Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A) Winter 2023

Week #11- **Summary**

Announcement

- **Last Quiz:** on March 31, 2023
- **Project Presentation:** on April 14, 2023
- **Project Paper** due April 23, 2023
- **Final exam**: on April 30, 2023

Today

- Intro to Trustworthy Machine Learning + Crash course on Algorithmic Fairness
- Recommender Systems: case study
- Q&A

Trustworthy ML

ML is everywhere!

amazon

Google

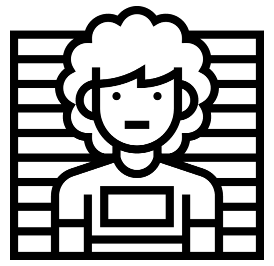
YouTube

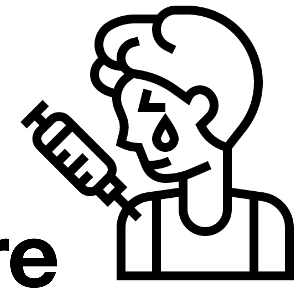
facebook

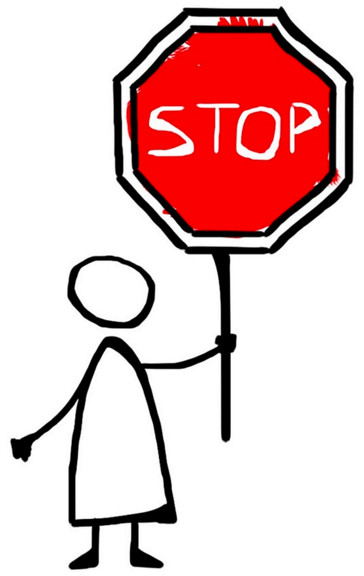
NETFLIX

Nowadays AI/ML algorithms determine

- **Who gets a job** 

- **Who goes to jail** 
- **Who receives loan from bank** 

- **Who gets diagnose and receive health care** 



Are these automated decision making systems trustworthy?

Trustworthy Machine Learning

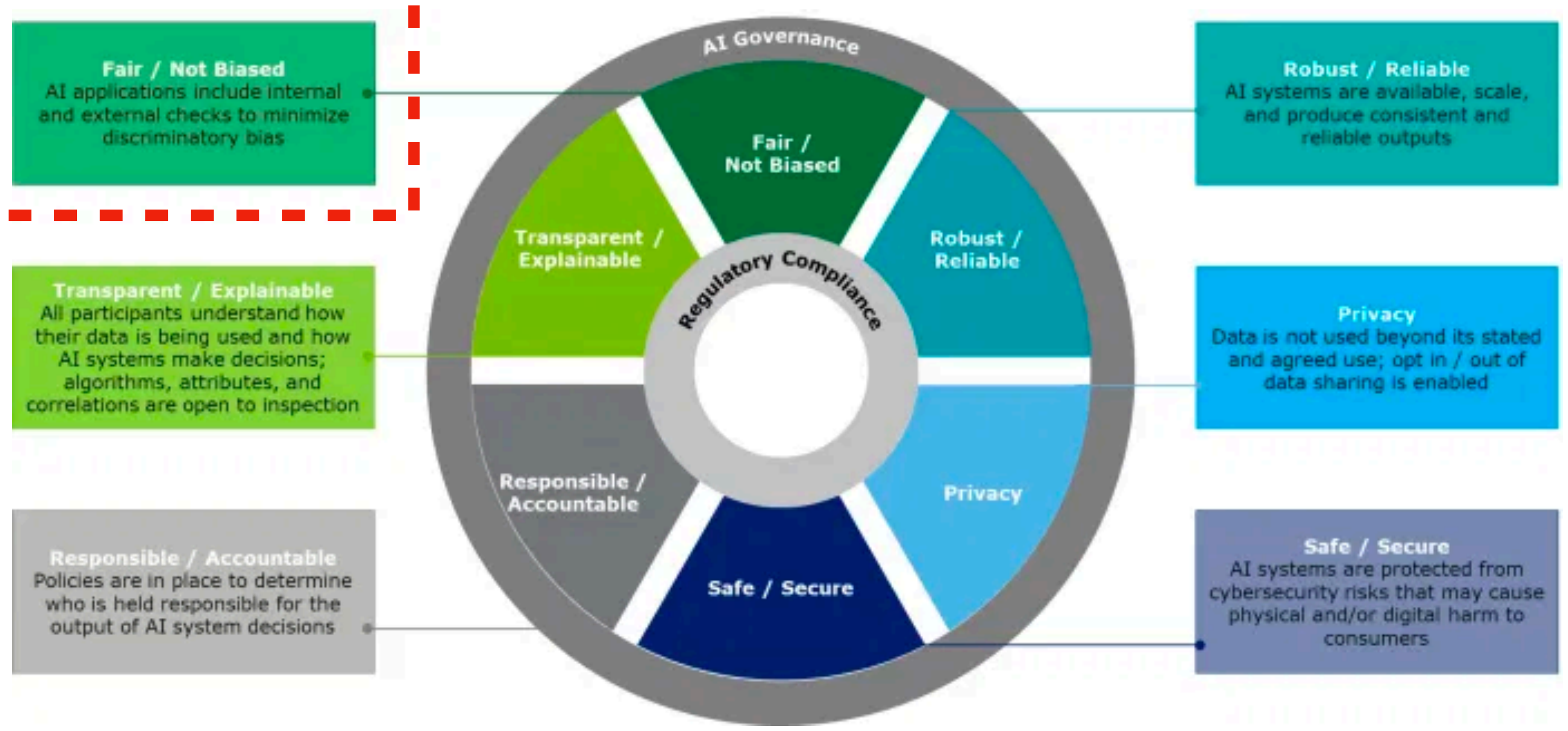
- Trustworthy ML is crucial to the widespread adoption of ML.

- Trustworthy ML is a framework to help address elements that ensure the ethical use of ML and sustain the trust of customers.

Fairness

Explainable

Accountability



Reliability/Robustness

Privacy

Safety

Source: Deloitte Consulting LLP

What is the Importance of Algorithmic Fairness?

Why Does Algorithmic Discrimination matter?

[Barocas, S., et al. ,2017]

- **Allocation harm**: E.g., Amazon Hiring system, COMPAS risk assessment
- **Quality of service harm**: E.g., gender shades, VMS make women sick
- **Stereotyping harm**, e.g., Black criminality in predictive policing, gender issues in NLP (in translation)
- **Denigration harm**, e.g., mislabeling images of Black women as Gorillas, Chatbot Tay for hate speech
- **Over and under-representation harm**, e.g., images of men in image search results

Gender-shades

- Let's hear about if from Joy Buolamwini!

http://gendershades.org/

Laws against Discrimination

Legally recognized 'protected classes'

Race (Civil Rights Act of 1964)
Color (Civil Rights Act of 1964)
Sex (Equal Pay Act of 1963; Civil Rights Act of 1964)
Religion (Civil Rights Act of 1964)
National origin (Civil Rights Act of 1964)
Citizenship (Immigration Reform and Control Act)
Age (Age Discrimination in Employment Act of 1967)
Pregnancy (Pregnancy Discrimination Act)
Familial status (Civil Rights Act of 1968)
Disability status (Rehabilitation Act of 1973; Americans with Disabilities Act of 1990)
Veteran status (Vietnam Era Veterans' Readjustment Assistance Act of 1974; Uniformed Services Employment and Reemployment Rights Act); **Genetic information** (Genetic Information Nondiscrimination Act)

sensitive attributes

Regulated domains

Credit (Equal Credit Opportunity Act)
Education (Civil Rights Act of 1964; Education Amendments of 1972)
Employment (Civil Rights Act of 1964)
Housing (Fair Housing Act)
Public Accommodation (Civil Rights Act of 1964)
Extends to marketing and advertising; not limited to final decision
This list sets aside complex web of laws that regulates the government

Article 14. Equality before law. -The State shall not deny to any person equality before the law or the equal protection of the laws within the territory of India. (1) **The State shall not discriminate against any citizen on grounds only of religion, race, caste, sex, place of birth or any of them**.

EU Charter of Fundamental Rights

1. Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited. 2.

https://www.refworld.org/pdfid/4d886bf02.pdf

Canadians have the right to be treated fairly in workplaces free from discrimination, and our country has laws and programs to protect this right. **The Canadian Human Rights Act** is a broad-reaching piece of legislation that prohibits discrimination on the basis of gender, race, ethnicity and other grounds. May 30, 2022

https://laws-lois.justice.gc.ca/eng/acts/h-6/fulltext.html

The basis for progressively redressing these conditions lies in the Constitution which, amongst others, upholds the values of human dignity, equality, freedom and social justice in a united, non-racial and non-sexist society where all may flourish;

South Africa also has international obligations under binding treaties and customary international law in the field of human rights which promote equality and prohibit unfair discrimination. Among these obligations are those specified in the Convention on the Elimination of All Forms of Discrimination Against Women and the Convention on the Elimination of All Forms of Racial Discrimination;

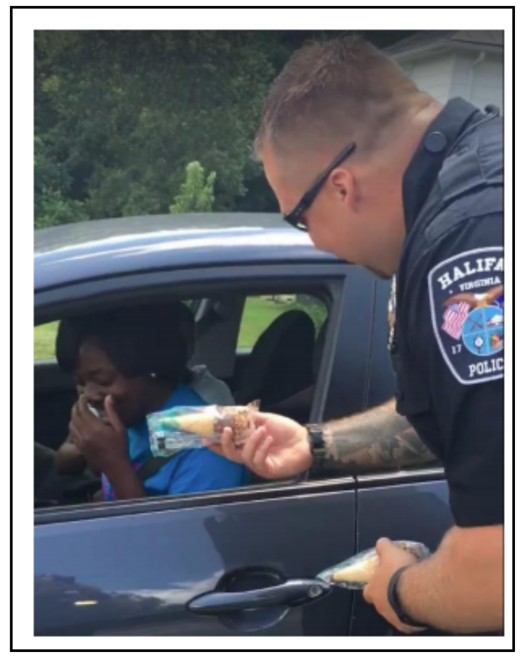
Initiating an Anti-Discrimination Regime in China

The 1982 Constitution has enshrined the principle of equality of all citizens before the law (Article 33). Articles 4, 36, 48, and 89 also guarantee the rights of ethnic minorities, religious freedom and gender equality and prohibits discrimination on those grounds.

Fairness in ML

2014

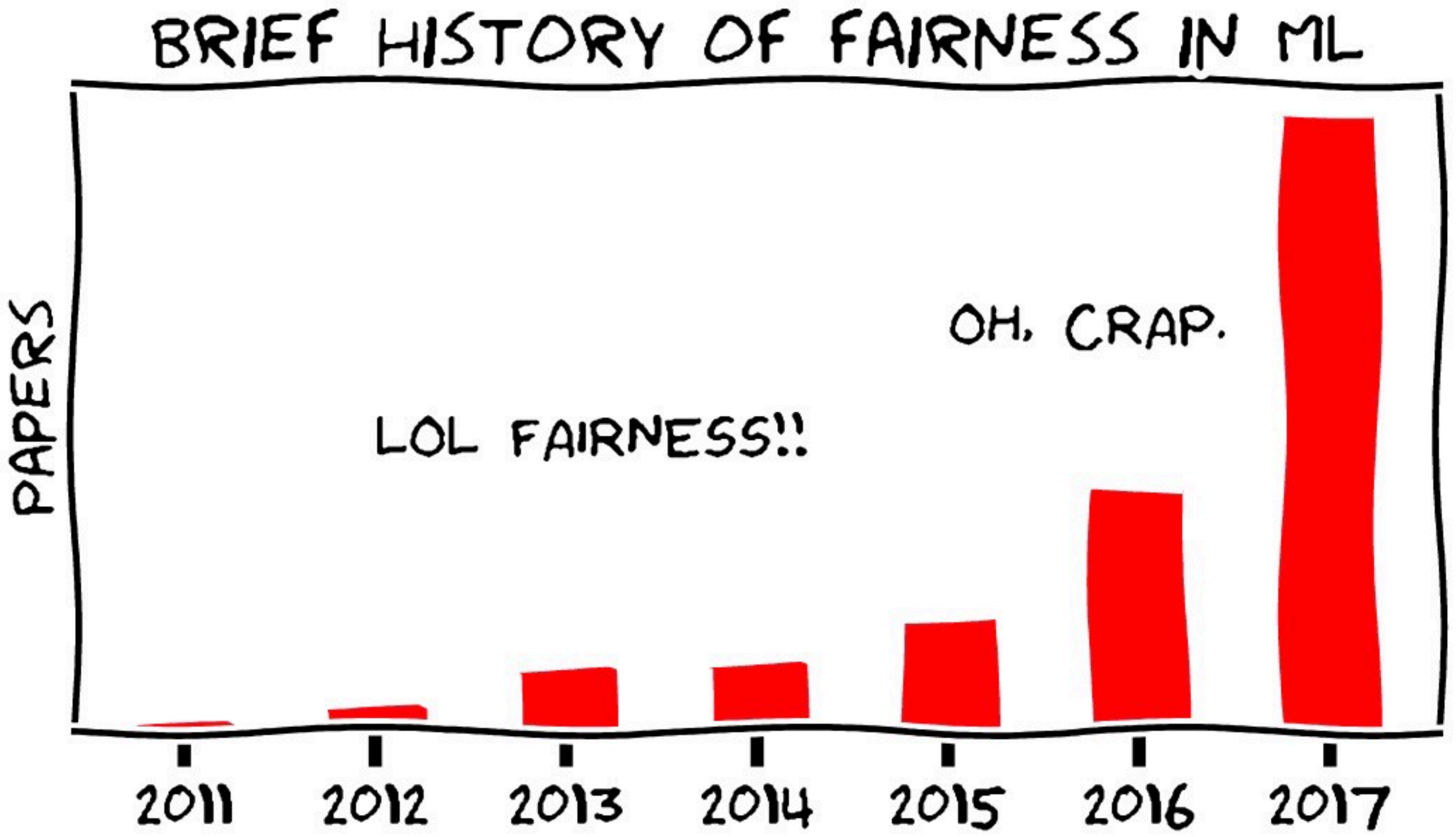
2015



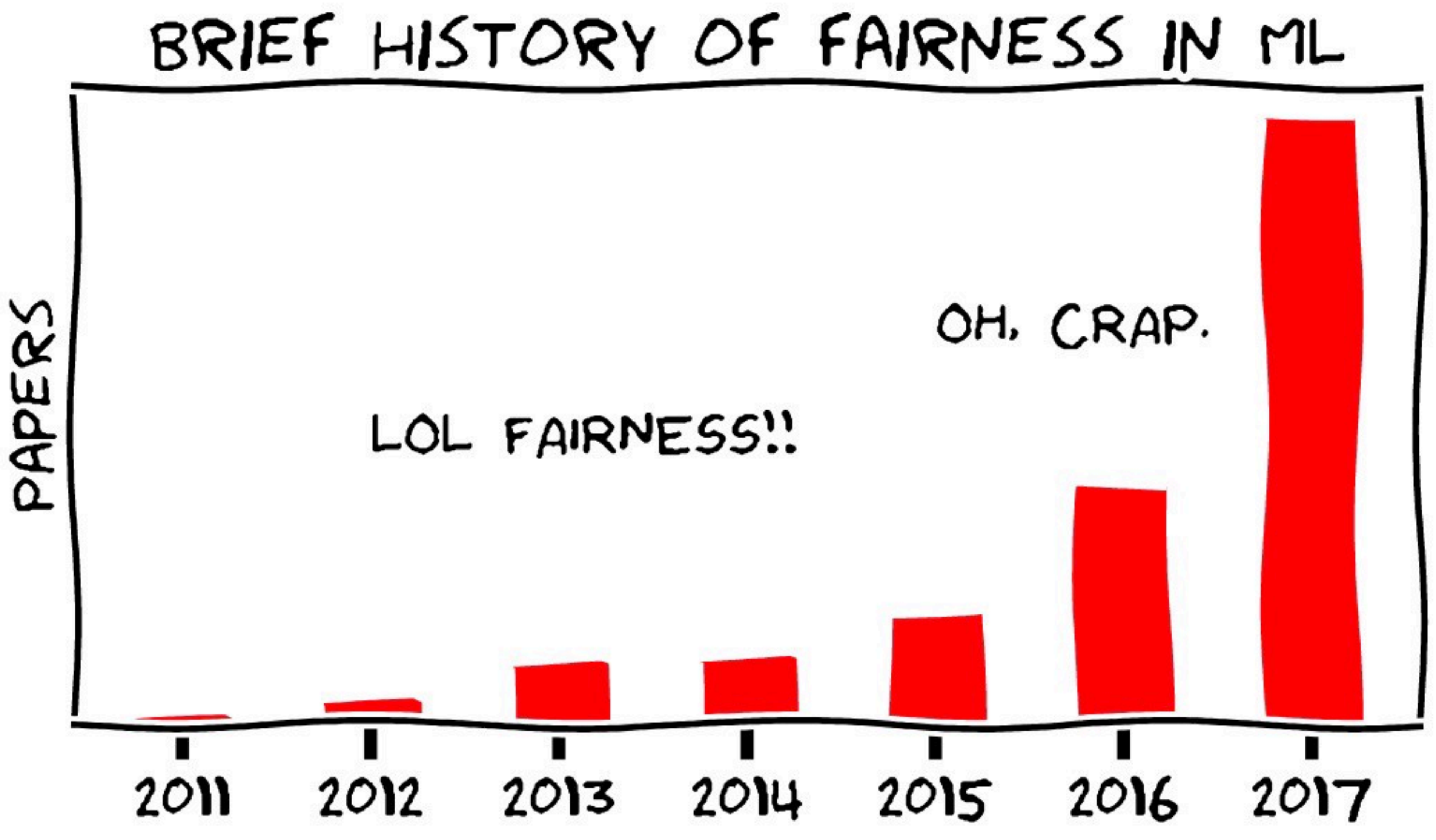
2016

2017

...



