Intersectional fair ranking via subgroup divergence

Eliana Pastor, Francesco Bonchi

Politecnico di Torino, CENTAI - Italy

ECML PKDD 2024 – Vilnius, Lithuania

Politecnico di Torino

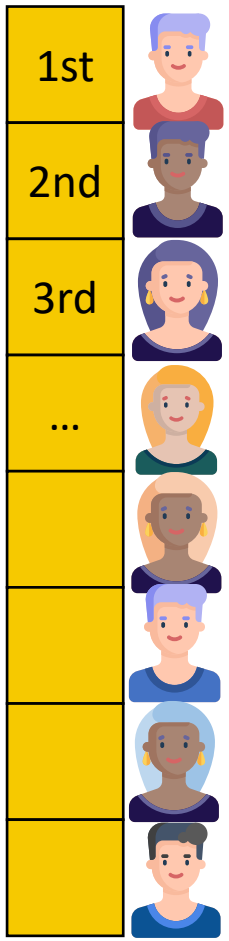
DBMG

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

eurecat

CENTAI

Rankings are adopted everywhere



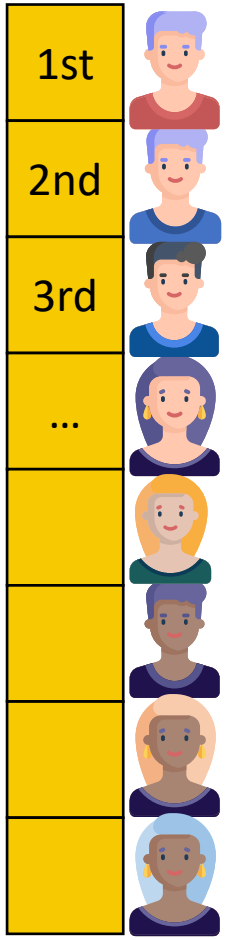
Internship access

University admissions

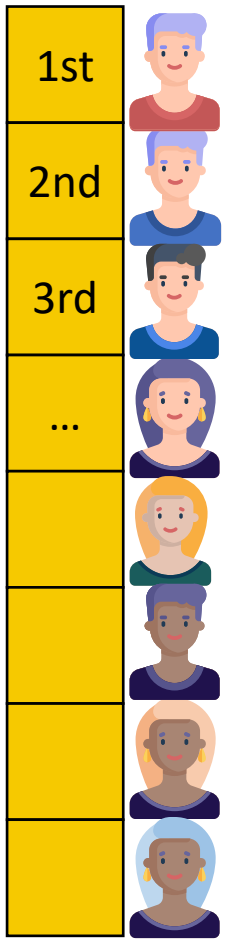
Job positions

Travel grants

Rankings are adopted everywhere



Rankings are adopted everywhere



Advantaged

Disadvantaged

Even more disadvantaged

→ **Intersectionality**

Fair ranking – current approaches

A priori knowledge disadvantaged protected groups

Impose fairness contraints

At least X for each gender, at least Y for each ethnicity

Fair ranking – current approaches

A priori knowledge disadvantaged protected groups

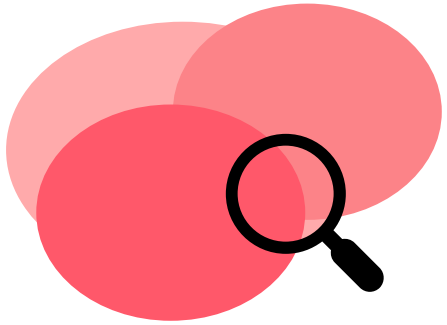
But.. which are the groups that need to be protected?

Impose fairness contraints

At least X for each gender, at least Y for each ethnicity

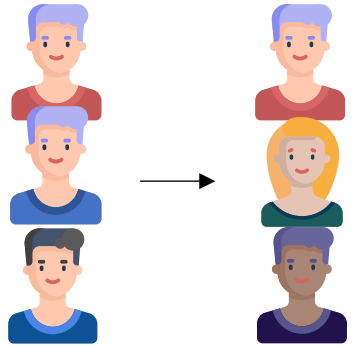
But.. how much representation is *enough*?

