Fairness and Transparency in Rankings

Carlos Castillo / UPF

chato@acm.org

Partially based on
https://doi.org/10.1145/3308774.3308783

WSSC
Research Group on Web Science
and Social Computing

BIAS Workshop @ ECIR
April 2021

Algorithmic Bias in Rankings

Contents

1. Can algorithms discriminate?
2. Algorithmic fairness in IR
3. Measuring fairness in rankings
4. Creating fair rankings
5. Transparency in ranking

Generic discrimination

X discriminates against someone Y in relation to Z if:

1. Y has property P and Z does not have P
2. X treats Y worse than s/he treats or would treat Z
3. It is because Y has P and Z does not have P that X treats Y worse than Z

(also applies if X believes Y has P and Z does not have P)

Kasper Lippert-Rasmussen: Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination. Oxford University Press, 2013.

Generic discrimination

X discriminates against someone Y in relation to Z if:

1. Y has property P and Z does not have P
2. X treats Y worse than s/he treats or would treat Z
3. It is because Y has P and Z does not have P that X treats Y worse than Z

Disadvantageous differential treatment

Kasper Lippert-Rasmussen: Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination. Oxford University Press, 2013.

Group discrimination

X group-discriminates against Y in relation to Z if:

1. X generically discriminates against Y in relation to Z
2. P is the property of belonging to a socially salient group
3. This makes people with P worse off relative to others
   or X is motivated by animosity towards people with P,
   or by the belief that people with P are inferior
   or should not intermingle with others

Kasper Lippert-Rasmussen: Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination. Oxford University Press, 2013.

Statistical discrimination

X statistically discriminates against Y in relation to Z if:

1. X group-discriminates against Y in relation to Z
2. P is statistically relevant
   (or X believes P is statistically relevant)

Kasper Lippert-Rasmussen: Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination. Oxford University Press, 2013.

Example (statistical / non-statistical)

a. Not hiring a highly-qualified woman because the interviewer believes women have a higher probability of taking parental leave (statistical discrimination)

b. Not hiring a highly-qualified woman because she has said that she intends to have a child and take parental leave (non-statistical discrimination)

Kasper Lippert-Rasmussen: Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination. Oxford University Press, 2013.

In statistical machine learning

An algorithm developed through statistical machine learning can statistically discriminate if we:

1. Disregard intentions/animosity
2. Understand statistically relevant as
   any information derived from training data

Algorithmic Bias in Rankings

Contents

1. Can algorithms discriminate?
2. Algorithmic fairness in IR
3. Measuring fairness in rankings
4. Creating fair rankings
5. Transparency in ranking

Ranking in IR

Objective: provide maximum relevance to searche**r**

Order by decreasing probability of being relevant

However, we sometimes care about the searche**d** items

Carbonell, J., & Goldstein, J. (1998, August). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 335-336). ACM.

When searche**d** utility matters

Finding a local business	Business success
Purchasing a product or service	Marketing success
Recruiting a candidate for a job	Career success
Discovering events or groups to join	Social success
Learning about a political candidate	Political success
Dating/mating	Affective/reproductive success

Fairness for those searched is ...

1. A **sufficient presence** of elements of the protected group

   Absence of statistical group discrimination

   Prevent allocative (distributional) harms

1. A **consistent treatment** of elements of both groups

   Prevent individual discrimination

2. A **proper representation** of protected groups

   Prevent representational harms

Castillo, C. (2019, January). Fairness and Transparency in Ranking. In ACM SIGIR Forum (Vol. 52, No. 1, pp. 64-71). ACM.

≃ "P-fairness"

… and for searchers, it is

4. An **equal level of satisfaction** across searcher groups

   Due to different intents or different resp. to relevance

   Prevent allocative harms

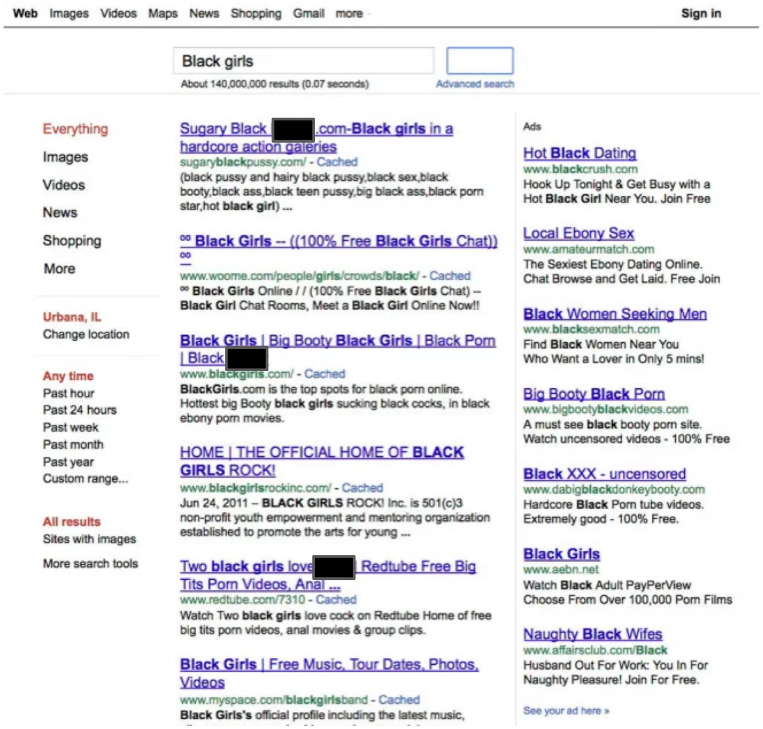
Mehrotra, R., Anderson, A., Diaz, F., Sharma, A., Wallach, H., & Yilmaz, E. (2017, April). Auditing search engines for differential satisfaction across demographics. In Proceedings of the 26th international conference on World Wide Web companion (pp. 626-633). IW3C2

≃ "C-fairness"

Representational harms

Representational harms occur when systems reinforce the subordination of some groups along the lines of identity (Kate Crawford)

- Sexualized search results
  Google ca. 2013, "black women" but in general "(race) women"



About 140,000,000 results (0.07 seconds) Advanced search

Sugary Black [redacted].com-Black girls in a hardcore action galeries
sugaryblackpussy.com/ - Cached
(black pussy and hairy black pussy,black sex,black booty,black ass,black teen pussy,big black ass,black porn star,hot black girl) ...