Fairness in ML



- “What is fair have been introduced in multiple disciplines for well over 50 years, including in education, hiring, and machine learning” [1].
- Statistics, Social Science, Economics, etc.

[1] Hutchinson, Ben, and Margaret Mitchell. "50 Years of Test (Un) fairness: Lessons for Machine Learning."

arXiv preprint arXiv:1811.10104 (2018).

What is the Importance of Algorithmic Fairness?

How to Define Fairness in Machine Learning?

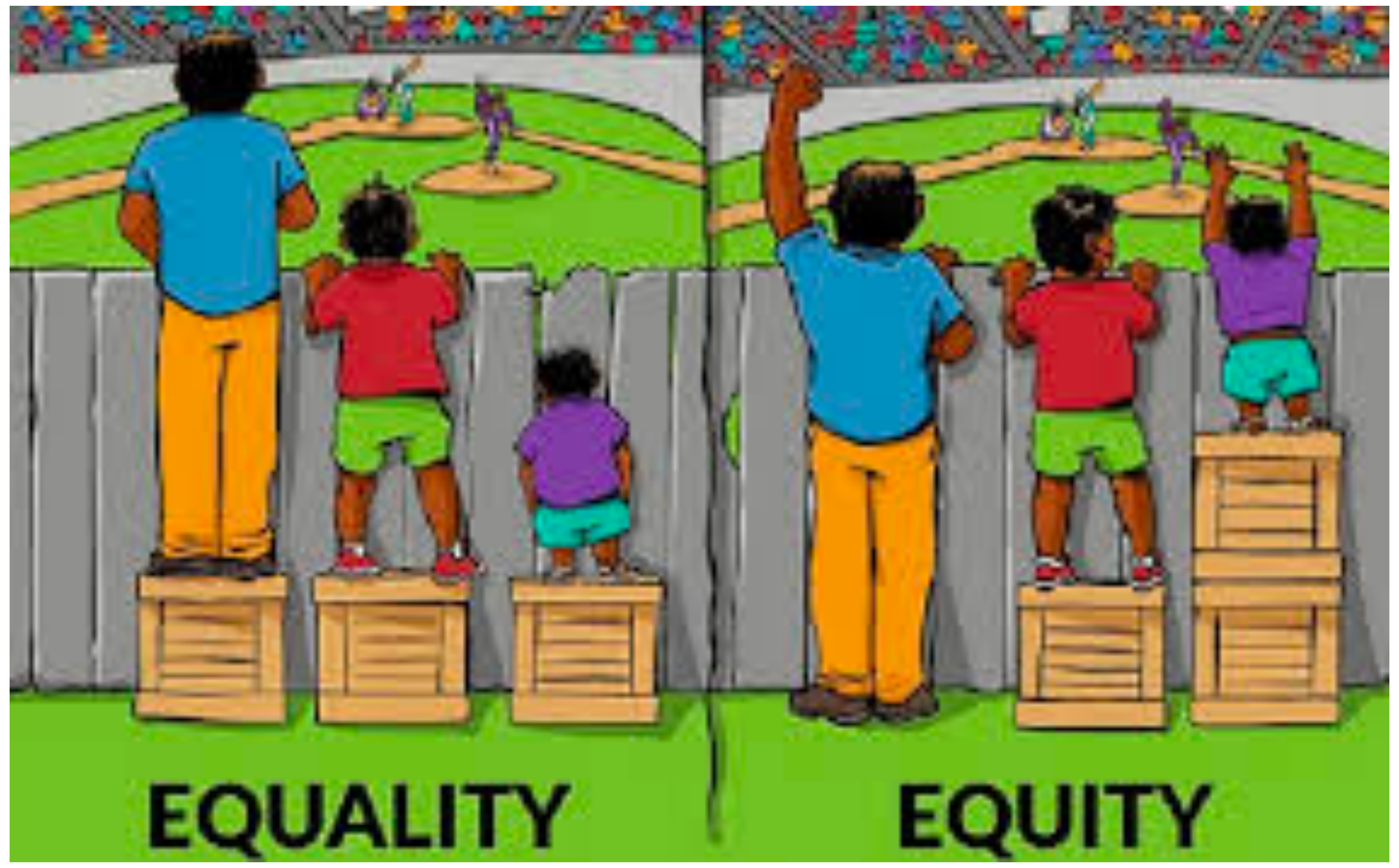
Why do we use fairness definitions?

- To make algorithmic systems support human
- To identify strengths and weakness of the sy
- To track improvement over time

To address Law Against Discrimination!

Why we have many notions of fairness?

By 2018, we had **21 definitions of fairness**!



An interesting tutorial by **Arvind Narayanan**:
Tutorial: 21 fairness definitions and their politics

Another interesting tutorial by **Jon Kleinberg**:
Inherent Trade-Offs in Algorithmic Fairness

Definition
Group fairness or statistical parity
Conditional statistical parity
Predictive parity
False positive error rate balance
False negative error rate balance
Equalised odds
Conditional use accuracy equality
Overall accuracy equality
Treatment equality
Test-fairness or calibration
Well calibration
Balance for positive class
Balance for negative class
Causal discrimination
Fairness through unawareness
Fairness through awareness
Counterfactual fairness
No unresolved discrimination
No proxy discrimination
Fair inference

Verma, Sahil, and Julia Rubin. "Fairness definitions explained." 2018 IEEE/ACM International Workshop on Software Fairness (FairWare). IEEE, 2018.

Why we don't have one definition?

- Correcting for **algorithmic bias** generally requires:
  - **knowledge** of how the measurement process is biased
  - **judgments** about properties to satisfy in an "unbiased" world

Hiring

Gender-biased

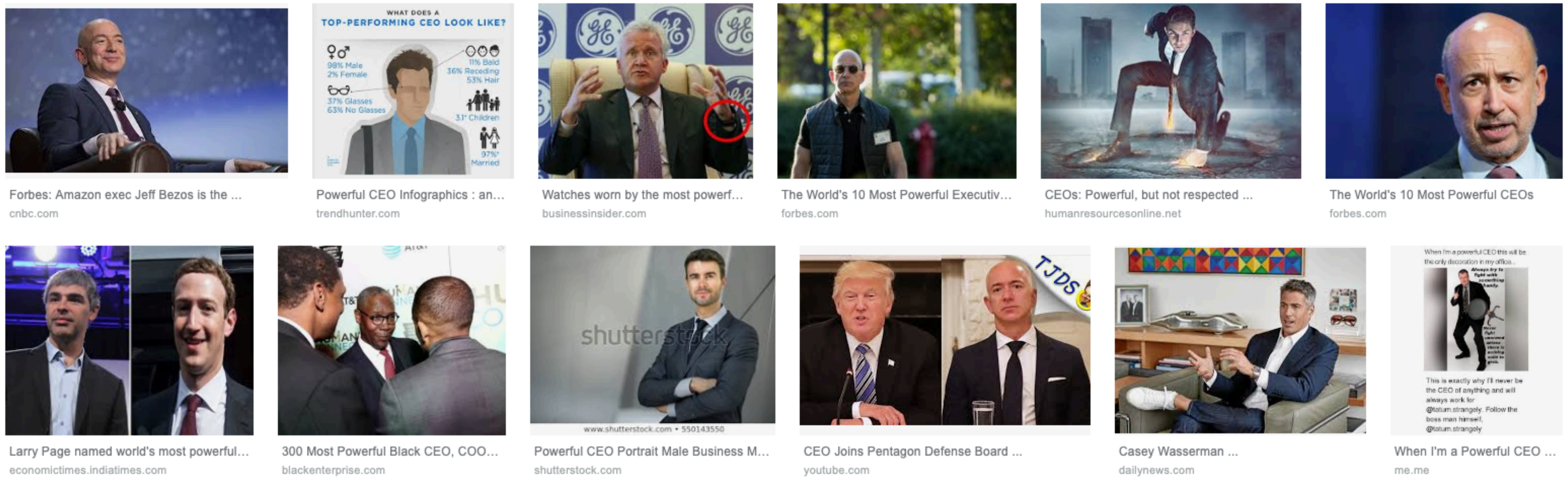
Medical diagnosis

Gender-biased

Fairness is not a general concept

Bias is **subjective** and must be considered **relative** to task

There is no agreed-upon measure

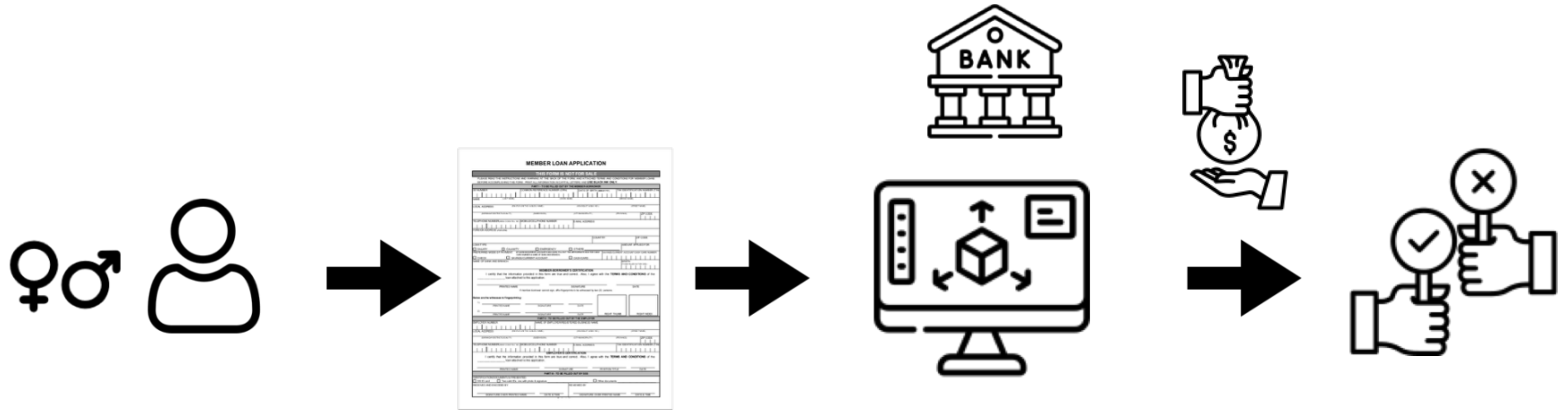


What is **fair?**

50% **female,** 50% **male?**
Based on the **population?**

Results for "CEO" in Google Images: 11% female, US
27% female CEOs

Running Example



Applicant

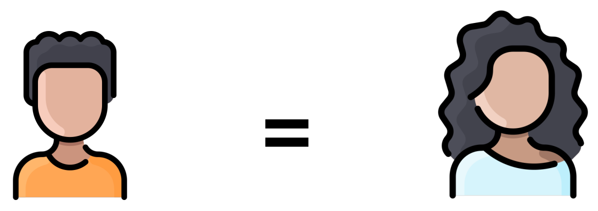
Loan Application

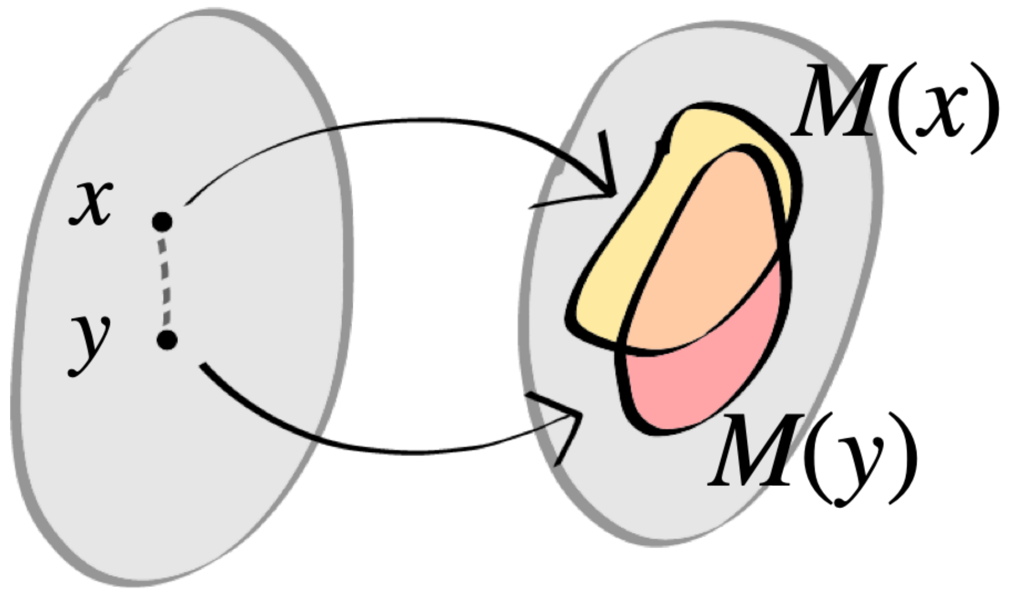
Loan Approval

S	X	$\hat{Y}$	Y
sensitive attribute	**non-sensitive attributes**	**Prediction decision**	**Actual Outcome**

How to Define Fairness in Machine Learning?

Individual Fairness: measures the impact that discrimination has on the individuals





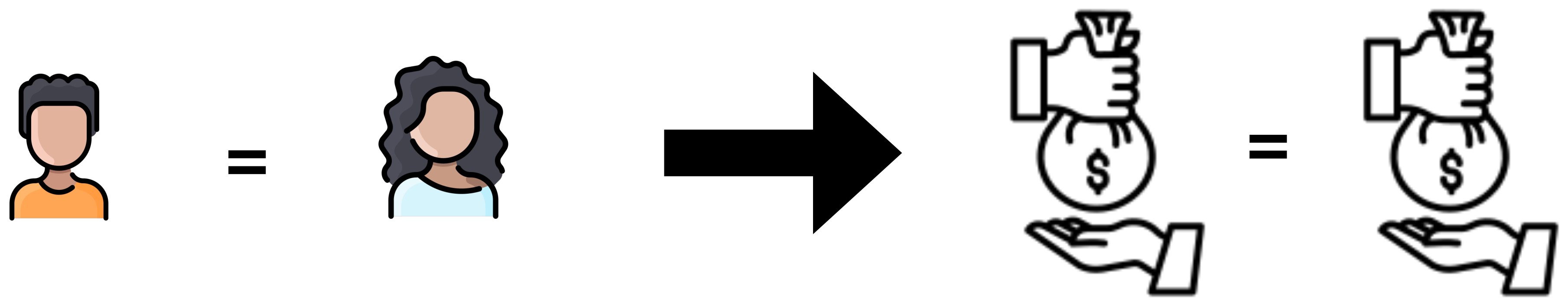
Lipschiz condition $||M(x) - M(y)|| \leq d(x, y)$

[Dwork, C., et al. ,2012]

The Lipschitz condition requires that any two individuals x, y that are at distance $d(x, y) \in [0, 1]$ map to distributions $M(x)$ and $M(y)$, respectively, such that the statistical distance between $M(x)$ and $M(y)$ is at most $d(x, y)$. In other words, the distributions over outcomes observed by x and y are indistinguishable up to their distance $d(x, y)$.