Our approach



Automatic identification of disadvantaged subgroups

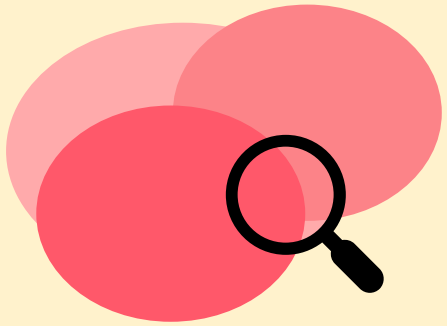
Via subgroup identification techniques



Mitigation of subgroup disadvantage

Based on their degree of disadvantage

Our approach



Automatic identification of disadvantaged subgroups

Via subgroup identification techniques

Automatic identification of disadvantaged subgroups

{gender=male , ethnicity=Cauc}

{gender=female}

{gender=female , ethnicity=Afr-Am}

- **Extract subgroups** over protected attributes via frequent pattern mining
  - Above frequency threshold to control the enumeration and statistical significance

Automatic identification of disadvantaged subgroups

{gender=male , ethnicity=Cauc}

{gender=female}

{gender=female , ethnicity=Afr-Am}

- **Extract subgroups** over protected attributes via frequent pattern mining
  - Above frequency threshold to control the enumeration and statistical significance
- **Compute their** degree of **disadvantage** (or advantage) in the ranking
  - Notion of subgroup divergence

Subgroup divergence Δ

$\gamma(c)$

$\gamma(g)$

g

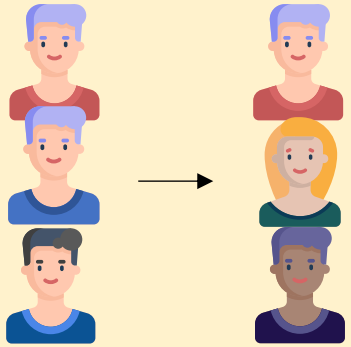
C: all candidates; g: a group;

$\gamma(c)$: **utility** of a candidate in the ranking, e.g., the score or ranking position

$\gamma(g)$: average utility of the candidates in group g

$$\Delta_\gamma(g) = \gamma(g) - \gamma(C)$$

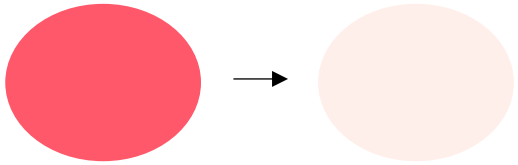
Disadvantaged group: $\Delta_\gamma(g) < 0$ & statistically significant



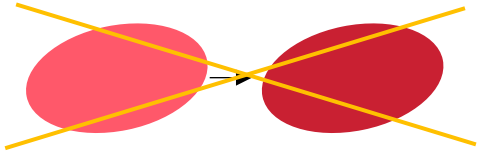
Mitigation of subgroup disadvantage

Based on their degree of disadvantage

Desired properties of the mitigation process



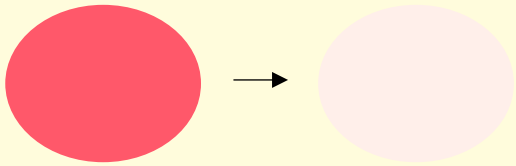
Mitigate the divergence of a subgroup g



Monotonicity constraint

Constant average overall behavior

Desired properties of the mitigation process



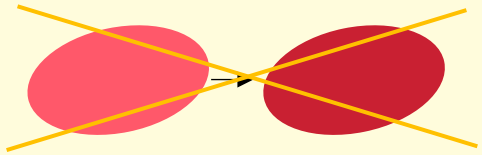
Mitigate the divergence of a subgroup g

$$\Delta_{\gamma'}(g) > \Delta_{\gamma}(g)$$

Where γ' are the scores for the new ranking r

Reduce the disadvantage

Desired properties of the mitigation process

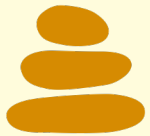


Monotonicity constraint

$$\min_{g \in G} \Delta_{\gamma'}(g) > \min_{g \in G} \Delta_{\gamma}(g)$$

Avoid that by mitigating the divergence of a subgroup, we worsen the condition of others