°° Black Girls -- ((100% Free Black Girls Chat)) °° –
www.woome.com/people/girls/crowds/black/ - Cached
°° Black Girls Online / / (100% Free Black Girls Chat) -- Black Girl Chat Rooms, Meet a Black Girl Online Now!!

Black Girls | Big Booty Black Girls | Black Porn | Black [redacted]
www.blackgirls.com/ - Cached
BlackGirls.com is the top spots for black porn online. Hottest big Booty black girls sucking black cocks, in black ebony porn movies.

HOME | THE OFFICIAL HOME OF BLACK GIRLS ROCK!
www.blackgirlsrockinc.com/ - Cached
Jun 24, 2011 – BLACK GIRLS ROCK! Inc. is 501(c)3 non-profit youth empowerment and mentoring organization established to promote the arts for young ...

Two black girls love [redacted] Redtube Free Big Tits Porn Videos, Anal ...
www.redtube.com/7310 - Cached
Watch Two black girls love cock on Redtube Home of free big tits porn videos, anal movies & group clips.

Black Girls | Free Music, Tour Dates, Photos, Videos
www.myspace.com/blackgirlsband - Cached
Black Girls's official profile including the latest music,

Ads

Hot Black Dating
www.blackcrush.com
Hook Up Tonight & Get Busy with a Hot Black Girl Near You. Join Free

Local Ebony Sex
www.amateurmatch.com
The Sexiest Ebony Dating Online. Chat Browse and Get Laid. Free Join

Black Women Seeking Men
www.blacksexmatch.com
Find Black Women Near You Who Want a Lover in Only 5 mins!

Big Booty Black Porn
www.bigbootyblackvideos.com
A must see black booty porn site. Watch uncensored videos - 100% Free

Black XXX - uncensored
www.dabigblackdonkeybooty.com
Hardcore Black Porn tube videos. Extremely good - 100% Free.

Black Girls
www.aebn.net
Watch Black Adult PayPerView Choose From Over 100,000 Porn Films

Naughty Black Wifes
www.affairsclub.com/Black
Husband Out For Work: You In For Naughty Pleasure! Join For Free.

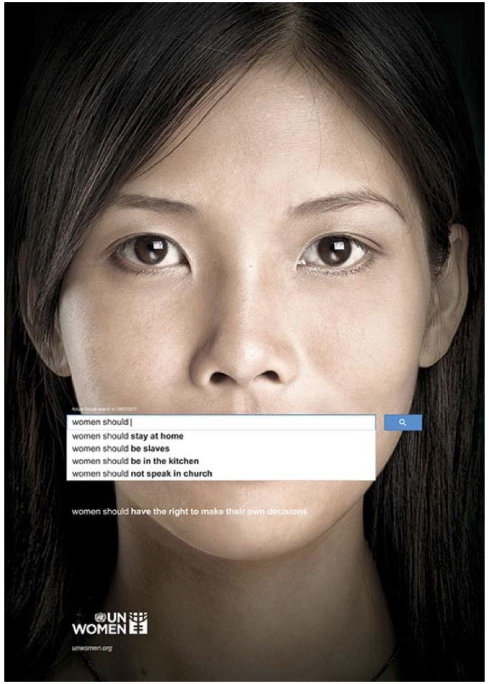
See your ad here »

Noble, S. U. (2018). Algorithms of Oppression: How search engines reinforce racism. NYU Press.
Crawford, K. (2017). The Trouble with Bias. Keynote at NIPS.

Representational harms (cont.)

Search suggestions reinforcing biases or stereotypes, spreading misinformation, manipulative, pointing to adult material, ...

- *{nationality|ethnicity|gender|...}* are [...]
- alexandria ocasio cortez [swimsuit]
- neil degrasse tyson [arrested]
- late term abortion [is never necessary]
- little girl in [miniskirt]


women should|
women should stay at home
women should be slaves
women should be in the kitchen
women should not speak in church
women should have the right to make their own decisions
UN WOMEN
unwomen.org