Individual Fairness Examples

Loan Approval



E.g., similar applicants with same application and only their gender differs, should have similar probability of receiving loan approval

Causal Discrimination: The **same decision** for any two subjects with the **exact same non-sensitive attributes.**

[Galhotra, S., et al. ,2017]

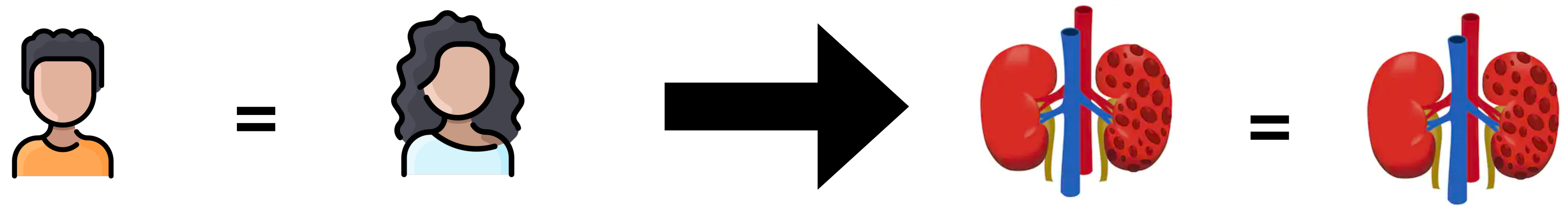
Kiarash Mohammadi (UdeM)

Aishwarya Sivaraman (UCLA)

Mohammadi, K., Sivaraman, A., & **Farnadi, G**. (2023). FETA: Fairness Enforced Verifying, Training, and Predicting Algorithms for Neural Networks. *arXiv preprint arXiv: 2206.00553*.

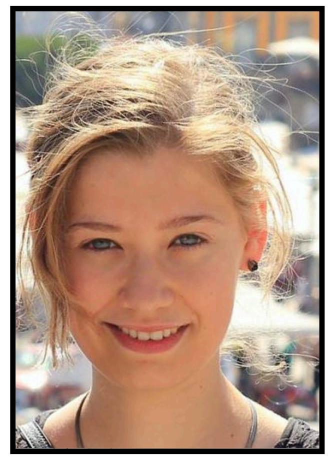
Individual Fairness Examples

Kidney Exchange Program



E.g., similar patients who are in a cycle to receive of a kidney transplant, should have similar chances of receiving a kidney transplant

William St-Arnaud (UdeM)



Margarida Carvalho (UdeM)

St-Arnaud, W.,Carvalho, M., and **Farnadi, G**. Adaptation, Comparison and Practical Implementation of Fairness Schemes in Kidney Exchange Programs. *arXiv preprint arXiv:2207.00241* (2023)

Farnadi, G., St-Arnaud, W., Babaki, B., & Carvalho, M. (2021, May). Individual Fairness in Kidney Exchange Programs. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 13, pp. 11496-11505).

Individual Fairness Examples

Salganik, R., Diaz, F., & **Farnadi, G**. (2022). Analyzing the Effect of Sampling in GNNs on Individual Fairness. In proceedings of FairRec Workshop at RecSsys 2022, arXiv preprint arXiv:2209.03904.

IVADO

Fernando Diaz (Google Research)

Rebecca Salganik (UdeM)

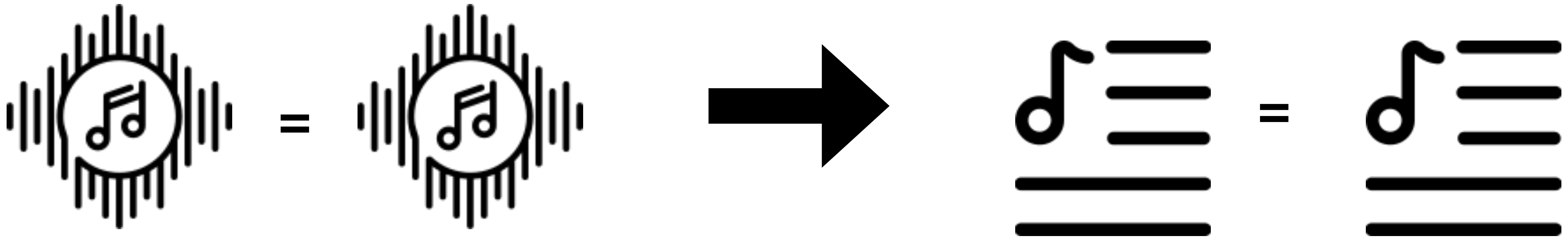
Ao Sun, J., Pentyala, S., De Cock, M., & **Farnadi, G**. (2023). Privacy-Preserving Fair Item Ranking. To appear in proceedings of the 45th European Conference on Information Retrieval (ECIR).

Jason Sun (UdeM)

Sikha Pentyala (UW)

Martine De Cock (UWT)

Recommender System



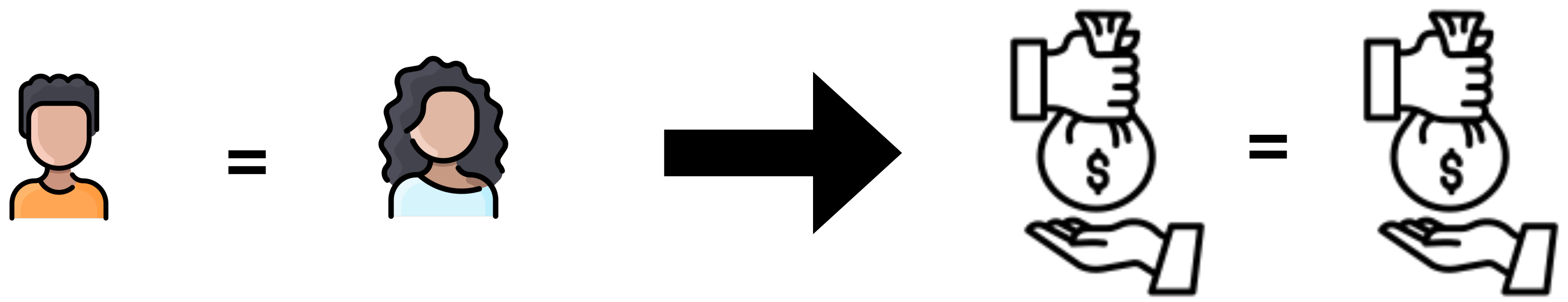
E.g., similar music items/artists with similar music features, should have similar probability of appearing in playlists

or

Similar Items should have similar chance of exposure in rankings

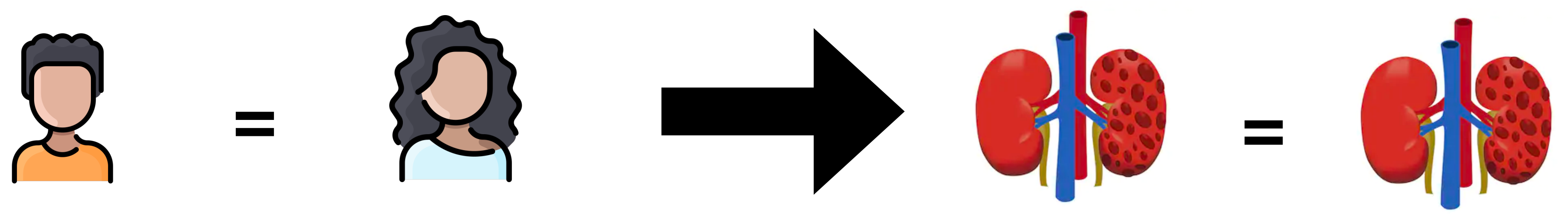
Individual Fairness Examples

Loan Approval



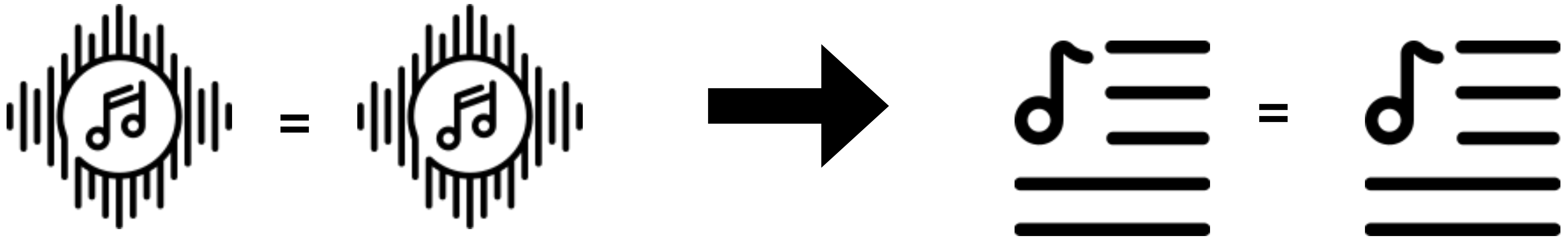
E.g., similar applicants with same application and only their gender differs, should have similar probability of receiving loan approval

Kidney Exchange Program



E.g., similar patients who are in a cycle to receive of a kidney transplant, should have similar chances of receiving a kidney transplant

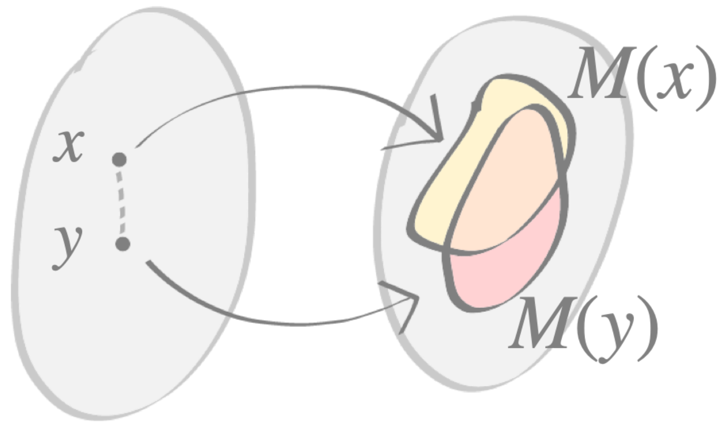
Recommender System



E.g., similar music items/artists with similar music features, should have similar probability of appearing in playlists

How to Define Fairness in Machine Learning?

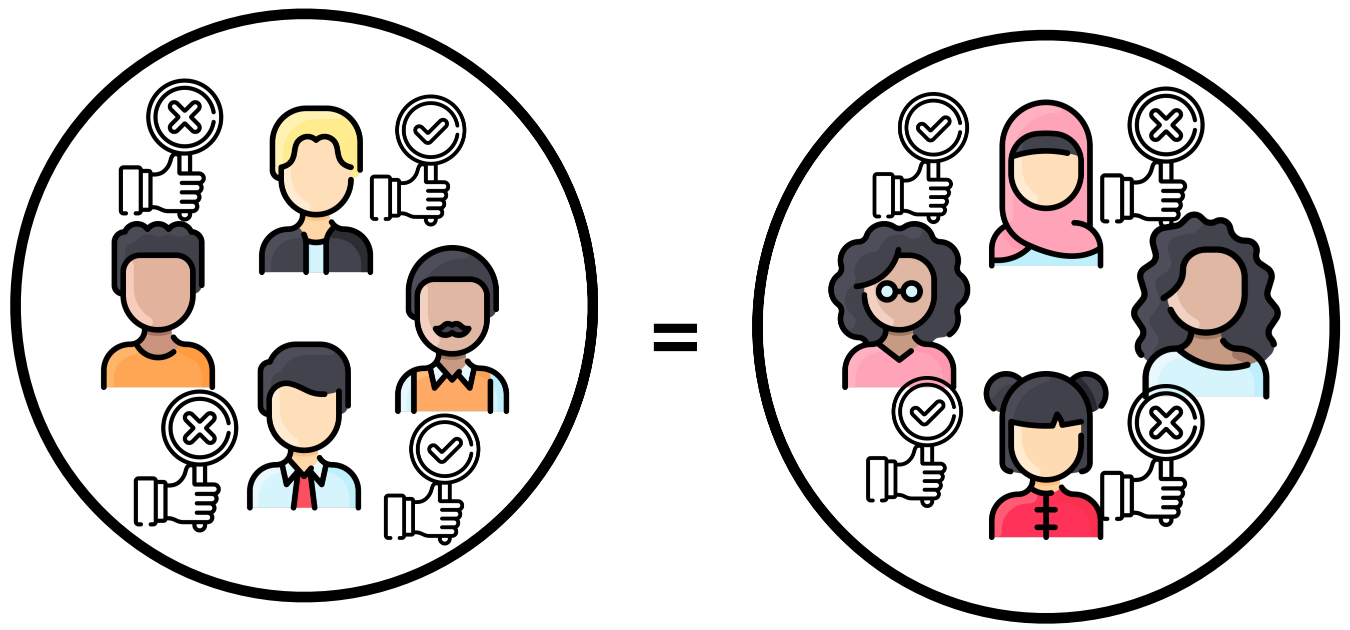
Individual Fairness: measures the impact that discrimination has on the individuals



[Dwork, C., et al. ,2012]

Lipschiz condition $||M(x) - M(y)|| \leq d(x, y)$

Group Fairness: measures the impact that the discrimination has on the groups of individuals



Group Fairness Examples

Demographic/Statistical Parity $P(\hat{Y} = 1 \mid S = 1) = P(\hat{Y} = 1 \mid S = 0)$

I.e., Equal probability of receiving a positive loan approval for female and male applicants

Group Fairness Examples

[Hardt, M., et al., 2016]

Demographic/Statistical Parity

$$P(\hat{Y} = 1 \mid S = 1) = P(\hat{Y} = 1 \mid S = 0)$$

I.e., Equal probability of receiving a positive loan approval for female and male applicants

Equal opportunity

$$P(\hat{Y} = 1 \mid Y = 1, S = 1) = P(\hat{Y} = 1 \mid Y = 1, S = 0)$$

I.e, Classifier should give similar results to applicants of both genders with actual positive loan approval.

Group Fairness Examples

[Hardt, M., et al., 2016]

Demographic/Statistical Parity $$P(\hat{Y} = 1 \mid S = 1) = P(\hat{Y} = 1 \mid S = 0)$$

I.e., Equal probability of receiving a positive loan approval for female and male applicants

Equal opportunity $$P(\hat{Y} = 1 \mid Y = 1, S = 1) = P(\hat{Y} = 1 \mid Y = 1, S = 0)$$

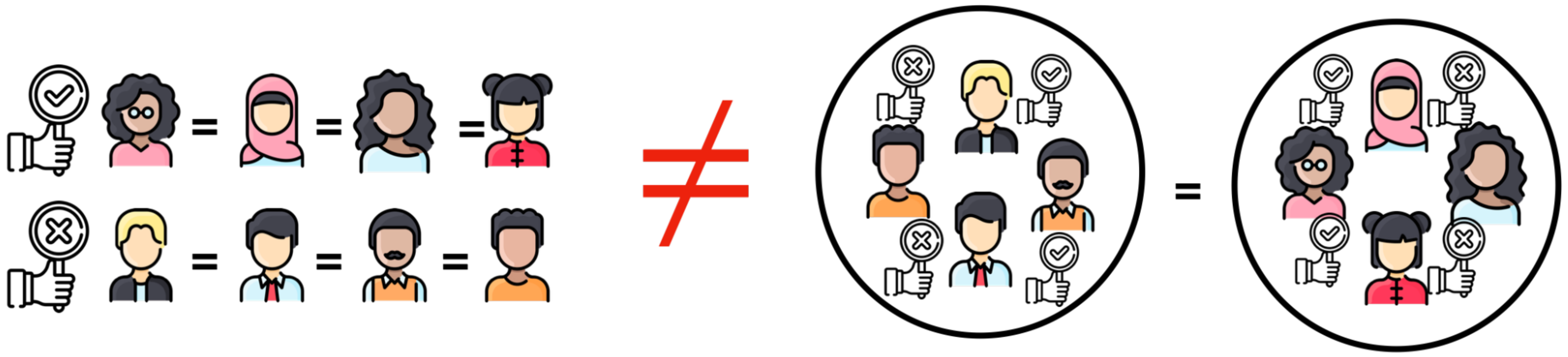
I.e, Classifier should give similar results to applicants of both genders with actual positive loan approval.

Equalized odds $$P(\hat{Y} = 1 \mid Y, S = 1) = P(\hat{Y} = 1 \mid Y, S = 0)$$

I.e., Applicants with a rejected loan application and applicants with an accepted loan application should have a similar classification, regardless of their gender.

Impossibility of Fairness

Impossibility wrt group and individual notions [Friedler, S., et al. ,2016]



Impossibility wrt various group fairness notions [Kleinberg, J., et al. ,2016]

- Demographic Parity
- Equal opportunity
- Equalized odds

Bias is **subjective** and must be considered **relative** to task

Fairness is a socio-technical challenge

Fairness is not a general concept

What is the Importance of Algorithmic Fairness?