Desired properties of the mitigation process



Constant average overall behavior

$$\gamma(C) = \gamma'(C)$$

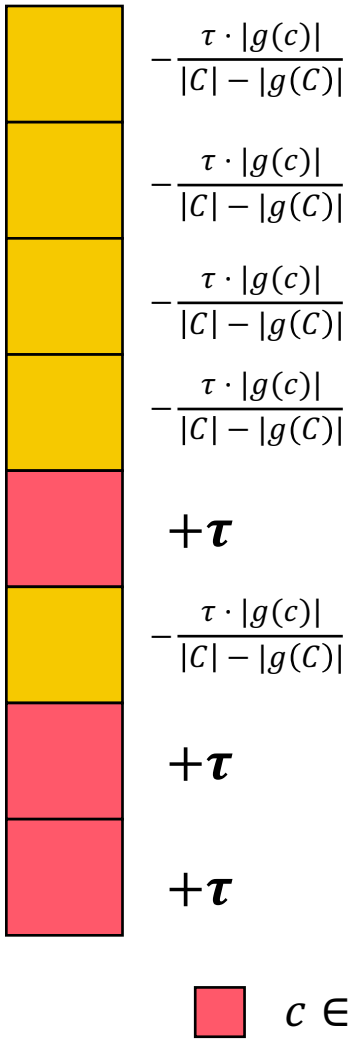
Overall ranking utility is maintained in the population

Step of mitigation for disadvantage group g

Update the scores γ' of the candidates $c \in C$ as follows

$$\gamma'(c) = \begin{cases} \gamma(c) + \tau & c \in g \\ \gamma(c) - \dfrac{\tau \cdot |g(c)|}{|C| - |g(C)|} & c \notin g \end{cases}$$

with $\tau \in \mathbb{R}_{>0}$



Step of mitigation for disadvantage group g

Update the scores γ' of the candidates $c \in C$ as follows

$$\gamma'(c) = \begin{cases} \gamma(c) + \tau & c \in g \\ \gamma(c) - \dfrac{\tau \cdot |g(c)|}{|C| - |g(C)|} & c \notin g \end{cases}$$

with $\tau \in \mathbb{R}_{>0}$

Counterbalance the increase of the scores
Satisfies **constant average overall behavior** property

Step of mitigation for disadvantage group g

Update the scores γ' of the candidates $c \in C$ as follows

$$\gamma'(c) = \begin{cases} \gamma(c) + \tau & c \in g \\ \gamma(c) - \dfrac{\tau \cdot |g(c)|}{|C| - |g(C)|} & c \notin g \end{cases}$$

with $\tau \in \mathbb{R}_{>0}$

For τ>0, $\gamma'(c) > \gamma(c) \rightarrow \Delta_{\gamma'}(g) > \Delta_{\gamma}(g)$

Implement the **mitigation of the divergence of a subgroup**

Step of mitigation for disadvantage group g

Update the scores γ' of the candidates $c \in C$ as follows

$$\gamma'(c) = \begin{cases} \gamma(c) + \tau & c \in g \\ \gamma(c) - \dfrac{\tau \cdot |g(c)|}{|C| - |g(C)|} & c \notin g \end{cases}$$

with $\tau \in \mathbb{R}_{>0}$

Does it satisfy the monotonicity property?
… Depends on τ

Ensuring the monotonicity property

For $c \notin g$, we decrease the score → we want to avoid decreasing it such that