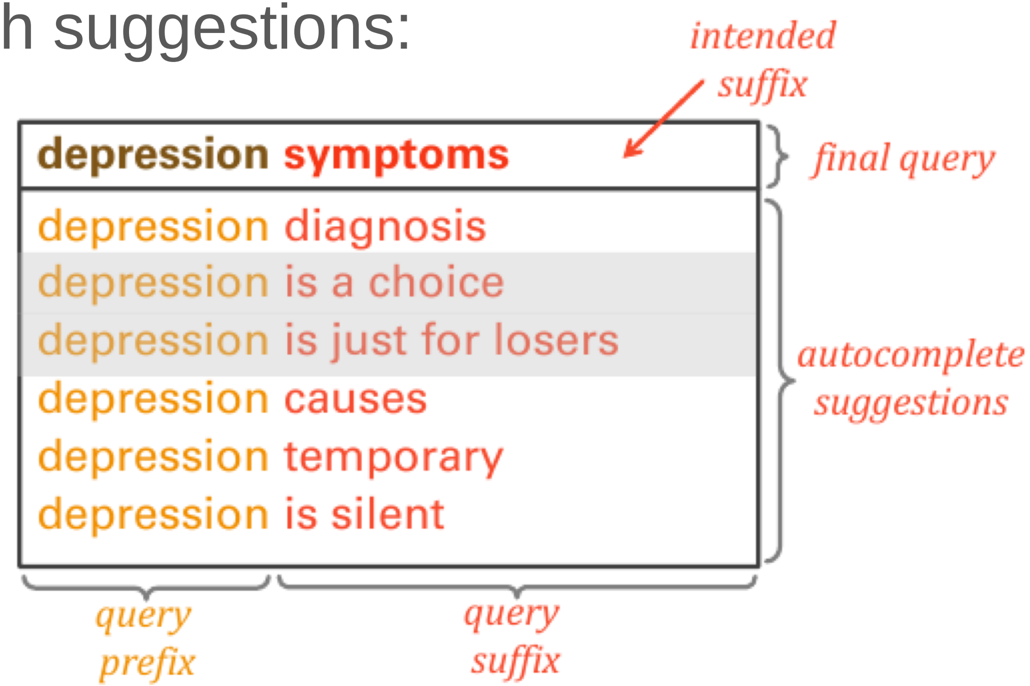
Olteanu, A., Diaz, F., & Kazai, G. (2020). When Are Search Completion Suggestions Problematic? *Proc. of CSCW*.

Baker, P., & Potts, A. (2013). Why do white people have thin lips? Google and the perpetuation of stereotypes via auto-complete search forms. *Critical discourse studies*.

Representational harms (cont.)

Types of problematic search suggestions:

- harmful speech
- potentially illicit
- misinformation
- stereotypes
- adult content
- ...



depression symptoms

depression diagnosis

depression is a choice

depression is just for losers

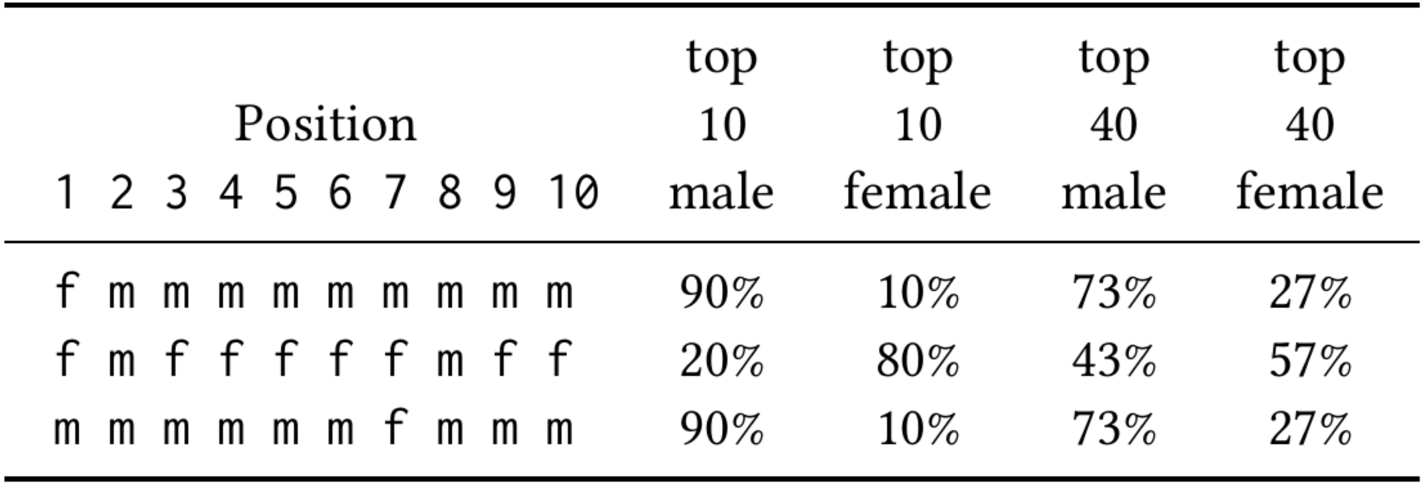
depression causes

depression temporary

depression is silent

Olteanu, A., Diaz, F., & Kazai, G. (2020). When Are Search Completion Suggestions Problematic? *Proc. of CSCW*.

Is this a *sufficient presence* of women?



	Position										top 10 male	top 10 female	top 40 male	top 40 female
	1	2	3	4	5	6	7	8	9	10				
Economist	f	m	m	m	m	m	m	m	m	m	90%	10%	73%	27%
Market analyst	f	m	f	f	f	f	f	m	f	f	20%	80%	43%	57%
Copywriter	m	m	m	m	m	m	f	m	m	m	90%	10%	73%	27%

Top-10 results for 3 professions in XING (a recruitment site, similar to LinkedIn, that is a market leader in Germany and Austria)

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). FA*IR: A fair top-k ranking algorithm. In Proc. of the ACM on Conference on Information and Knowledge Management (pp. 1569-1578). ACM.

Two different goals

Reduce discrimination when

a protected group has **higher utility but lower rankings**

E.g.: a university admittance test gives lower scores to economically disadvantaged applicants, but they have better academic performance if admitted

Provide equal opportunity when

a protected group has **lower utility and lower rankings**

E.g.: a university admittance test gives lower score to some applicants, who also have lower academic performance if admitted

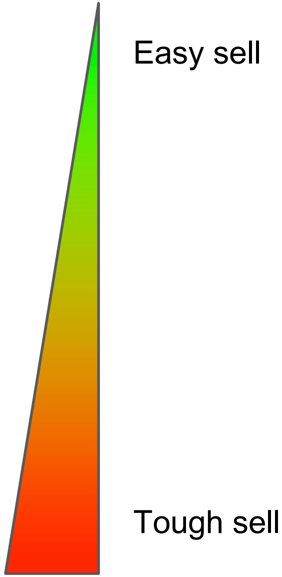
John E. Roemer (2000). Equality of Opportunity. Harvard University Press.