How to Define Fairness in Machine Learning?

How to Ensure Algorithmic Fairness in Machine Learning?

How to Ensure Algorithmic Fairness in Machine Learning?

Machine Learning Pipeline

bias

Pre-processing	In-processing	Post-processing
Data is unbalanced	Unfair Objective	Unfair outcome
Historical discrimination	Lack of algorithmic design knowledge	Black-box models
Encoded of protected attributes	Biased loss functions	No user feedback

How to Ensure Algorithmic Fairness in Machine Learning?

Machine Learning Pipeline

bias

Data is unbalanced

Historical discrimination

Encoded of protected attributes

Pre-processing

Unfair Objective

Lack of algorithmic design knowledge

Biased loss functions

In-processing

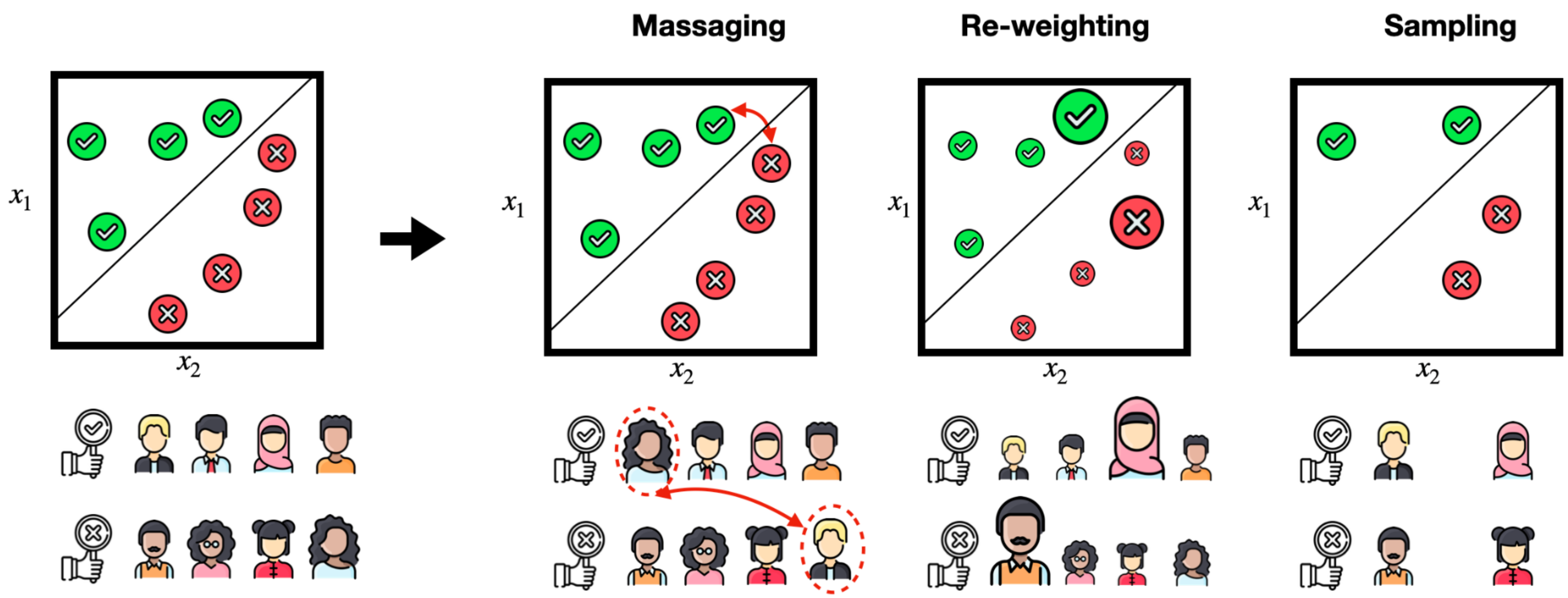
Unfair outcome

Black-box models

No user feedback

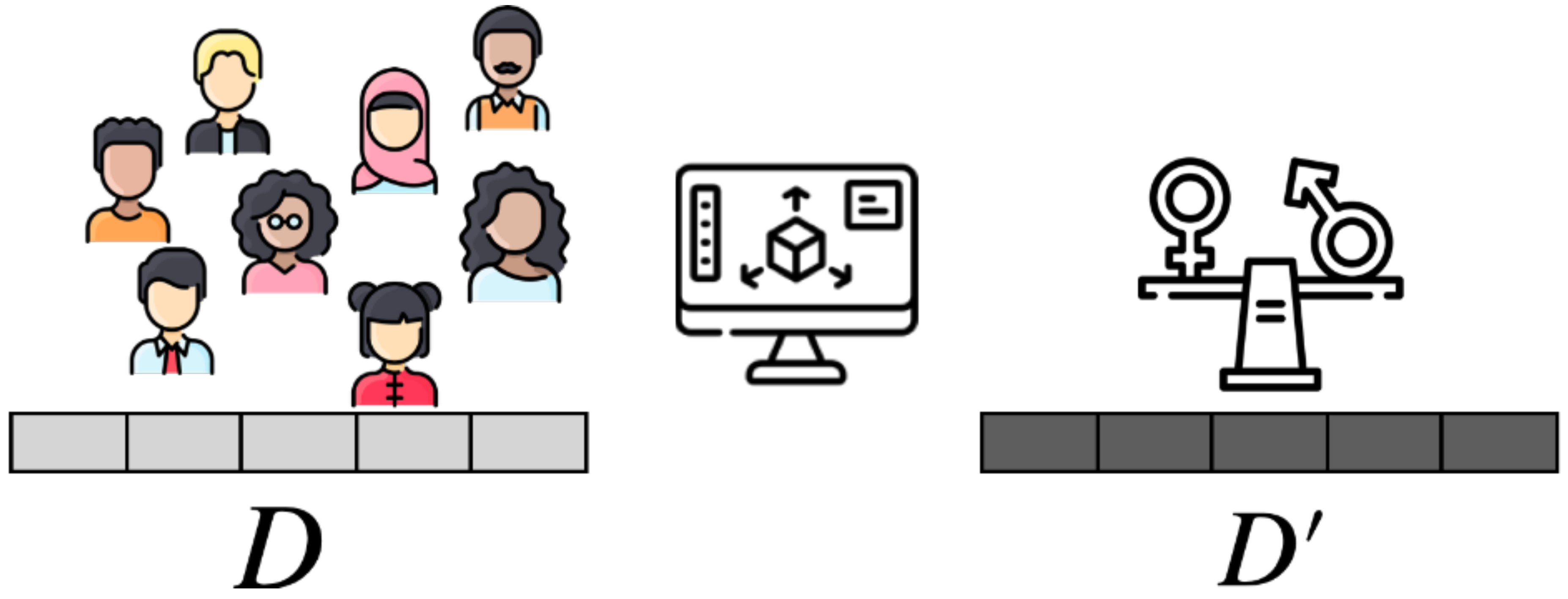
Post-processing

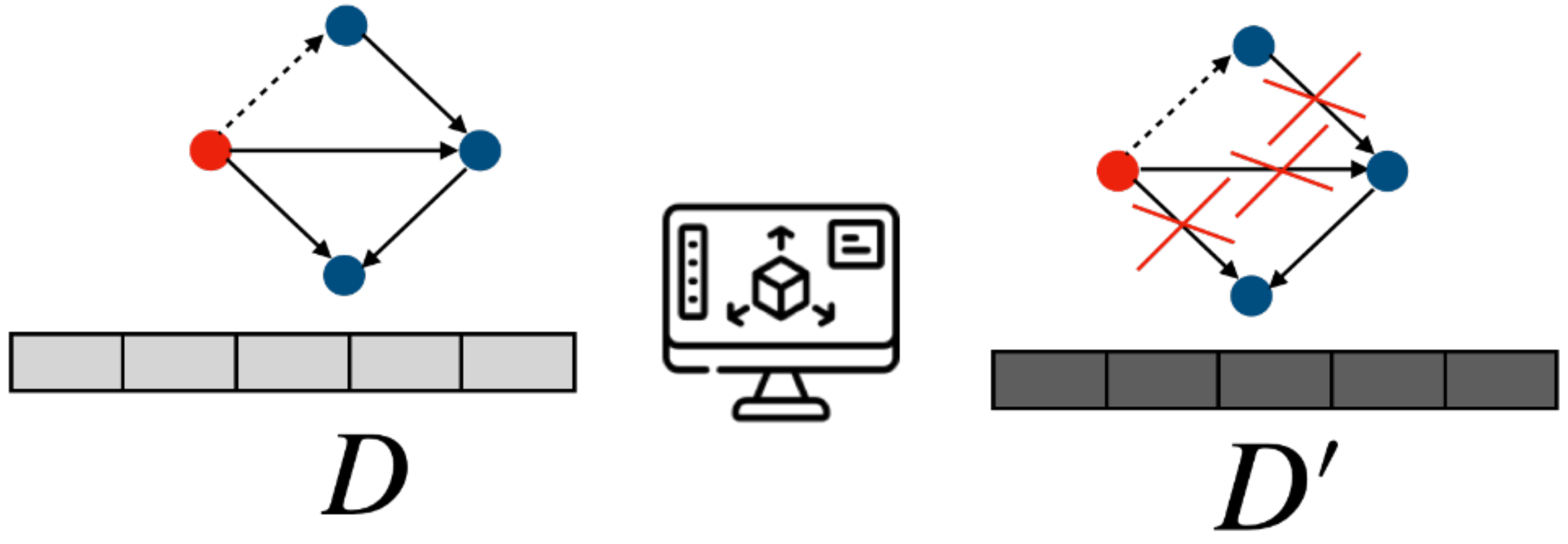
Fairness in Pre-Processing: Data De-Biasing



[Kamiran, F., & Calders, T. (2012)]

Fairness in Pre-Processing: Data Generative Models





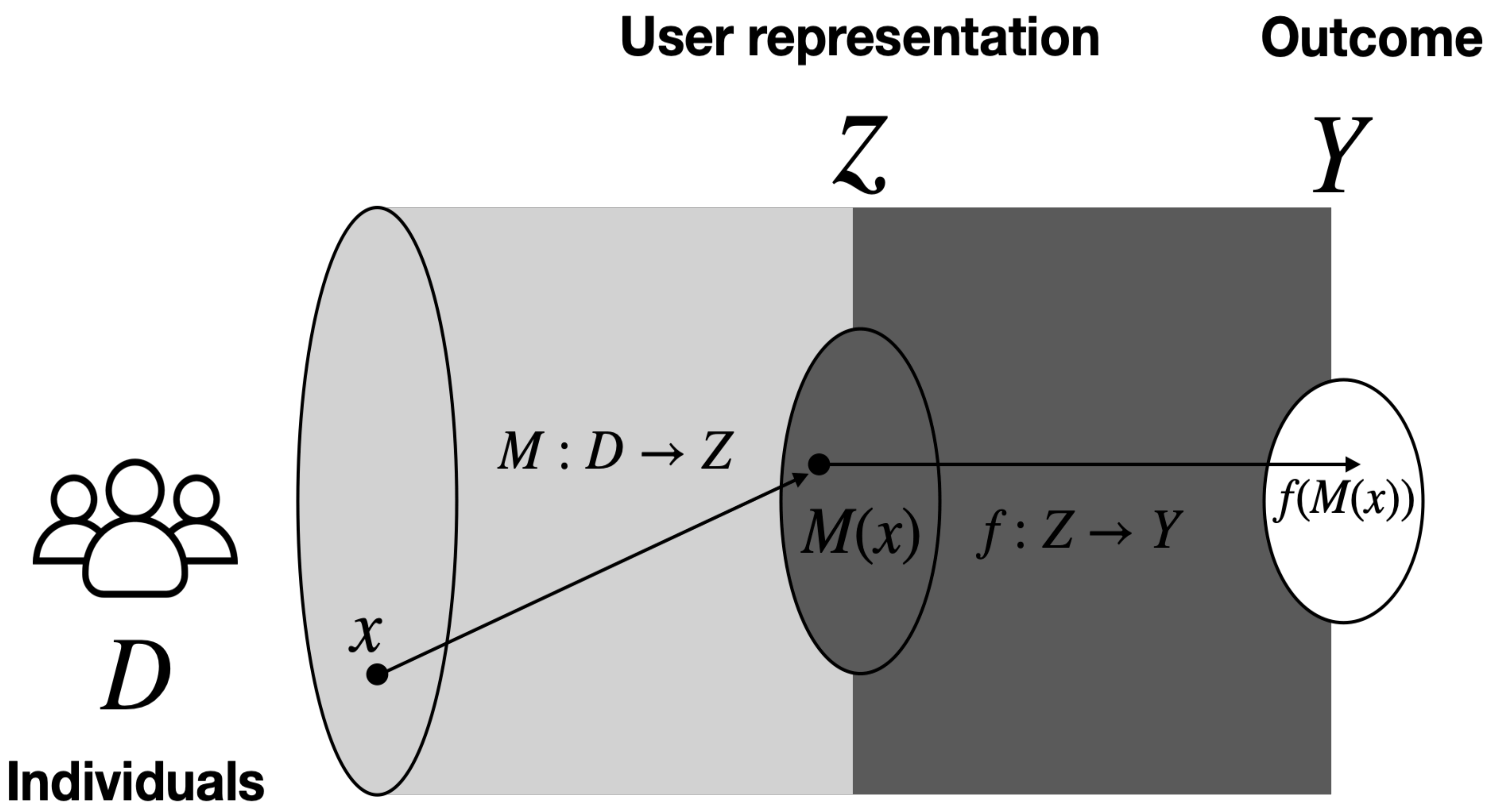
E.g., Using Generative Adversarial Networks (GANs), Variational Autoencoders, etc.

E.g., Using SCMs (by removing paths from sensitive attributes)

Sattigeri, Prasanna, et al. "Fairness GAN: Generating datasets with fairness properties using a generative adversarial network." *IBM Journal of Research and Development* 63.4/5 (2019): 3-1.

van Breugel, Boris, et al. "Decaf: Generating fair synthetic data using causally-aware generative networks." *Advances in Neural Information Processing Systems* 34 (2021): 22221-22233.

Fair Representation Learning



Goal of Representation learning

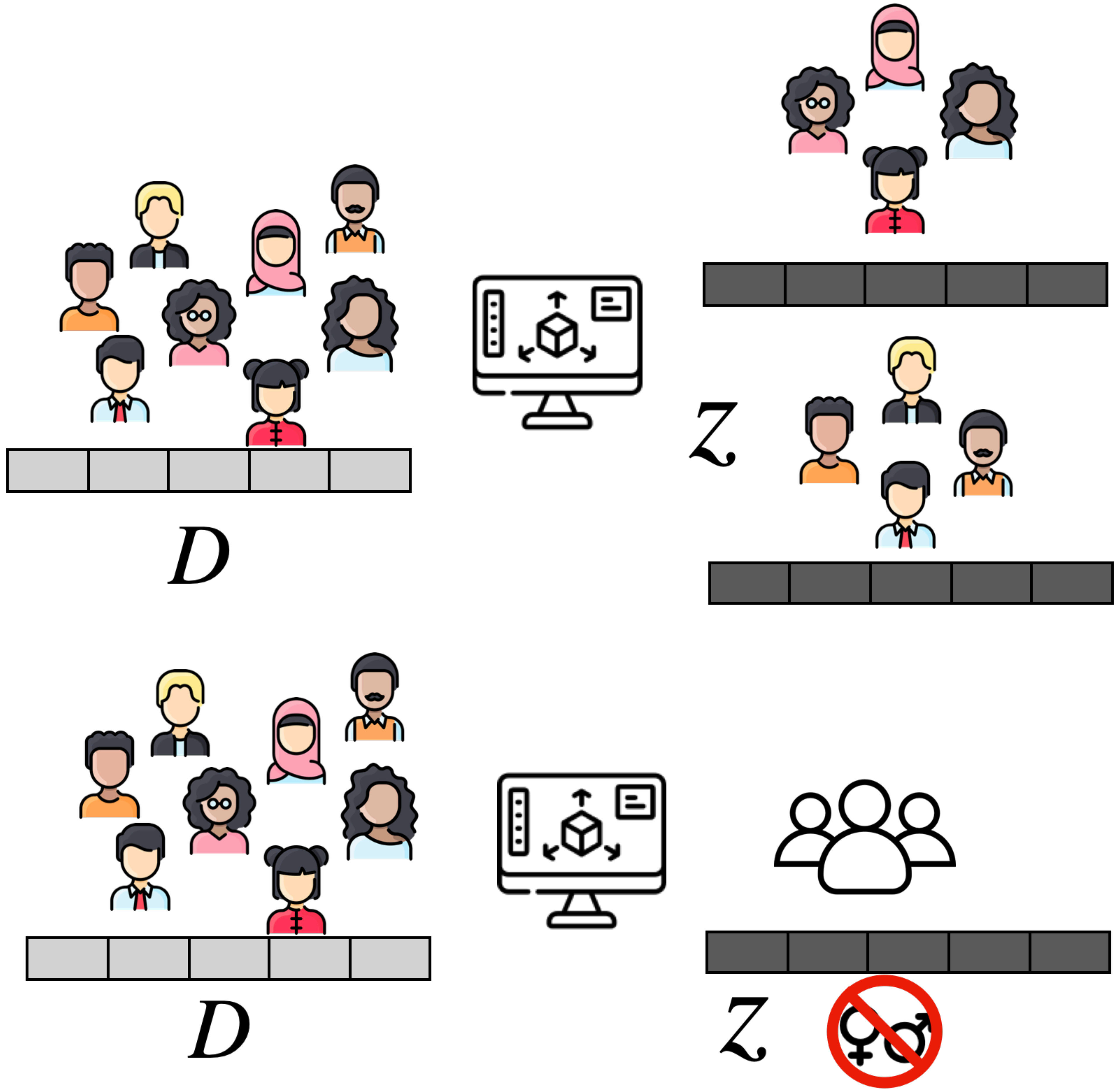
Preserve Performance:

