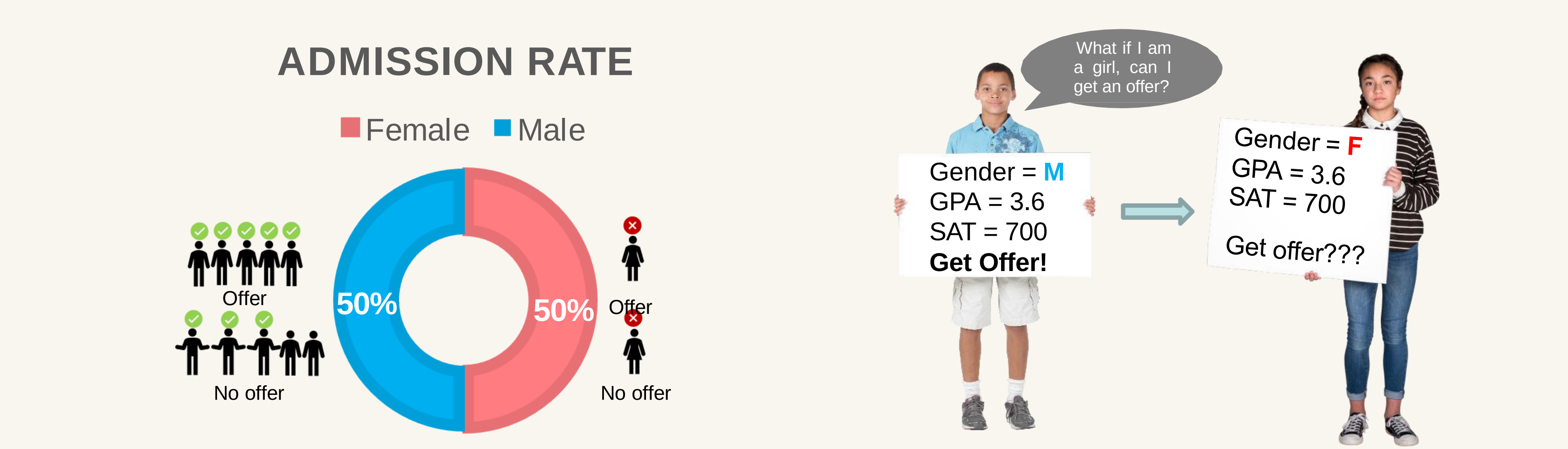


Figure Source: Makhlouf, Karima, et al. "Survey on Causal-based Machine Learning Fairness Notions." *arXiv preprint arXiv:2010.09553* (2020).

Counterfactual fairness

- Counterfactual fairness is an individual-level causal-based fairness notion. It requires that for any possible individual, the predicted result of the learning system should be the **same** in th**e counterfactual world** as in the **real world.**



Associative Group Fairness

Counterfactual Fairness

Associative Fairness in Recommendation

- **Method**:

$$\min_{\boldsymbol{P},\boldsymbol{Q},\boldsymbol{u},\boldsymbol{v}} \underbrace{\boxed{J(\boldsymbol{P},\boldsymbol{Q},\boldsymbol{u},\boldsymbol{v})}}_{\text{Loss for recommender model}} + \underbrace{\boxed{U}}_{\text{Fairness constraint}}$$

- **Experiment Results**: the experiments on synthetic and real data show that minimization of these forms of unfairness is possible with no significant increase in reconstruction error.

Unfairness	Error	Value	Absolute	Underestimation	Overestimation	Non-Parity
None	0.887 ± 1.9e-03	0.234 ± 6.3e-03	0.126 ± 1.7e-03	0.107 ± 1.6e-03	0.153 ± 3.9e-03	0.036 ± 1.3e-03
Value	0.886 ± 2.2e-03	**0.223 ± 6.9e-03**	0.128 ± 2.2e-03	**0.102 ± 1.9e-03**	**0.148 ± 4.9e-03**	0.041 ± 1.6e-03
Absolute	0.887 ± 2.0e-03	0.235 ± 6.2e-03	**0.124 ± 1.7e-03**	0.110 ± 1.8e-03	0.151 ± 4.2e-03	0.023 ± 2.7e-03
Under	0.888 ± 2.2e-03	0.233 ± 6.8e-03	0.128 ± 1.8e-03	**0.102 ± 1.7e-03**	0.156 ± 4.2e-03	0.058 ± 9.3e-04
Over	**0.885 ± 1.9e-03**	0.234 ± 5.8e-03	**0.125 ± 1.6e-03**	0.112 ± 1.9e-03	**0.148 ± 4.1e-03**	0.015 ± 2.0e-03
Non-Parity	0.887 ± 1.9e-03	0.236 ± 6.0e-03	0.126 ± 1.6e-03	0.110 ± 1.7e-03	0.152 ± 3.9e-03	**0.010 ± 1.5e-03**

Yao, Sirui, and Bert Huang. “Beyond Parity: Fairness Objectives for Collaborative Filtering” NIPS’17

Causal Fairness in Recommendation

- **Fairness Concerns**: Counterfactual fairness for users in recommendations.

- **Definition:** A recommender model is *counterfactually fair* if for any possible user u with features $X = x$ and $Z = z$, for all L, and for any value z' attainable by Z:

$$P\left(\boxed{L_{\boldsymbol{z}}} \mid \boxed{\boldsymbol{X}} = \boldsymbol{x}, \boxed{\boldsymbol{Z}} = \boldsymbol{z}\right) = P\left(L_{\boldsymbol{z}'} \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}\right)$$

Top-N recommendation list for user u with sensitive features z

Insensitive features

Sensitive features

Li. Y et al. "Towards Personalized Fairness based on Causal Notion" SIGIR'21

Fairness in RS, further reading

- https://link.springer.com/article/10.1007/s11257-020-09285-1
- https://dl.acm.org/doi/pdf/10.1145/3383313.3411545
- https://www.sciencedirect.com/science/article/pii/S0306457321001503
- https://arxiv.org/abs/1908.06708
- https://dl.acm.org/doi/pdf/10.1145/3450614.3461685
- https://arxiv.org/abs/2006.05255
- https://dl.acm.org/doi/pdf/10.1145/3184558.3186949