Making a case to create fair rankings

1. Biases harming searcher utility
   (i.e., reduce discrimination)

2. Legal mandates and voluntary commitments
   (i.e., provide equal opportunity)

3. Ensuring technology embodies certain values



Easy sell

Tough sell

Some possible biases in input data

Biases in expert-provided training data

Expert or editorially provided rankings

(e.g., all protected items ranked lower than nonprotected)

Biases in user-provided training data

Clicks and user feedback

(e.g., if women preferred ads for jobs that pay less)

Biases in document construction

(e.g., completion of different CV sections by men/women)

Olteanu, A., Castillo, C., Diaz, F., Kiciman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Frontiers in Big Data 2(13).

Contents

1. Can algorithms discriminate?
2. Algorithmic fairness in IR
3. Measuring fairness in rankings
4. Creating fair rankings
5. Transparency in ranking

Measuring fairness in rankings

Rank-weighted exposure

Singh and Joachims 2018, ...

Randomized merging (probability-based)

Yang and Stoyanovich 2017, Zehlike et al. 2017, ...

Pairwise comparisons

Kallus and Zhou 2019, Beutel et al. 2019, ...

Measuring fairness in rankings

Rank-weighted exposure

Singh and Joachims 2018, ...

C.f. “retrievability” concept, 10 years earlier:

Azzopardi, L., & Vinay, V.. Retrievability: An evaluation measure for higher order information access tasks. In *Proc. CIKM 2008*.

Randomized merging (probability-based)

Yang and Stoyanovich 2017, Zehlike et al. 2017, ...

Pairwise comparisons

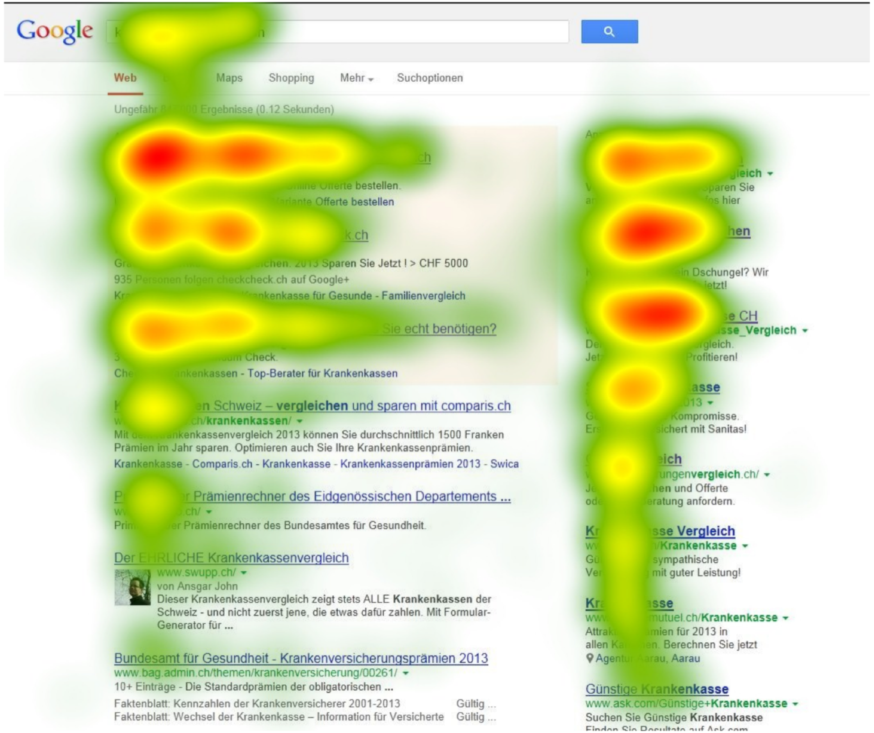
Kallus and Zhou 2019, Beutel et al. 2019, ...

Disparate exposure

Each position in a ranking has a certain value (e.g., probability of being examined) v_i

A ranking is fair if

$$\underset{i \in G_0}{E}(v_i) \simeq \underset{i \in G_1}{E}(v_i)$$



Google

Web Maps Shopping Mehr Suchoptionen

Ungefähr Ergebnisse (0.12 Sekunden)

Online Offerte bestellen.

Variante Offerte bestellen

k.ch

chen. 2013 Sparen Sie Jetzt ! > CHF 5000

935 Personen folgen checkcheck.ch auf Google+

Krankenkasse für Gesunde - Familienvergleich

Sie echt benötigen?

Check.

Krankenkassen - Top-Berater für Krankenkassen

en Schweiz – **vergleichen** und sparen mit comparis.ch

ch/**krankenkassen**/

Mit Krankenkassenvergleich 2013 können Sie durchschnittlich 1500 Franken Prämien im Jahr sparen. Optimieren auch Sie Ihre Krankenkassenprämien.

Krankenkasse - Comparis.ch - Krankenkasse - Krankenkassenprämien 2013 - Swica

Prämienrechner des Eidgenössischen Departements ...

.ch/

Prämienrechner des Bundesamtes für Gesundheit.

Der EHRLICHE Krankenkassenvergleich

www.swupp.ch/

von Ansgar John

Dieser Krankenkassenvergleich zeigt stets ALLE **Krankenkassen** der Schweiz - und nicht zuerst jene, die etwas dafür zahlen. Mit Formular-Generator für ...

Bundesamt für Gesundheit - Krankenversicherungsprämien 2013

www.bag.admin.ch/themen/krankenversicherung/00261/

10+ Einträge - Die Standardprämien der obligatorischen ...

Faktenblatt: Kennzahlen der Krankenversicherer 2001-2013 Gültig ...