Reconstruction term: the learned representation should resemble the original data

Utility terms: the learned representation should predict target variable

+ Fairness

Fair Representation Learning: Group Fairness



Fairness

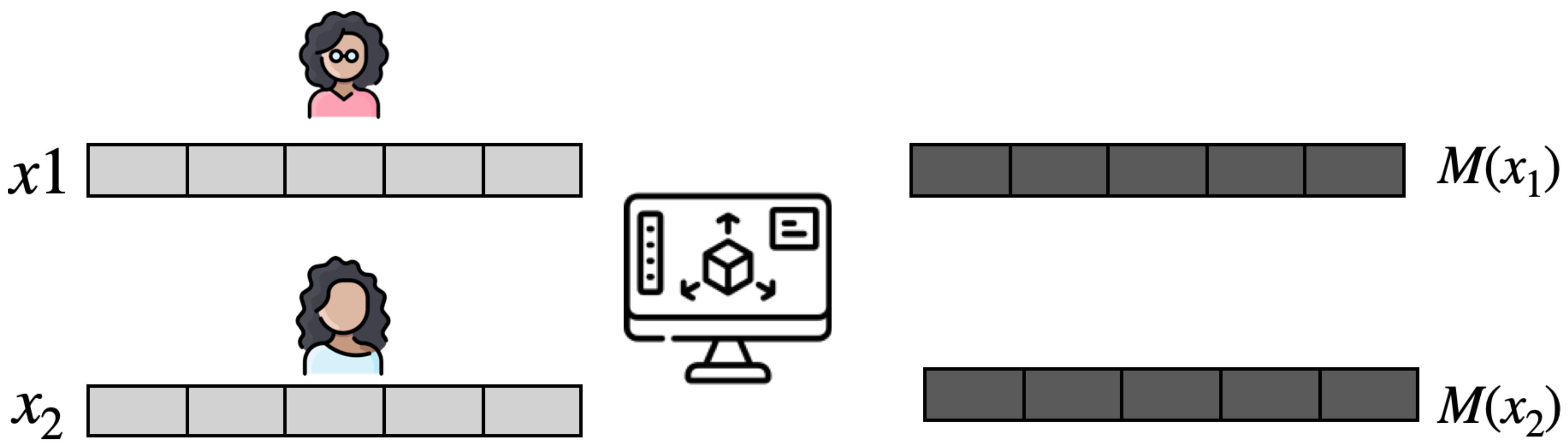
- **Balancing the distribution** among various groups
- **Remove sensitive attributes** (common approach is to use deep learning: VAE, adversarial learning, or disentangled learning)

Louizos, C., Swersky, K., Li, Y., Welling, M., & Zemel, R. (2015). The variational fair autoencoder. arXiv preprint arXiv:1511.00830.

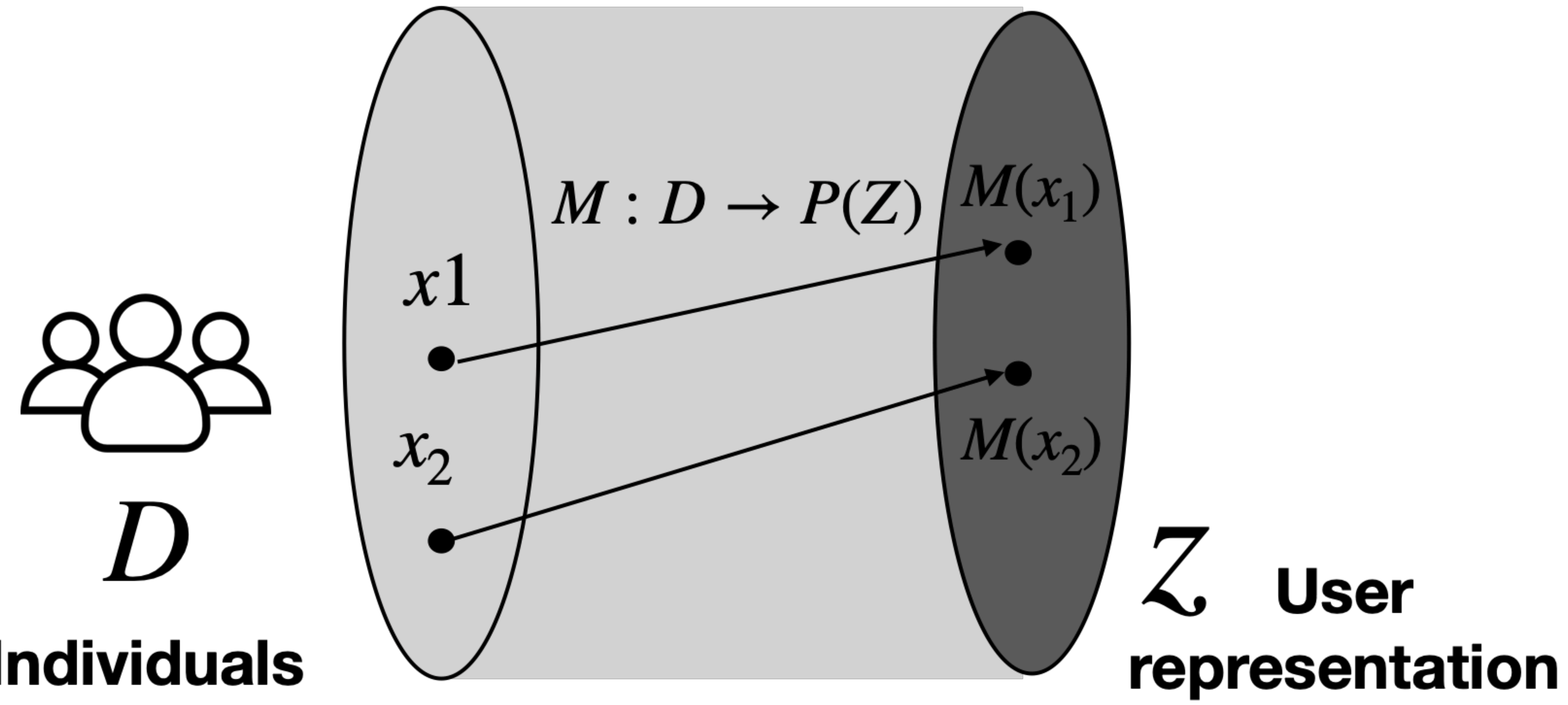
Madras, D., Creager, E., Pitassi, T., & Zemel, R. (2018, July). Learning adversarially fair and transferable representations. In International Conference on Machine Learning (pp. 3384-3393). PMLR.

Locatello, F., Abbati, G., Rainforth, T., Bauer, S., Schölkopf, B., & Bachem, O. (2019). On the fairness of disentangled representations. *Advances in Neural Information Processing Systems*, *32*.

Representation Learning-Individual Fairness
Pre-processing approach



Lipschiz condition $||M(x_1) - M(x_2)|| \leq d(x_1, x_2)$



- **Similar individuals** should map to **similar distributions.**
- **Task-specific** similarity metric. Ideally captures ground truth or society's best approximation
- Many applications: E.g., music playlist continuation.

IVADO

Fernando Diaz (Google Research)

Rebecca Salganik (UdeM)

Salganik, R., Diaz, F., & **Farnadi, G**. (2022). Analyzing the Effect of Sampling in GNNs on Individual Fairness. arXiv preprint arXiv:2209.03904.

How to Ensure Algorithmic Fairness in Machine Learning?

Machine Learning Pipeline

Input →  → Output

bias ········▶

Data is unbalanced

Historical discrimination

Encoded of protected attributes

Pre-processing

Unfair Objective

Lack of algorithmic design knowledge

Biased loss functions

In-processing

Unfair outcome

Black-box models

No user feedback

Post-processing

In-processing techniques

- Machine learning tasks are often expressed as optimization problems (i.e., Empirical Risk Minimization (ERM) framework)

$$minimize_{\theta} \quad f(X, Y; \theta)$$

- The optimization problem: finding the parameters that give the best model w.r.t the desired properties

$$g(X, Y; \theta)$$

Fairness in yet another desired property of the learned models

Accuracy is not enough!

A hypothetical (extreme) situation:

Born and raised in Canada

90% of population

- data describes them accurately
- accurate predictions (95% accurate)

The model is still ~90.5% accurate!

Migrated to Canada in recent years

10% of population

- data describes them poorly
- poor predictions (50% accurate)

In-processing techniques

- **Not all optimization problems are the same!**
- Some problems are **computational easy**
- Some problems are **hard**, but **behave well** (approximation methods work well)
- Some problems are **hard**, but have **structure**. And we can exploit this structure.

Adding fairness can change these properties!

In-processing techniques

Fairness as Constrained Optimization	Fairness as Regularizer	Fairness as Multi-objective Optimization
$\text{minimize}_{\theta} \; f(X, Y; \theta)$ subject to $\boxed{g(X, Y; \theta)}$	$\text{minimize}_{\theta} \; f(X, Y; \theta) + \lambda \boxed{g(X, Y; \theta)}$	$\text{minimize}_{\theta} \; f(X, Y; \theta) \times \boxed{g(X, Y; \theta)}$

How to Ensure Algorithmic Fairness in Machine Learning?