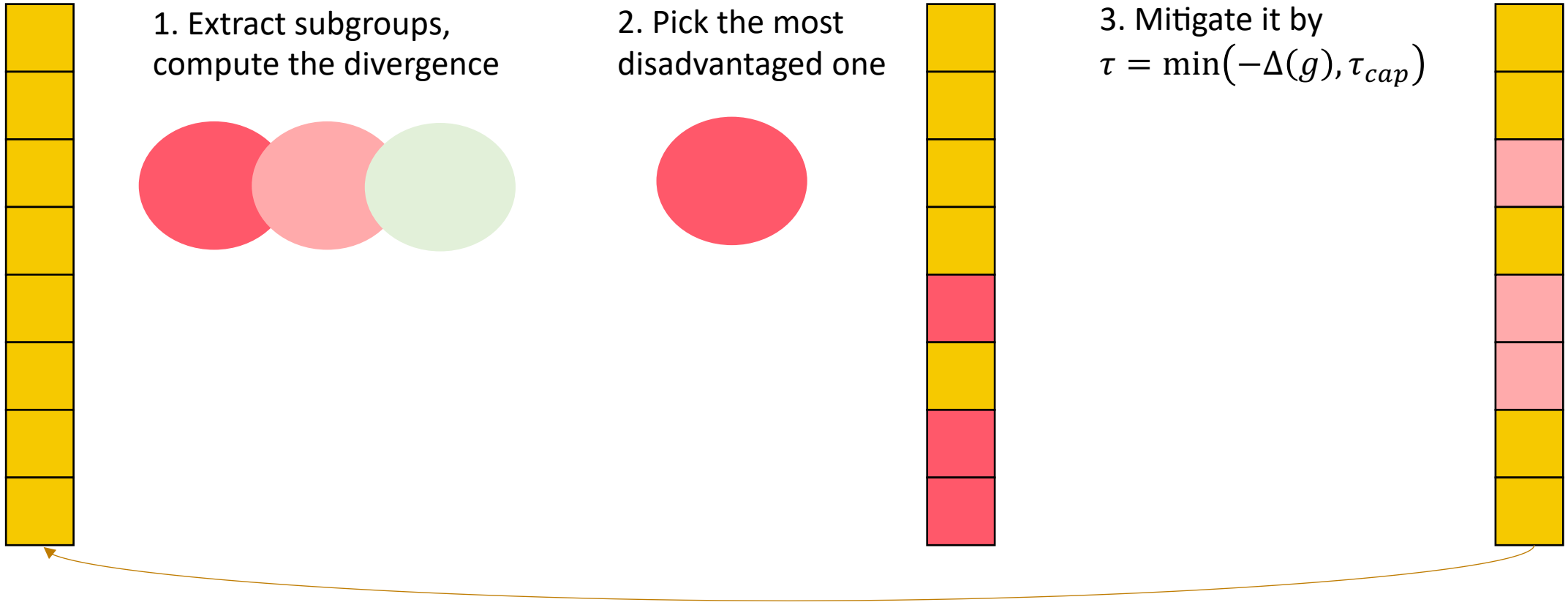
$$\min_{g \in G} \Delta_{\gamma}(g) > \min_{g \in G} \Delta_{\gamma'}(g)$$

We compute the maximum supported τ_{cap} for all extracted subgroups

$$\tau = \min\left(-\Delta(g), \tau_{cap}\right)$$

For a full mitigation

Iterative process



We repeat the mitigation step until
- No disadvantaged group
- No $\tau > 0$

Example for LSAT dataset

21,791 law students; we rank them by LSAT score

Gender and ethnicity as protected attributes

There are 11 disadvantage and 4 advantaged groups

Disadvantaged group	**Highest position**	**50% by position**	**Δ**
ethinicity=African-American, gender=female	777	20,163	-7,7
ethinicity=African-American	402	19,968	-7,35