Faktenblatt: Wechsel der Krankenkasse – Information für Versicherte Gültig ...

gleich

sparen Sie

Infos hier

hen

ein Dschungel? Wir

jetzt!

se CH

sse_Vergleich

rgleich.

Profitieren!

asse

2013

Kompromisse.

sichert mit Sanitas!

eich

rungen**vergleich**.ch/

hen und Offerte

Beratung anfordern.

sse Vergleich

/**Krankenkasse**

sympathische

mit guter Leistung!

asse

mutuel.ch/**Krankenkasse**

mien für 2013 in

allen Kantonen. Berechnen Sie jetzt

Agentur Aarau, Aarau

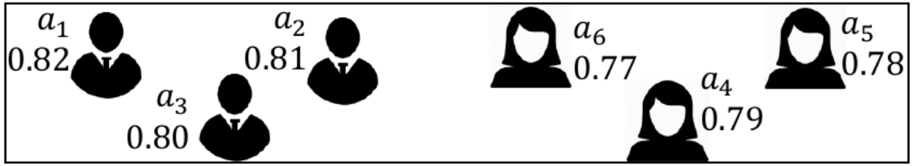
Günstige Krankenkasse

www.ask.com/Günstige+**Krankenkasse**

Suchen Sie Günstige **Krankenkasse**

Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2219-2228). ACM.

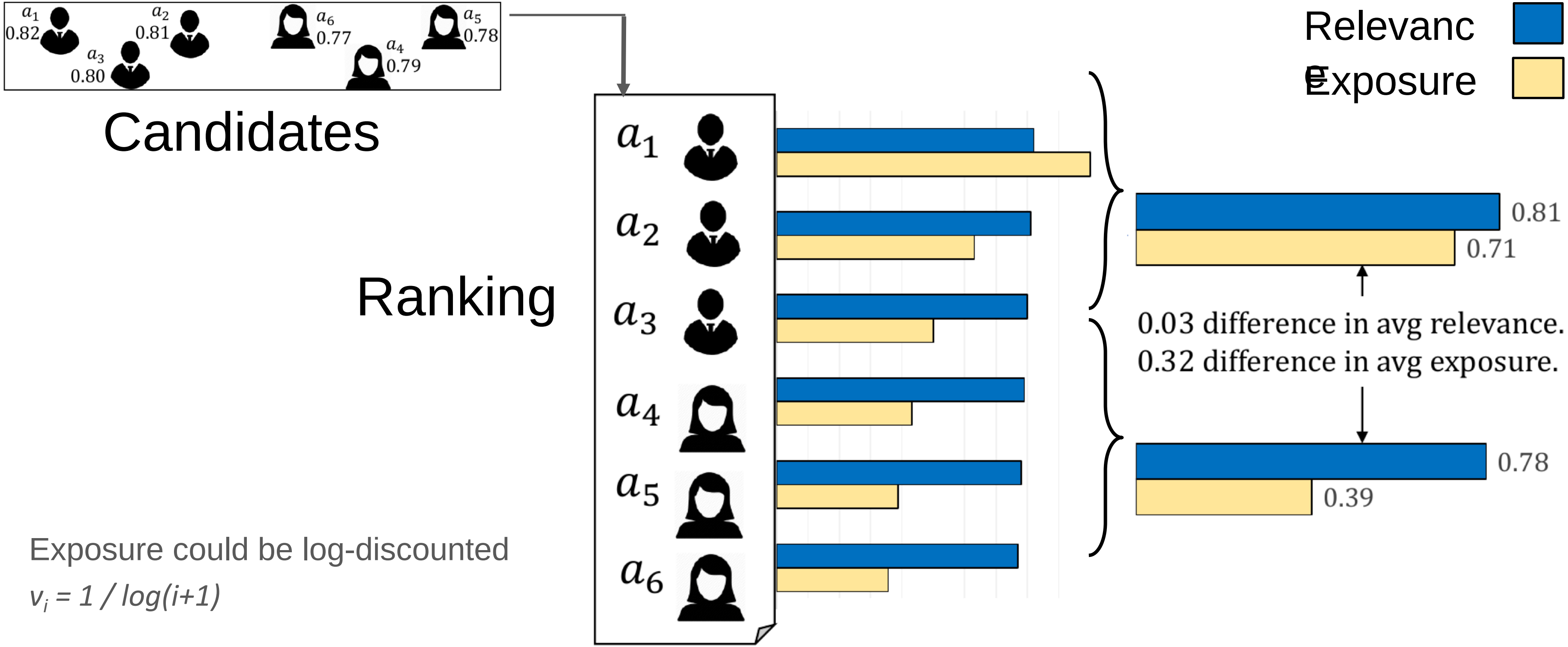
Disparate exposure: example



Candidates
(and their relevance)

Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2219-2228). ACM.

Disparate exposure: example



Exposure could be log-discounted

v_i = 1 / log(i+1)

Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2219-2228). ACM.

Disparate exposure

Utility-normalized exposure disparity ("Disparate Treatment Ratio"):

$$\mathrm{DTR}(G_0, G_1|\mathbf{P}, q) = \frac{\mathrm{Exposure}(G_0|\mathbf{P})/\mathrm{U}(G_0|q)}{\mathrm{Exposure}(G_1|\mathbf{P})/\mathrm{U}(G_1|q)}$$

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

Expected click-through rate disparity ("Disparate Impact Ratio"):

$$\mathrm{DIR}(G_0, G_1|\mathbf{P}, q) = \frac{\mathrm{CTR}(G_0|\mathbf{P})/\mathrm{U}(G_0|q)}{\mathrm{CTR}(G_1|\mathbf{P})/\mathrm{U}(G_1|q)}$$

$$\mathrm{CTR}(G_k|\mathbf{P}) = \frac{1}{|G_k|} \sum_{i \in G_k} \sum_{j=1}^{N} \mathbf{P}_{i,j} \mathbf{u}_i \mathbf{v}_j.$$

Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2219-2228). ACM.