Machine Learning Pipeline

bias

Data is unbalanced

Historical discrimination

Encoded of protected attributes

Pre-processing

Unfair Objective

Lack of algorithmic design knowledge

Biased loss functions

In-processing

Unfair outcome

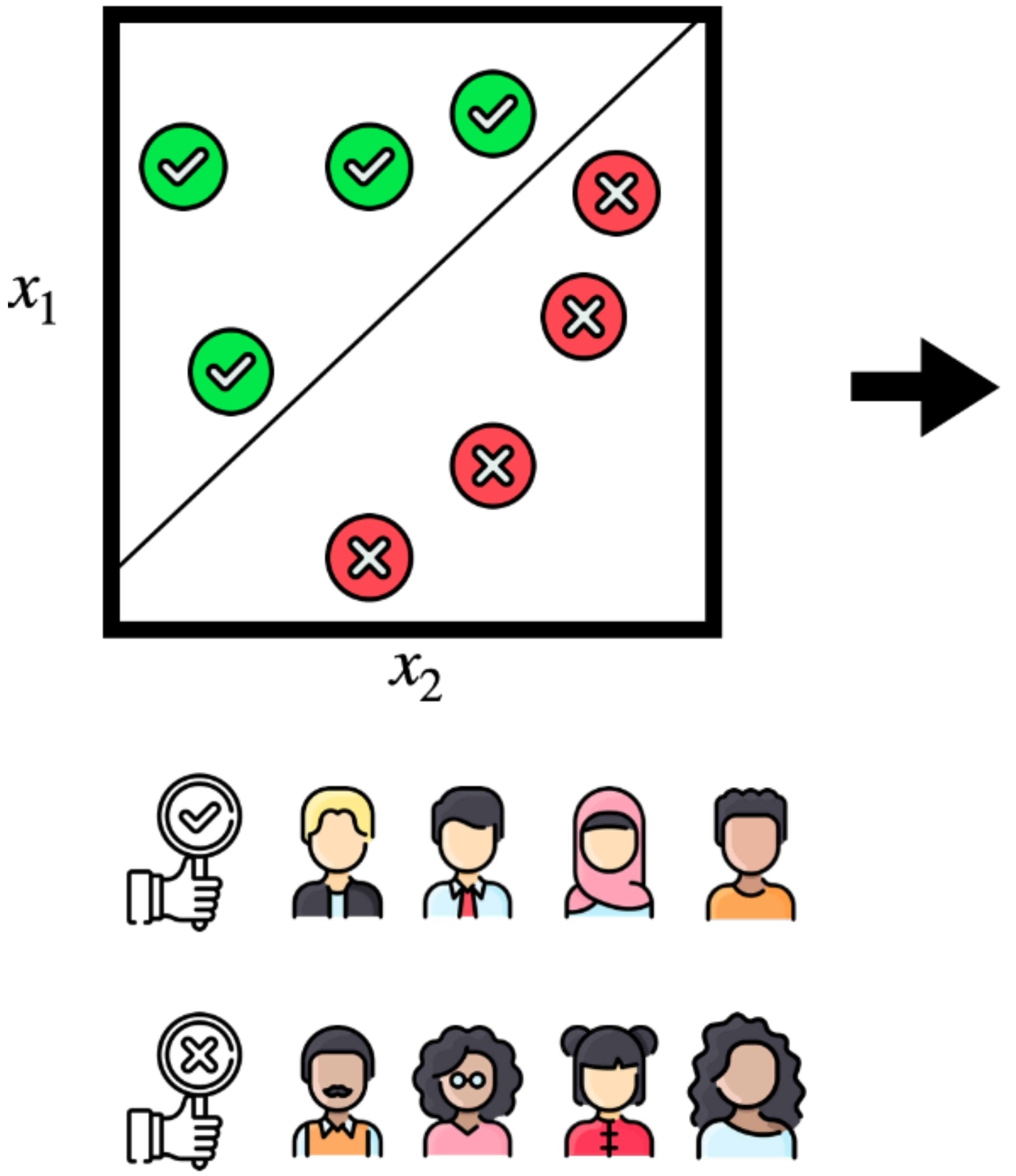
Black-box models

No user feedback

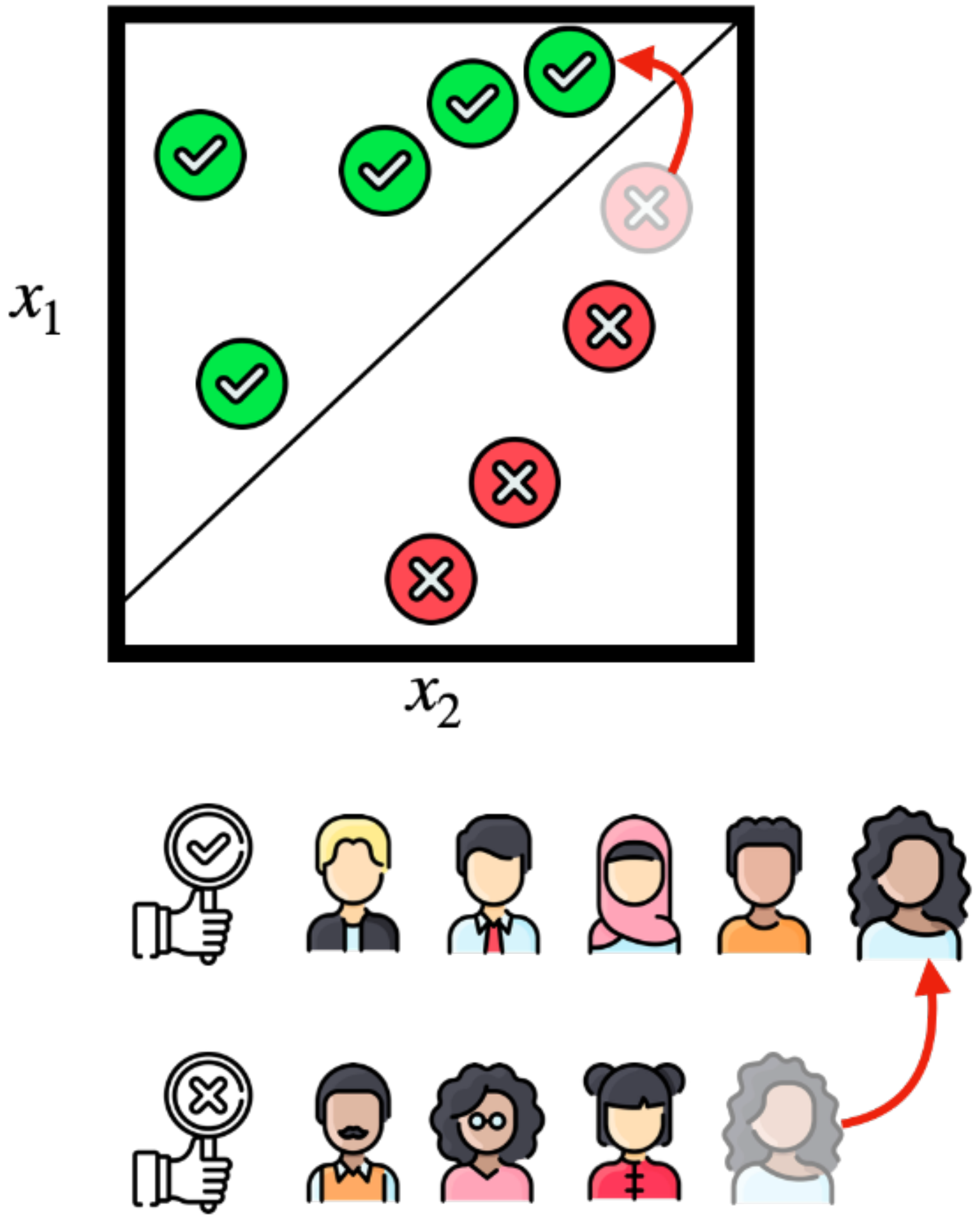
Post-processing

Fairness in Post-Processing

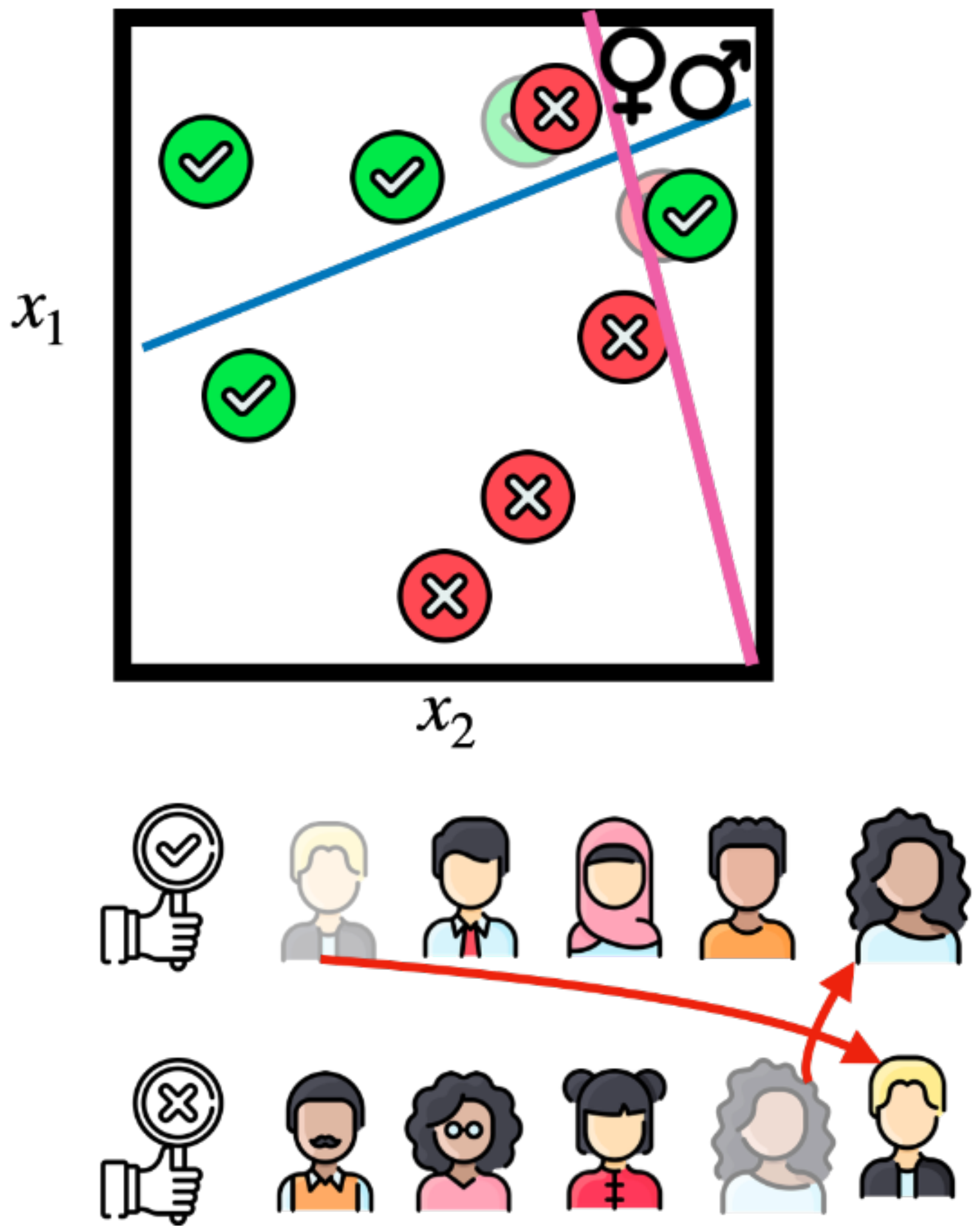
[Hardt, M., et al., 2016]



Rejection Option



Thresholding

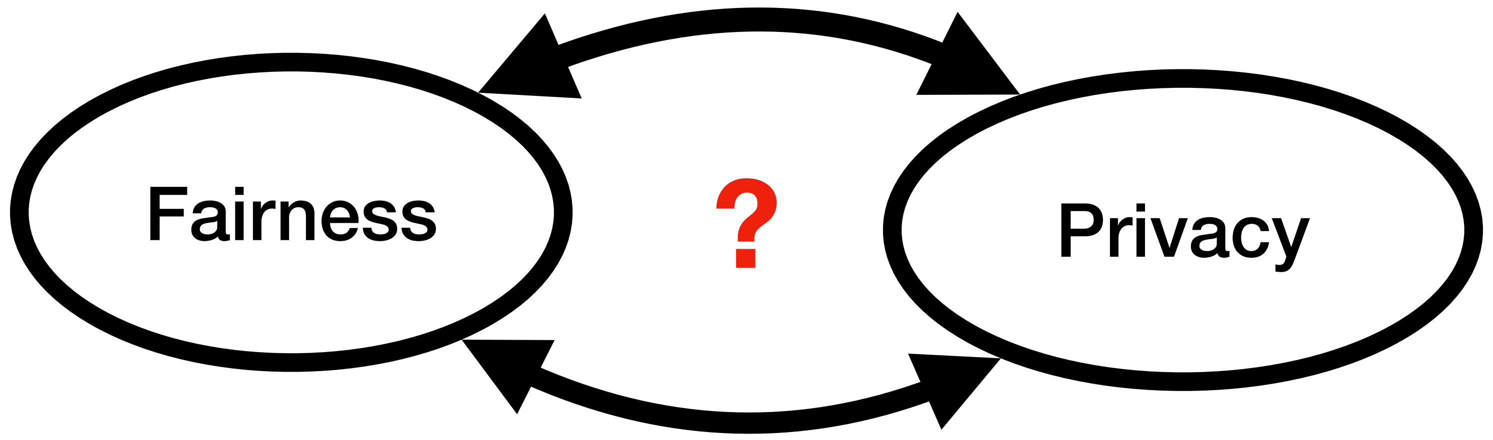


Trade-offs

	Ease of implementation and (re)-use	Scalability	Ease of auditing	Fairness/ Performance tradeoff	Generalization
Pre-processing, e.g., representation learning	✓	✓	✓		✓
In-processing, e.g., fairness regularizer			✓	✓	✓
Post-processing, e.g., thresholding		✓	✓		

Inspired by Sanmi Koyejo's talk on fair representation learning tutorial at NeurIPS 2019

Challenges & Opportunities: Fairness Vs. Privacy



Law against discrimination

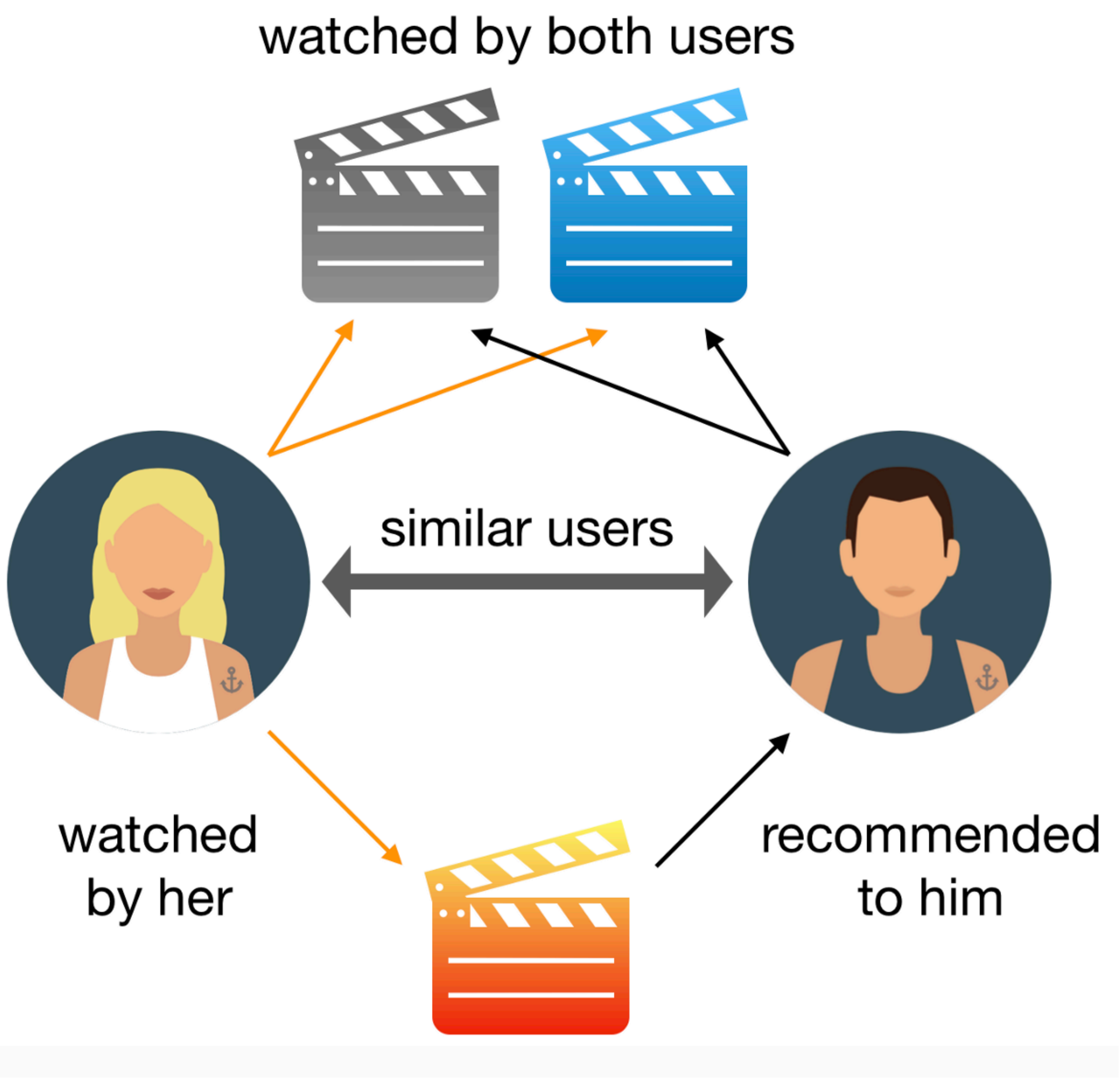
Legally recognized ‘protected classes’

Regulated domains

Data protection laws

Challenges & Opportunities: Fairness is not the same in different fields!

Two-sided fairness in recommender systems



Take aways

Bias happens throughout the automated systems:

- Educate people about **discrimination**
- How to **define fairness** in your set-up?
- Ask who is **using** the model?
- What is **the purpose** of the system?

Recommendation Systems for Décathlon - Discussion

Case developed by Jérémi DeBlois-Beaucage supervised by Laurent Charlin & Renaud Legoux

Presentation of the case

Q1: Which model(s) for a recommendation system would the data science team need to choose, and why?

Session in small groups

- ~15 minutes, groups of ~4, prepare 3—4 slides
    - Suggestion: designate one scribe and one presenter by team
- Then: we will discuss your answers in class

Discussions

Plan

- Choose the right model for a recommendation task
- Models chosen by Décathlon
    - Basic models, as reference
    - Model 1: Based on similarity between product images
    - Model 2: Collaborative filtering based on products
    - Model 3: Matrix factorization
    - Model 4: Recursive neuronal networks
    - Chosen metrics
    - Results and final choice
- Limits of the current models and the next steps being considered by Décathlon

How to Choose the Right Model?

- A more subjective task than most other common machine learning tasks
- Imperative: the model must be able to process a large amount of data
- Started with simpler models and then moved on to more complex ones
- Final choice according to performance on chosen metrics and logistical considerations
- 4 models have been selected

Basic Models: Reference Point

- Random recommendation
- Recommending the same 10 most popular items to every user

Types of Recommender Systems

Content Filtering

- *Example*: Pandora.com music recommendations (Music Genome Project)
- **Con:** Assumes access to side information about items (e.g. properties of a song)
- **Pro:** Got a new item to add? No problem, just be sure to include the side information

Collaborative Filtering

- *Example*: Netflix movie recommendations
- **Pro:** Does not assume access to side information about items (e.g. does not need to know about movie genres)
- **Con:** Does not work on new items that have no ratings

Model 1: Based on Similarity Between Product Images

Each product is first assigned an image and a vector that represents this image, created through a pre-trained VGG-type convolutional neural network.

1. For each user, a list of all the products with which they have interacted is extracted.

2. For each product, the 10 most “similar” products are chosen, based on the cosine distance between their image vectors.

3. The most similar product gets 10 “points”, the second one gets 9, and so on. Points are added up, and the 5 products with the highest scores get recommended.