Advantaged group	**Highest position**	**50% by position**	**Δ**
ethinicity=Caucasian, gender=male	1	9,690	0,96
ethinicity=Caucasian	1	10,028	0,76

Example for LSAT dataset - mitigation

Disadvantaged group	Δ_γ
ethinicity=African-American, gender=female	**-7,7**
ethinicity=African-American	**-7,35**

Advantaged group	Δ_γ
ethinicity=Caucasian, gender=male	0,96
ethinicity=Caucasian	0,76

Example for LSAT dataset - mitigation

Disadvantaged group	Δ_{γ}	$\Delta_{\gamma'}$ **1st iteration**
ethinicity=African-American, gender=female	**-7,7**	0
ethinicity=African-American	**-7,35**	**-2,66**

Advantaged group	Δ_{γ}	$\Delta_{\gamma'}$ **1st iteration**
ethinicity=Caucasian, gender=male	0,96	0,67
ethinicity=Caucasian	0,76	0,46

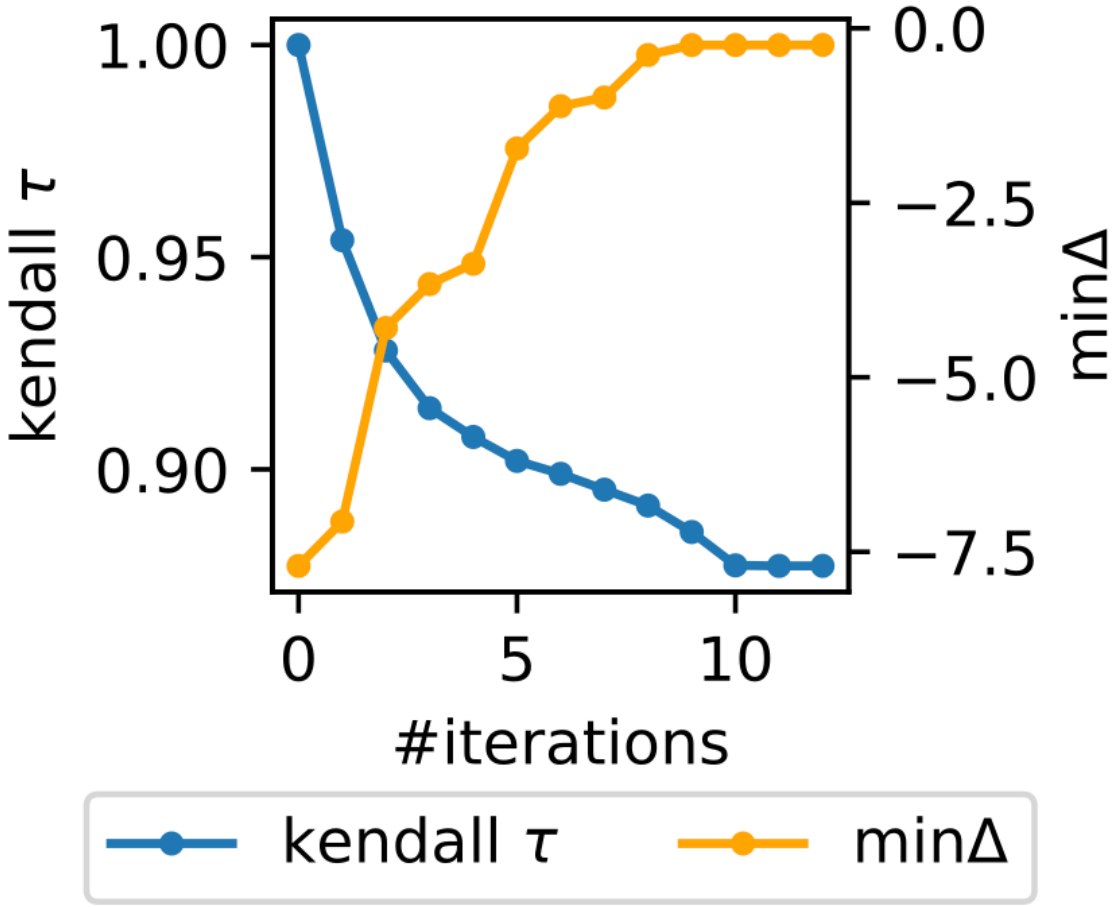
Example for LSAT dataset - mitigation

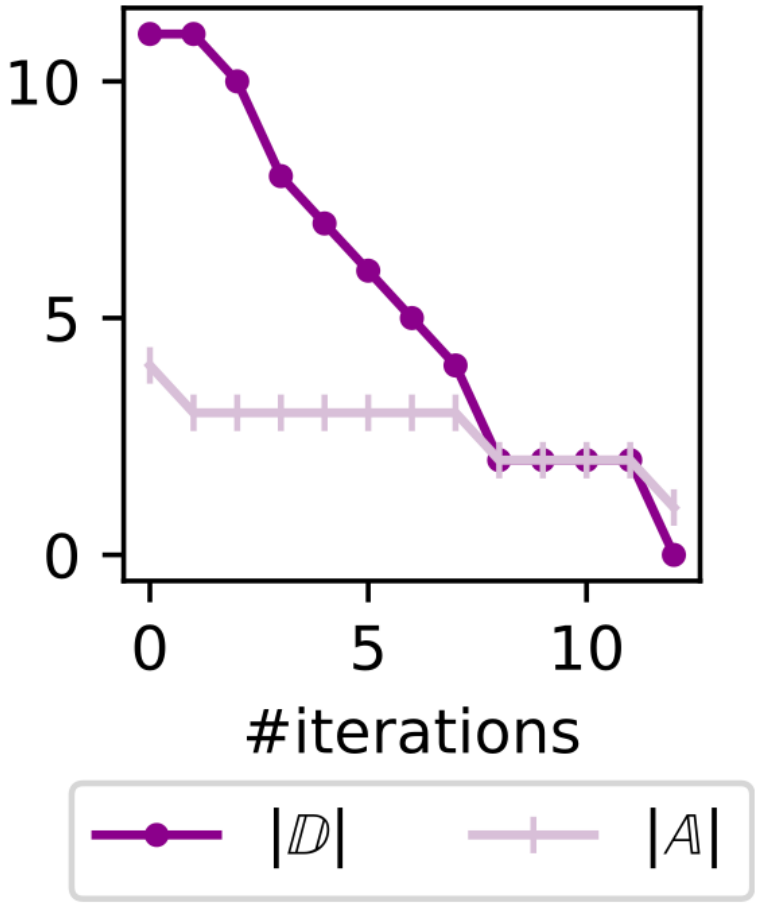
Disadvantaged group	Δ_{γ}	$\Delta_{\gamma'}$ **1st iteration**	$\Delta_{\gamma'}$ **mitigation**
ethinicity=African-American, gender=female	**-7,7**	0	0,15
ethinicity=African-American	**-7,35**	**-2,66**	0,02

Advantaged group	Δ_{γ}	$\Delta_{\gamma'}$ **1st iteration**	$\Delta_{\gamma'}$ **mitigation**
ethinicity=Caucasian, gender=male	0,96	0,67	0,15
ethinicity=Caucasian	0,76	0,46	0,06