Amortized fairness

Every element should receive attention or exposure (a_i) proportional to its utility (r_i)

$$\frac{\sum_{l=1}^{m} a_{i1}^{l}}{\sum_{l=1}^{m} r_{i1}^{l}} = \frac{\sum_{l=1}^{m} a_{i2}^{l}}{\sum_{l=1}^{m} r_{i2}^{l}}, \forall u_{i1}, u_{i2}.$$

This should be achieved across *m* queries

At every query, consider past accumulated attention/utility deficits or surpluses, and correct them to the extent possible while honoring quality constraints

Biega, A. J., Gummadi, K. P., & Weikum, G. (2018). Equity of Attention: Amortizing Individual Fairness in Rankings. Proc. of SIGIR.

More variants

Inverse log-weighted KL divergence of prefixes
[Geyik et al. KDD 2019]

...

Measuring fairness in rankings

Rank-weighted exposure

Singh and Joachims 2018, ...

Randomized merging (probability-based)

Yang and Stoyanovich 2017, Zehlike et al. 2017, ...

Pairwise comparisons

Kallus and Zhou 2019, Beutel et al. 2019 ...

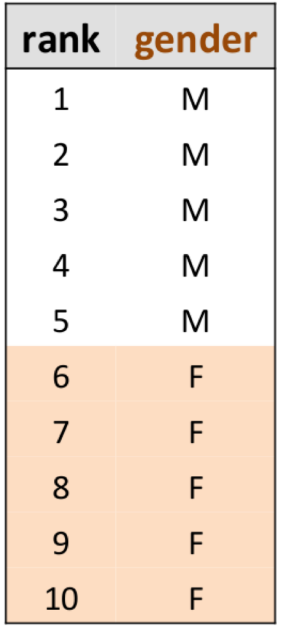
Ranking as randomized merging

1. Rank protected and unprotected separately

2. For each position:

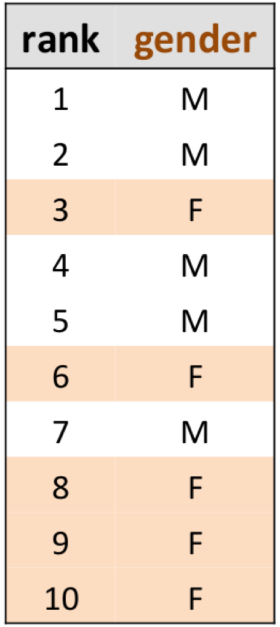
- Pick protected with probability *p*
- Pick nonprotected with probability *1-p*

Continue until exhausting both lists



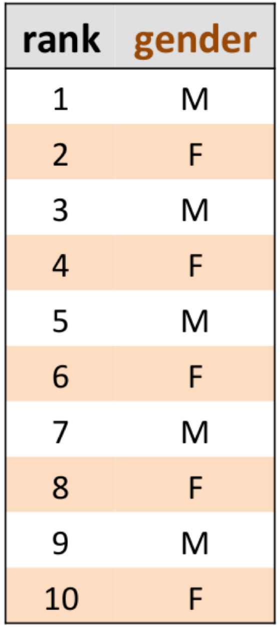
rank	gender
1	M
2	M
3	M
4	M
5	M
6	F
7	F
8	F
9	F
10	F

p=0



rank	gender
1	M
2	M
3	F
4	M
5	M
6	F
7	M
8	F
9	F
10	F

p=0.3



rank	gender
1	M
2	F
3	M
4	F
5	M
6	F
7	M
8	F
9	M
10	F

p=0.5

Yang, K., & Stoyanovich, J. (2017). Measuring fairness in ranked outputs. In Proc. of the 29th International Conference on Scientific and Statistical Database Management (p. 22). ACM.

Fair representation condition

Given parameters p, α and a set of size k

Let $F(x;p,k)$ be the cumulative distribution function of a binomial distribution with parameters p, k

A ranking of k elements having x protected elements satisfies the **fair representation condition** with probability p and significance α if $F(x;p,k) > \alpha$

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). FA*IR: A fair top-k ranking algorithm. In Proc. of the ACM on Conference on Information and Knowledge Management (pp. 1569-1578). ACM.

Example: fair representation condition

Suppose $p=0.5$, $k=10$, $\alpha=0.10$

$F(1, 0.5, 10) = 0.01 < 0.10 \Rightarrow$ if 1 protected element, fail

$F(2, 0.5, 10) = 0.05 < 0.10 \Rightarrow$ if 2 protected elements, fail

$F(3; 0.5, 10) = 0.17 > 0.10 \Rightarrow$ if 3 protected elements, pass

$F(4; 0.5, 10) = 0.37 > 0.10 \Rightarrow$ if 4 protected elements, pass

...