User A has interacted in the past with items **3** and **5**

For each item, we identify the 10 items in the rest of the catalog that are most similar. We give 10 pts to the most similar one, 9 pts to the second most, and so on

3	5	
7	11	10 pts
12	7	9 pts
37	22	8 pts
11	30	7 pts
22	1	6 pts
6	26	5 pts
1	27	4 pts
28	12	3 pts
29	15	2 pts
30	9	1 pts

We sum all the scores, and recommend the top five items to our user

Recommendations to user **A**:

Item **7** (19 pts)
Item **11** (17 pts)
Item **22** (14 pts)
Item **12** (12 pts)
Item **1** (10 pts)

Pros and Cons

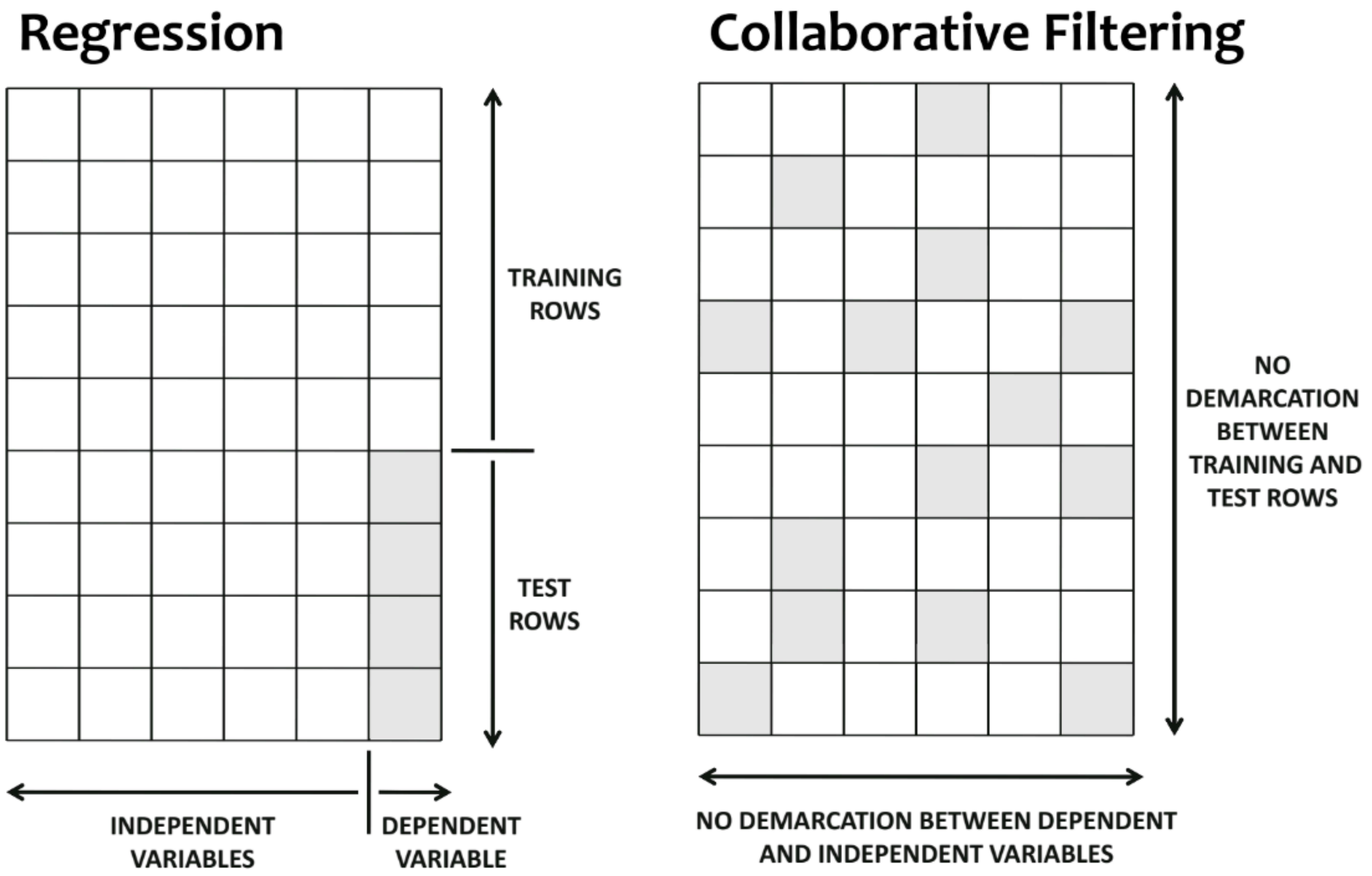
- (+) Even new or unpopular products can get recommended
- (+) Explanations can be offered (because you bought product X, you could find these products interesting)
- (+) Easy and quick to implement

- (-) The recommended products are very similar to previously purchased items
- (-) *Cold-start* problem for the users

Collaborative Filtering

- **Everyday Examples of Collaborative Filtering...**
  - Bestseller lists
  - Top 40 music lists
  - The "recent returns" shelf at the library
  - Unmarked but well-used paths thru the woods
  - The printer room at work
  - "Read any good books lately?"
  - ...
- **Common insight:** personal tastes are correlated
  - If Alice and Bob both like X and Alice likes Y then Bob is more likely to like Y
  - especially (perhaps) if Bob knows Alice

Regression vs. Collaborative Filtering



Model 2: Collaborative Filtering Based on Products

Very similar model to the previous one. Instead of calculating the similarity of image vectors, similarity is calculated according to a user-product interaction matrix.

1. For each user, a list of all the products with which they have interacted is extracted.

2. For each product, the 10 most "similar" products are chosen, based on a cosine distance between their respective lines in the user-product interaction matrix.

3. The final score is similar to the one produced by model 1. However, instead of attributing arbitrary points (10 pts for the 1st, 9 for the 2nd, etc.), the similarity scores are used directly. Points are added, and the 5 products with the highest scores get recommended.

Example: similarity between two products

- In this matrix, lines represent users and columns represent products. A value of 1 indicates an interest, and 0 means no interaction.

  - For example, the first column shows that only User 4 has interacted with Product 1.

- The cosine similarity between the first (a =[0 0 0 1]) and the second product (b=[0 1 0 0]) would be 0. No user has interacted with both products.

- The similarity between the third (c=[1 0 1 0]) and the fifth product (e=[1 0 1 1]) is 0.82. The nearer the value is to 1, the greater the similarity between the products.

Products

```
[0, 0, 1, 0, 1, 0]
[0, 1, 0, 0, 0, 0]
[0, 0, 1, 0, 1, 0]
[1, 0, 0, 0, 1, 0]
```

Users

Pros and Cons

- (+) Explanations can be offered (because you bought product X, you could find these products interesting)
- (+) Easy and quick to implement
- (+) Collaborative filtering usually yields good results

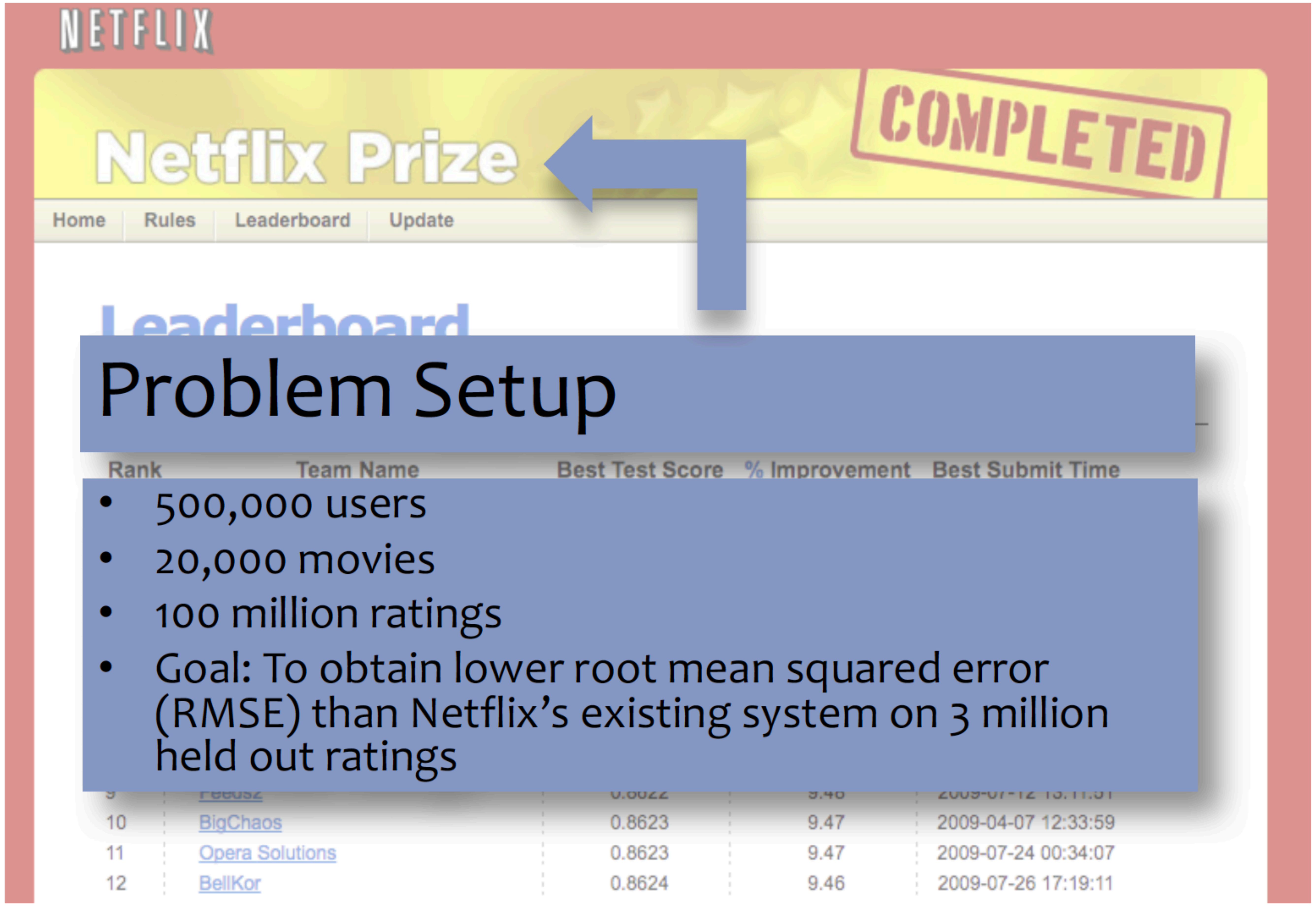
- (-) Cold-start problem for the users and the products

Model 3 : Matrix Factorization

- Completion of the user-product interaction matrix through matrix factorization
- Several methodologies have been used, like Singular Value Decomposition and Non-Negative Matrix Factorization.
- It predicts the probability of an interaction with each product. Recommended products are the one with the highest probabilities of interaction.

Netflix Prize

Problem Setup



HEC MONTRÉAL

Leaderboard

NETFLIX

Netflix Prize

COMPLETED

Home | Rules | Leaderboard | Update

Leaderboard

Showing Test Score. Click here to show quiz score

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11

Matrix Factorization

MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS

Yehuda Koren, *Yahoo Research*
Robert Bell and Chris Volinsky, *AT&T Labs—Research*

As the Netflix Prize competition has demonstrated, matrix factorization models are superior to classic nearest-neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels.

Modern consumers are inundated with choices. Electronic retailers and content providers offer a huge selection of products, with unprecedented opportunities to meet a variety of special needs and tastes. Matching consumers with the most appropriate products is key to enhancing user satisfaction and loyalty. Therefore, more retailers have become interested in recommender systems, which analyze patterns of user interest in products to provide personalized recommendations that suit a user's taste. Because good personalized recommendations can add another dimension to the user experience, e-commerce leaders like Amazon.com and Netflix have made recommender systems a salient part of their websites.

Such systems are particularly useful for entertainment products such as movies, music, and TV shows. Many customers will view the same movie, and each customer is likely to view numerous different movies. Customers have proven willing to indicate their level of satisfaction with particular movies, so a huge volume of data is available about which movies appeal to which customers. Companies can analyze this data to recommend movies to particular customers.

RECOMMENDER SYSTEM STRATEGIES

Broadly speaking, recommender systems are based on one of two strategies. The *content filtering* approach creates a profile for each user or product to characterize its nature. For example, a movie profile could include attributes regarding its genre, the participating actors, its box office popularity, and so forth. User profiles might include demographic information or answers provided on a suitable questionnaire. The profiles allow programs to associate users with matching products. Of course, content-based strategies require gathering external information that might not be available or easy to collect.

A known successful realization of content filtering is the Music Genome Project, which is used for the Internet radio service Pandora.com. A trained music analyst scores

 Published by the IEEE Computer Society 0018-9162/09/$26.00 © 2009 IEEE

Yehuda Koren, Yahoo Research

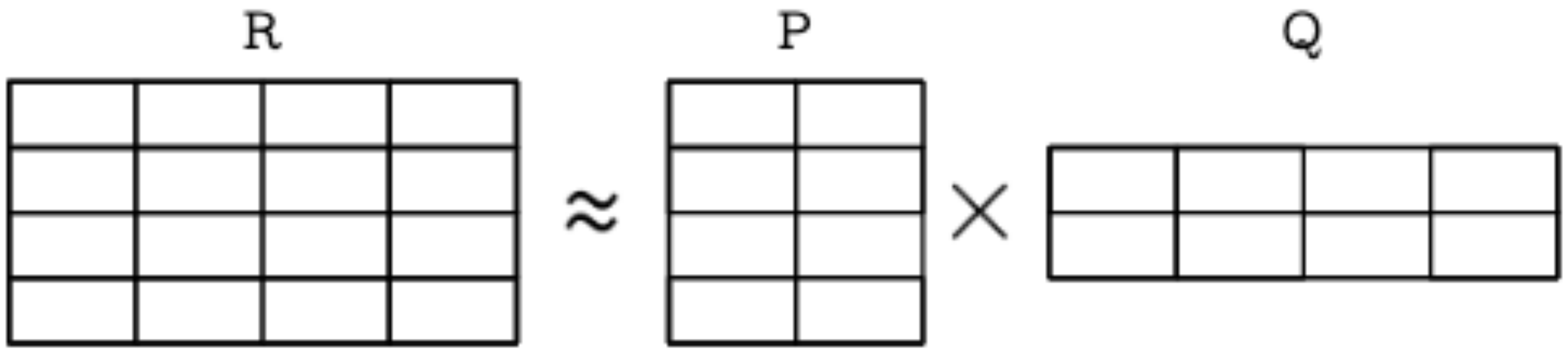
Robert Bell and Chris Volinsky, AT&T Labs-Research

Paper published in August 2009

Authors won the grand Netflix Prize in September 2009

Matrix Factorization

the completion is driven by a factorization



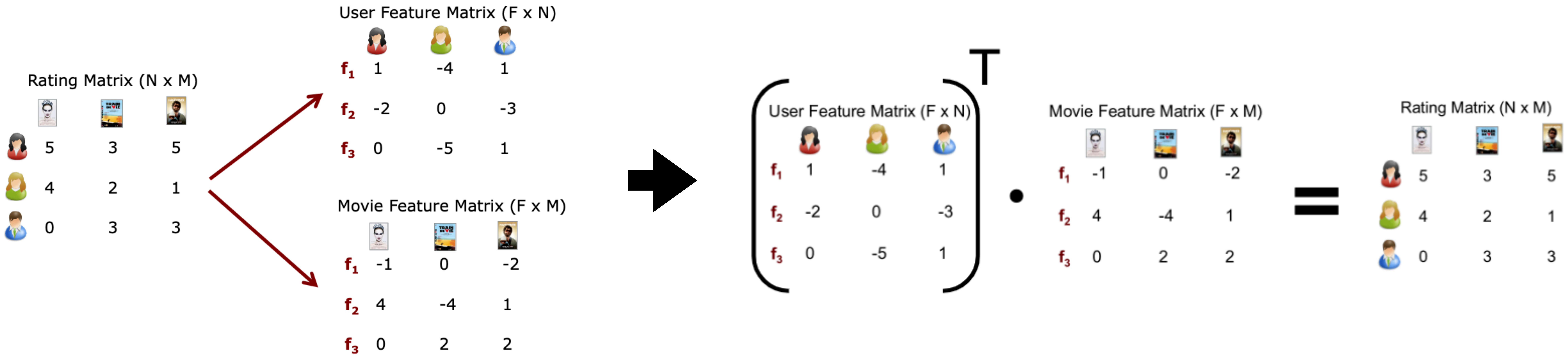
associate a latent factor vector with each user and each item

missing entries are estimated through the dot product

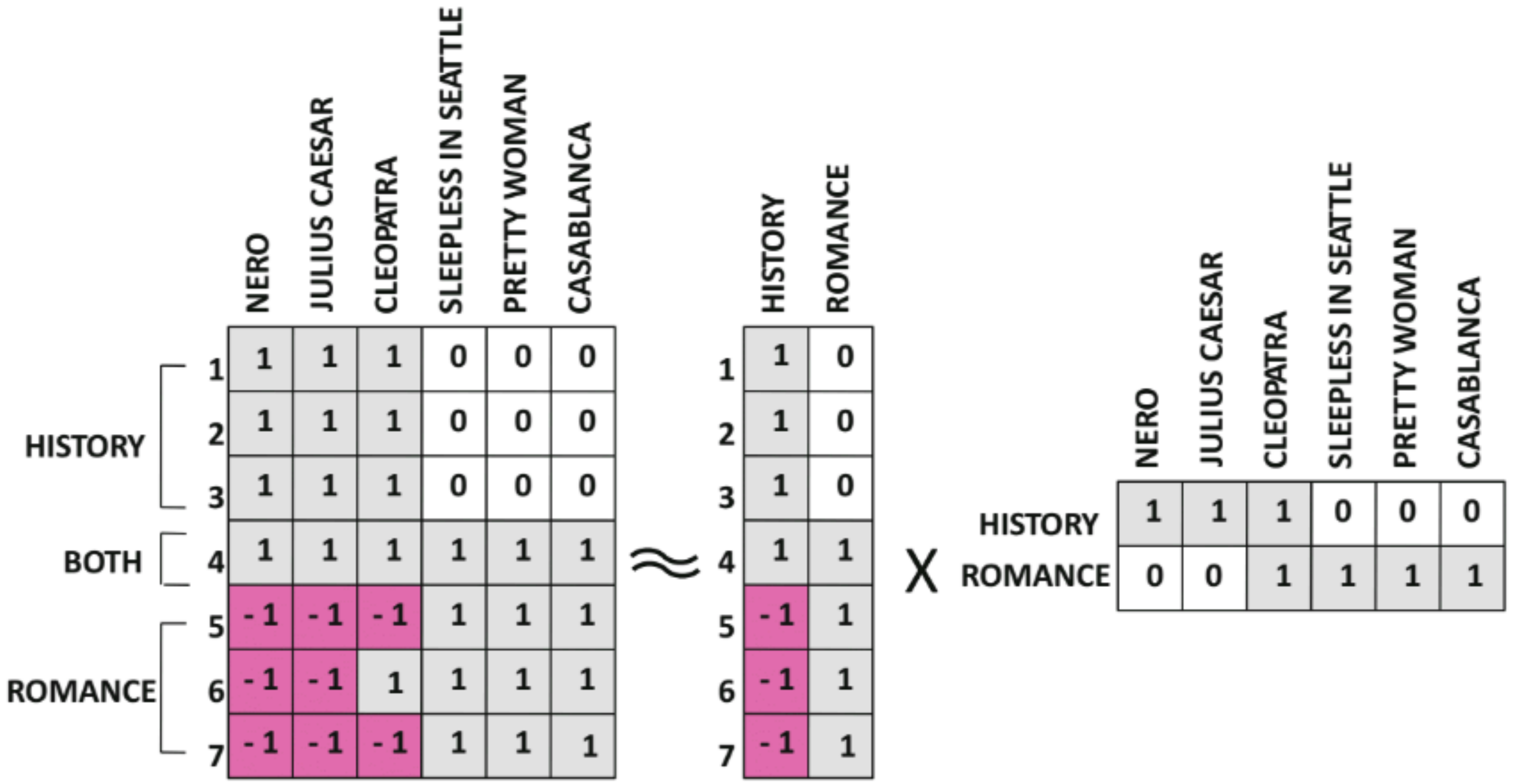
$$r_{ij} \approx p_i q_j$$

Slides are collected from:
Matt Gormley at CMU
Markus Freitag, Jan-Felix Schwarz from University Potsdam
Jure Leskovec from Stanford