Example for LSAT dataset – impact of mitigation

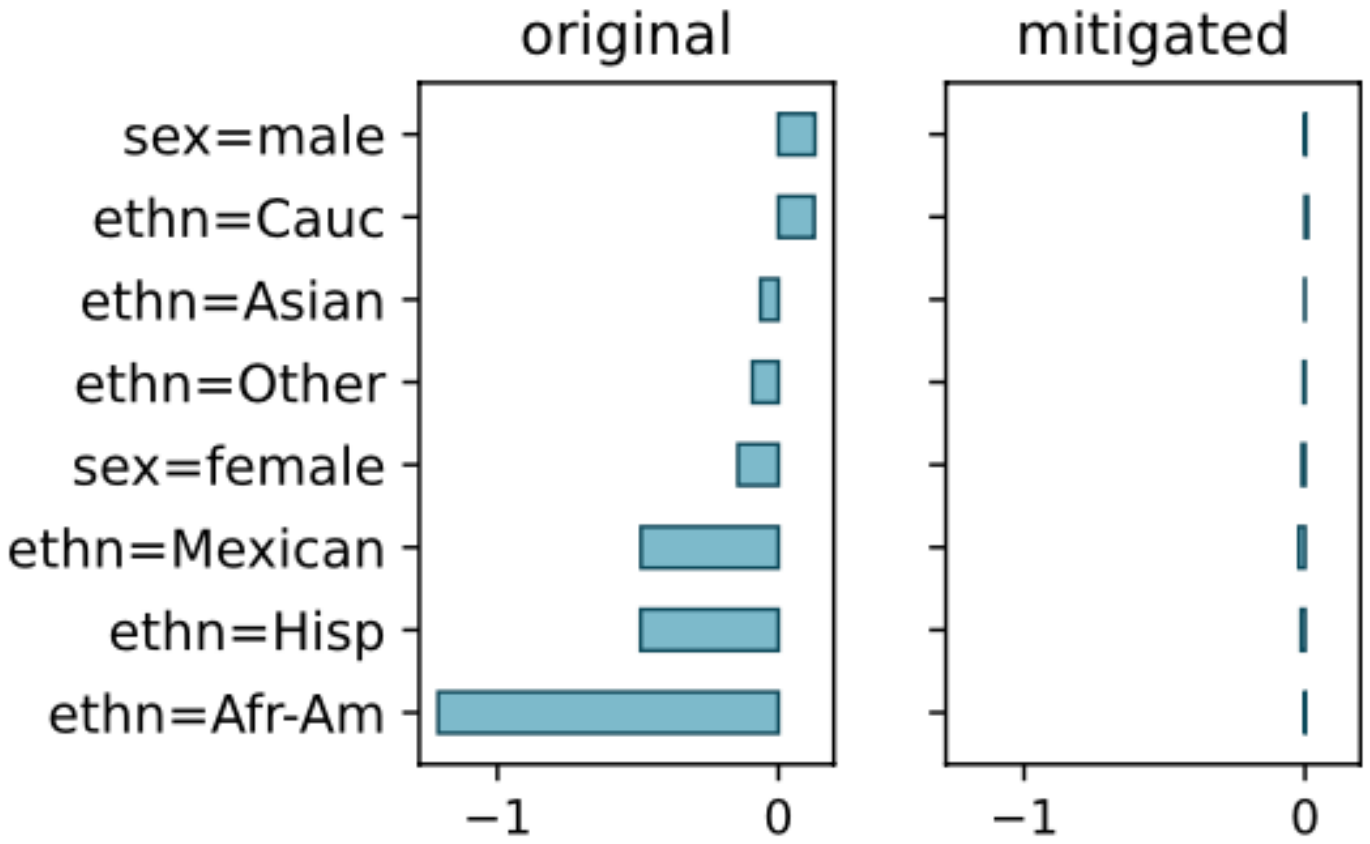
Min disadvantage, closeness to original ranking, dis&advantage groups





Example for LSAT dataset – impact of mitigation

Contribution to the divergence via Global Shapley value



Outline of the experimental results

5 real-world (LSAT, COMPAS, folktables, german credit, IIT-JEE) + 1 synthetic datasets

3 baseline approaches – also addressing intersectionality

Our approach always mitigates the disadvantages, reducing disadvantaged subgroups to 0 **with high closeness to the original ranking**

THANK YOU!


elianap.github.io/


eliana.pastor@polito.it


eliana__pastor