Ranked group fairness (unadjusted)

Given parameters p, α and a list of size k

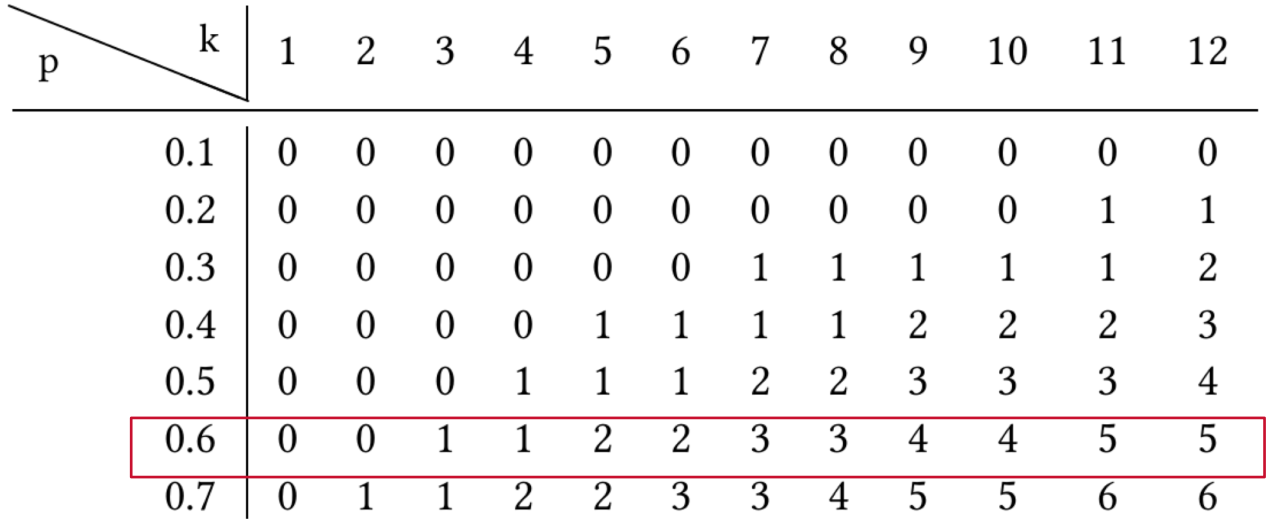
The list satisfies the **ranked group fairness** condition if

for every $i \leq k$

the prefix of size i of the list
satisfies the fair representation condition (i, p, α)

Examples: ranked group fairness

Can be expressed with a vector



p \ k	1	2	3	4	5	6	7	8	9	10	11	12
0.1	0	0	0	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0	0	1	1
0.3	0	0	0	0	0	0	1	1	1	1	1	2
0.4	0	0	0	0	1	1	1	1	2	2	2	3
0.5	0	0	0	1	1	1	2	2	3	3	3	4
0.6	0	0	1	1	2	2	3	3	4	4	5	5
0.7	0	1	1	2	2	3	3	4	5	5	6	6

Problem: **multiple hypothesis testing**

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). FA*IR: A fair top-k ranking algorithm. In Proc. of the ACM on Conference on Information and Knowledge Management (pp. 1569-1578). ACM.

Ranked group fairness (adjusted)

Given parameters p, α and a list of size k

The list satisfies the **ranked group fairness** condition if

for every $i \leq k$

the prefix of size i of the list

satisfies the fair representation condition (i, p, α_c)

Where $\alpha_c > \alpha$ is adjusted to make the failure probability of a ranking generated by randomized merging equal to α

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). FA*IR: A fair top-k ranking algorithm. In Proc. of the ACM on Conference on Information and Knowledge Management (pp. 1569-1578). ACM.

Probability-based measure

Given a ranking of *k* elements …

… and a significance *α*:

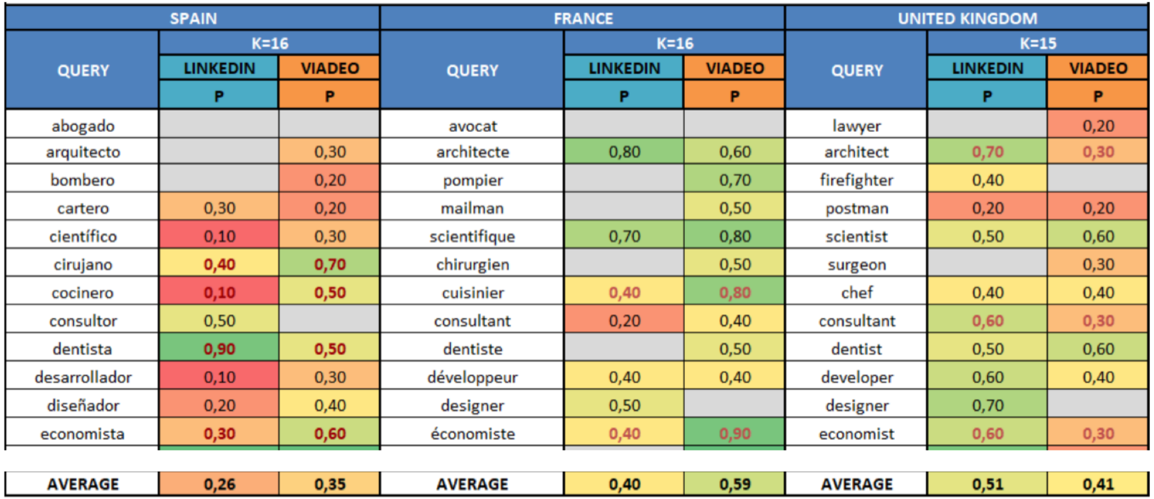
its **ranked group fairness is the maximum *p*** such that the ranking passes the ranked group fairness at *p*, *α*

… and a probability *p*:

its ranked group fairness is the minimum *α* such that the ranking passes the ranked group fairness at *p*, *α*

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). FA*IR: A fair top-k ranking algorithm. In Proc. of the ACM on Conference on Information and Knowledge Management (pp. 1569-1578). ACM.