Matrix Factorization



Example of MF for Netflix problem



Matrix Factorization

$$\min_{q^*,p^*} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2$$

r_{ui} : known rating of user u for item i

remember:
predicted rating $\hat{r}_{ui} = q_i^T p_u$

Regularization to avoid overfitting

Idea: penalize complexity

$$\min_{q^*, p^*} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$

known as Funk SVD

λ : constant to control the extend of regularization

→ determined by cross-validation

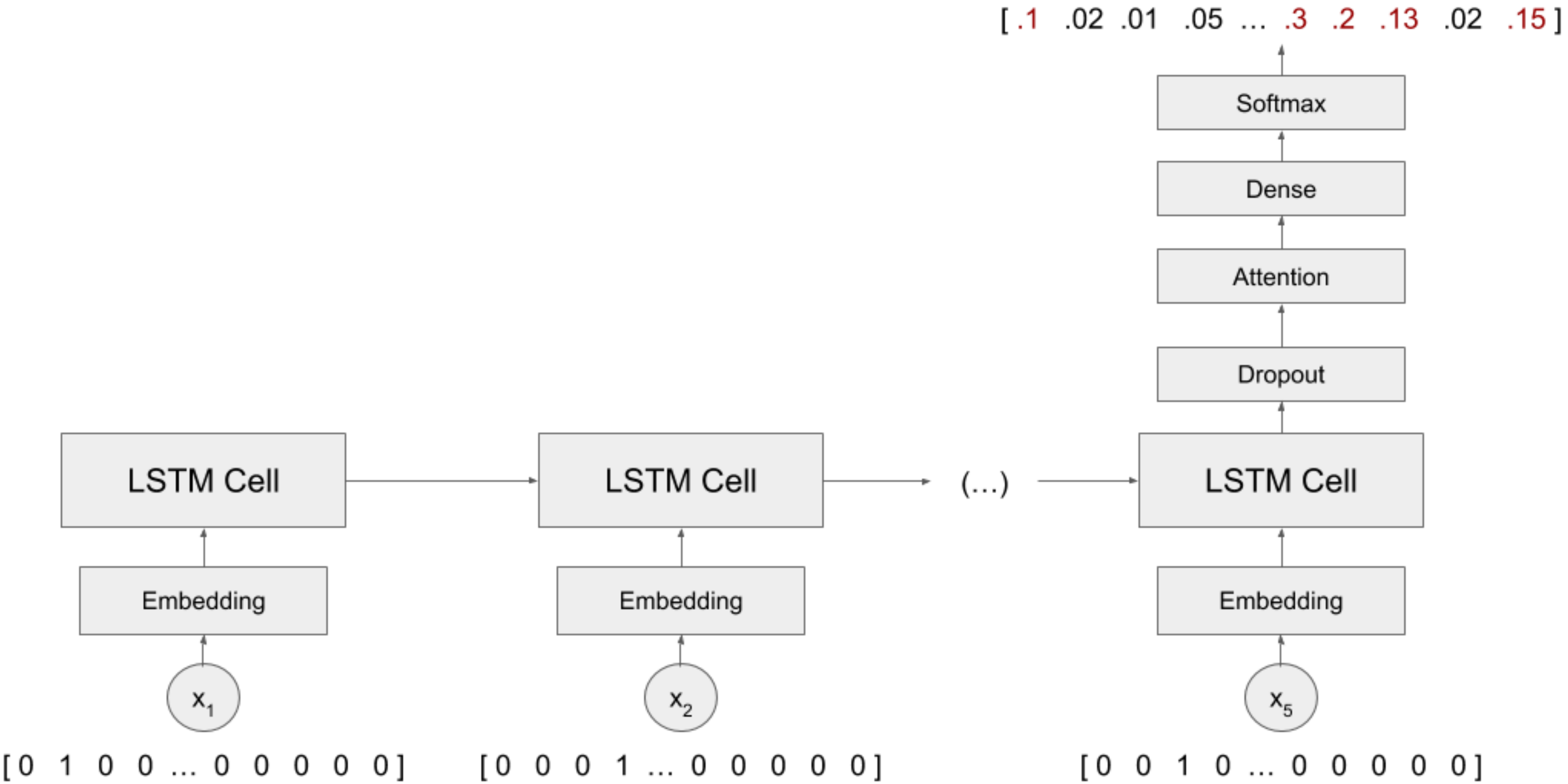
Pros and Cons

- (+) Matrix factorization usually yields good results
- (+) It can reveal interesting underlying characteristics

- (-) Cold-start problem for the users and the products
- (-) Important computational costs
- (-) Deploying the models requires a more complex virtual infrastructure*

Model 4: Recurrent Neural Networks

- This model tackles the problem as a sequence of products with which the user interacts, and tries to predict what the next one could be.
  - Each product is represented by a one-hot vector.
    - Products are entered into the network sequentially, and the network predicts the next one.
  - The network then outputs a probability distribution for every product in the catalogue.
  - Long Short-Term Memory (LSTM) neurons are used, with dropout-type regularization and attention principles.



[.1 .02 .01 .05 … .3 .2 .13 .02 .15]
Softmax
Dense
Attention
Dropout
LSTM Cell
LSTM Cell
(…)
LSTM Cell
Embedding
Embedding
Embedding
x1
x2
x5
[0 1 0 0 … 0 0 0 0 0]
[0 0 0 1 … 0 0 0 0 0]
[0 0 1 0 … 0 0 0 0 0]

Pros and Cons

- (+) New users can be easily added
- (+) Recurrent networks usually give very good results

- (-) Cold-start problem for the products
- (-) Works better if a user has interacted with several products

Metrics

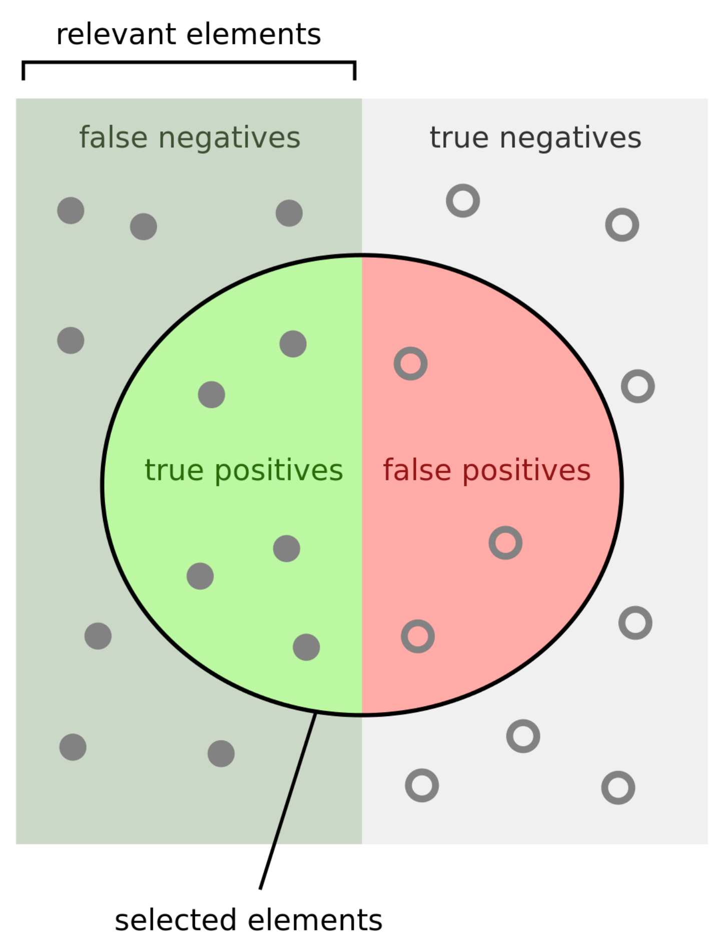
Only off-line metrics: the team does not have access to on-line evaluation or user studies

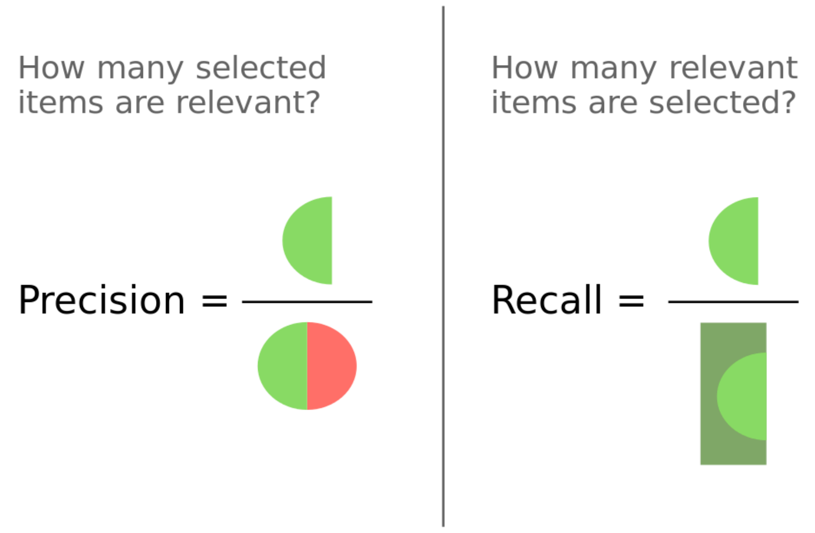
A. Accuracy: proportion of the recommended products that actually get bought by the users

B. Recall: proportion of products that were actually bought which were recommended

C. Coverage: proportion of products that were recommended to at least 1 user

Option to not consider diversity or serendipity

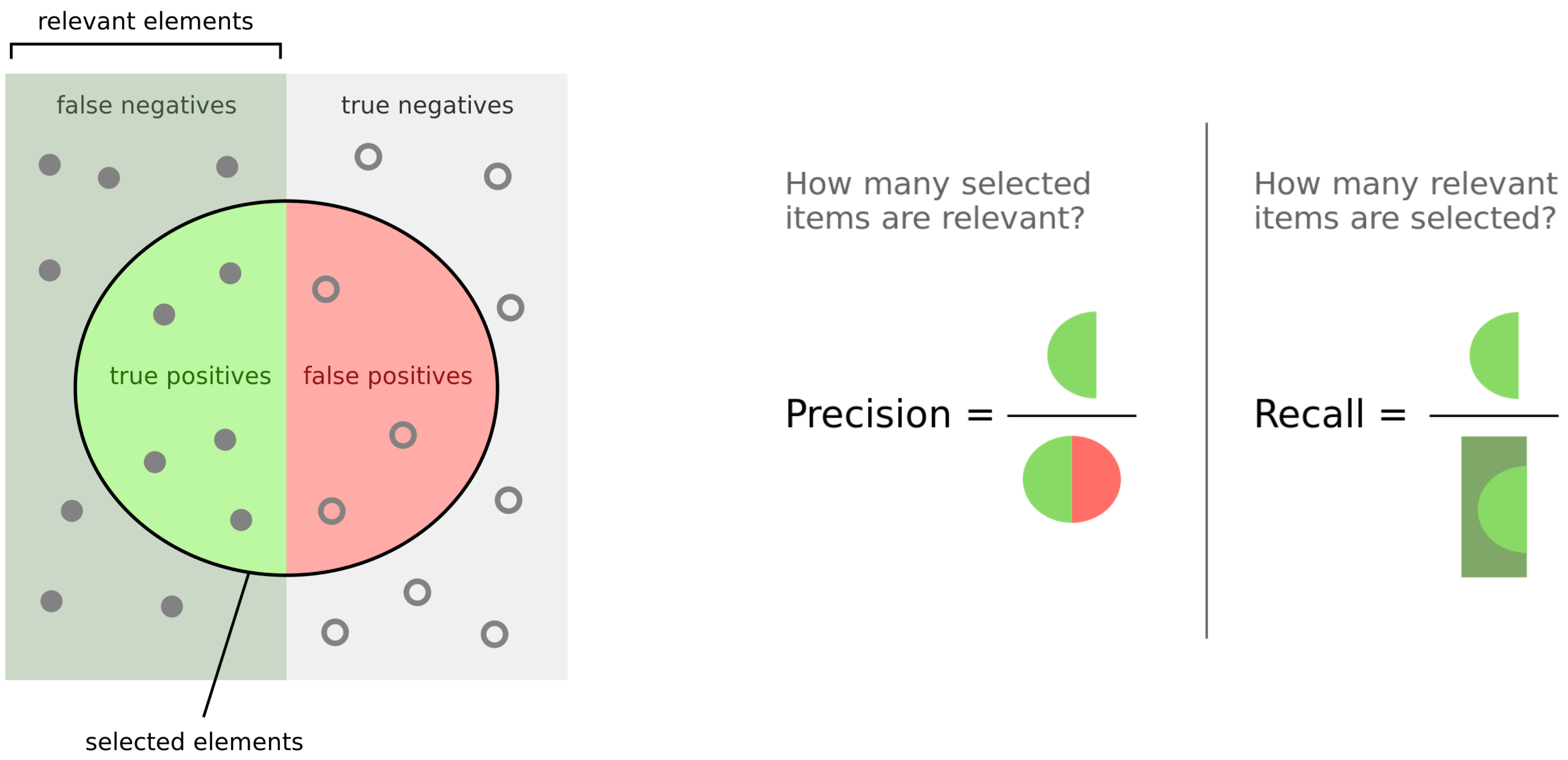




https://en.wikipedia.org/wiki/Precision_and_recall

Results

Model	Random	Most popular	1. Visual Similarity	2. Collaborative, based on products	3. Matrix factorization	4. Recurrent Neural Networks (RNNs)
Precision	0,06%	1,5%	1,9%	3,4%	3,9%	4%
Recall	0,07%	1,8%	2,3%	4,1%	5,3%	5,7%
Coverage	91%	0,07%	37,1%	74%	69%	57%



https://en.wikipedia.org/wiki/Precision_and_recall

Final Choice

First, simpler models were implemented:

1. Most popular products

2. Collaborative based on products (Model 2)

The visual similarity model (Model 1) did not show sufficient performance to justify its implementation.

Matrix factorization (Model 3) was put aside: the performance boost did not justify the investment in more complex logistical architecture.

The recurrent neural network model is in use now (Model 4)

Limits

- There’s an intuition that another model might perform better for the chosen metrics
  - The team is looking for a model that uses both the user-product interaction data and product characteristics.
- The current model focuses on short-term performance
  - The team would like a model that can further explore the diversity of user preference, and potentially offer pleasantly surprising products to users

To go further

What improvements could be made, or what more advanced models could be prioritized for testing?

Session in Small Groups

If enough time…

Improvements and New Models Being Considered

1. **Curiosity in recurrent neuronal networks**
   - Adding functionalities to the networks being used: adding curiosity techniques
   - Allows for a more thorough exploration of diversity in user preferences
2. **Graph neural networks**
   - Structure data differently: heterogeneous graph, which allows for the use of interaction data and product characteristics
   - Prediction of the links between the graph’s knots
3. **Learning through reinforcement**
   - Model the task differently: sequential and interactive process, considered as a loop between product recommendations and user feedback
   - Allows for a more thorough exploration of diversity in user preferences