Example: job search



SPAIN			FRANCE			UNITED KINGDOM		
QUERY	K=16		QUERY	K=16		QUERY	K=15	
	LINKEDIN	VIADEO		LINKEDIN	VIADEO		LINKEDIN	VIADEO
	P	P		P	P		P	P
abogado			avocat			lawyer		0,20
arquitecto		0,30	architecte	0,80	0,60	architect	0,70	0,30
bombero		0,20	pompier		0,70	firefighter	0,40	
cartero	0,30	0,20	mailman		0,50	postman	0,20	0,20
científico	0,10	0,30	scientifique	0,70	0,80	scientist	0,50	0,60
cirujano	0,40	0,70	chirurgien		0,50	surgeon		0,30
cocinero	0,10	0,50	cuisinier	0,40	0,80	chef	0,40	0,40
consultor	0,50		consultant	0,20	0,40	consultant	0,60	0,30
dentista	0,90	0,50	dentiste		0,50	dentist	0,50	0,60
desarrollador	0,10	0,30	développeur	0,40	0,40	developer	0,60	0,40
diseñador	0,20	0,40	designer	0,50		designer	0,70	
economista	0,30	0,60	économiste	0,40	0,90	economist	0,60	0,30
AVERAGE	0,26	0,35	AVERAGE	0,40	0,59	AVERAGE	0,51	0,41

There are large differences in the presence of women across professions, countries *and platforms*

Plus: treatment of masculine as *neutral* gender in queries in Spanish and French is inconsistent across and within platforms

Sara Galindo: "Evaluating potential biases in commercial people search engines". MSc Thesis, UPF, July 2019.
Data: https://github.com/sgalinma/job-search-discrimination-data

Measuring fairness in rankings

Rank-weighted exposure

Singh and Joachims 2018, ...

Randomized merging (probability-based)

Yang and Stoyanovich 2017, Zehlike et al. 2017, ...

Pairwise comparisons ⇦

Kallus and Zhou 2019, Beutel et al. 2019 ...

Cross-AUC (xAUC, ΔxAUC)

If R_1 is the ranking of a relevant item and R_0 the ranking of an irrelevant item:

$AUC = Pr[R_1 > R_0]$

Pr[Relevant item ranked above irrelevant item]

The cross-AUC between groups a and b is defined as:

$xAUC = Pr[R^a_1 > R^b_0]$

$\Delta xAUC = Pr[R^a_1 > R^b_0] - Pr[R^b_1 > R^a_0]$

Kallus, Nathan, and Angela Zhou. "The fairness of risk scores beyond classification: Bipartite ranking and the xAUC metric." In Advances in Neural Information Processing Systems, pp. 3438-3448. 2019.

Pairwise success

If $R^a_1 > R^b_1$ are the rankings of two relevant items from different groups:

- If $\text{clicks}(R^a_1) > \text{clicks}(R^b_1)$ then we count a success
- Otherwise, we count a failure

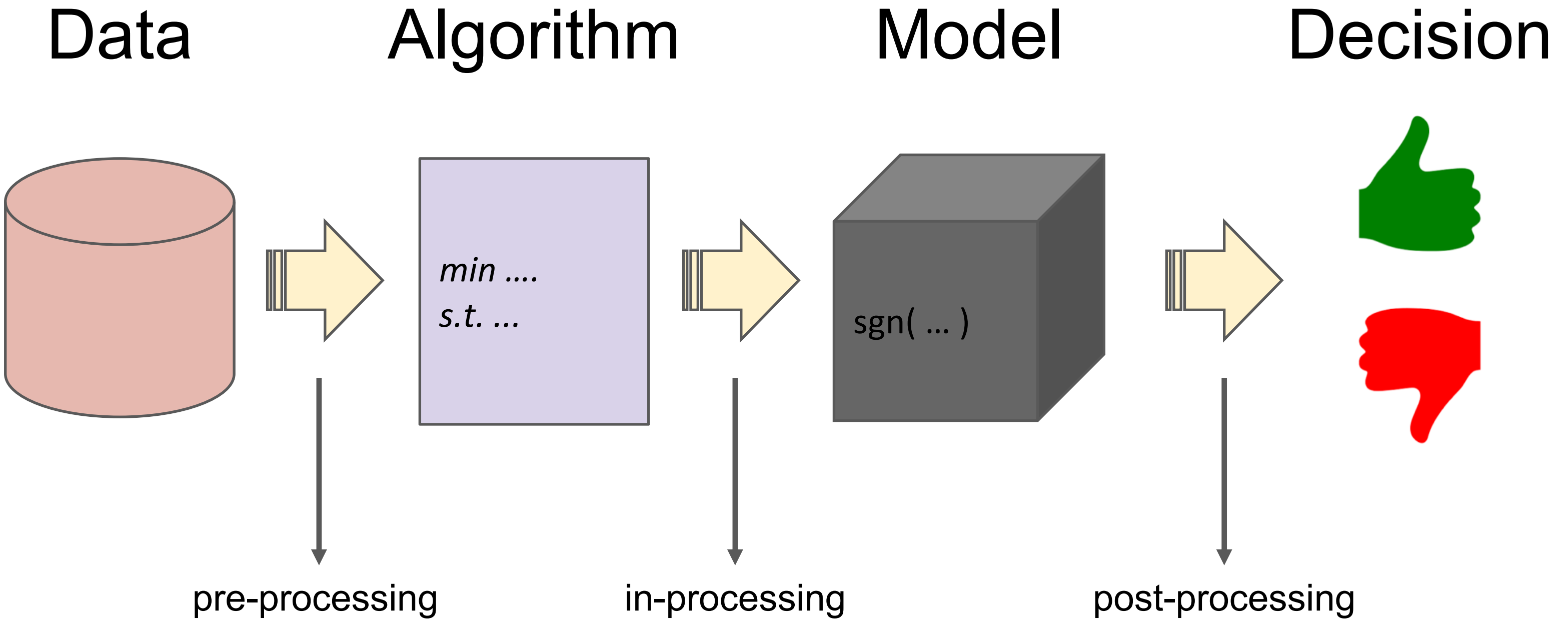
A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, C. Goodrow (2019). Fairness in Recommendation Ranking through Pairwise Comparisons. arXiv:1903.00780.

Algorithmic Bias in Rankings

Contents

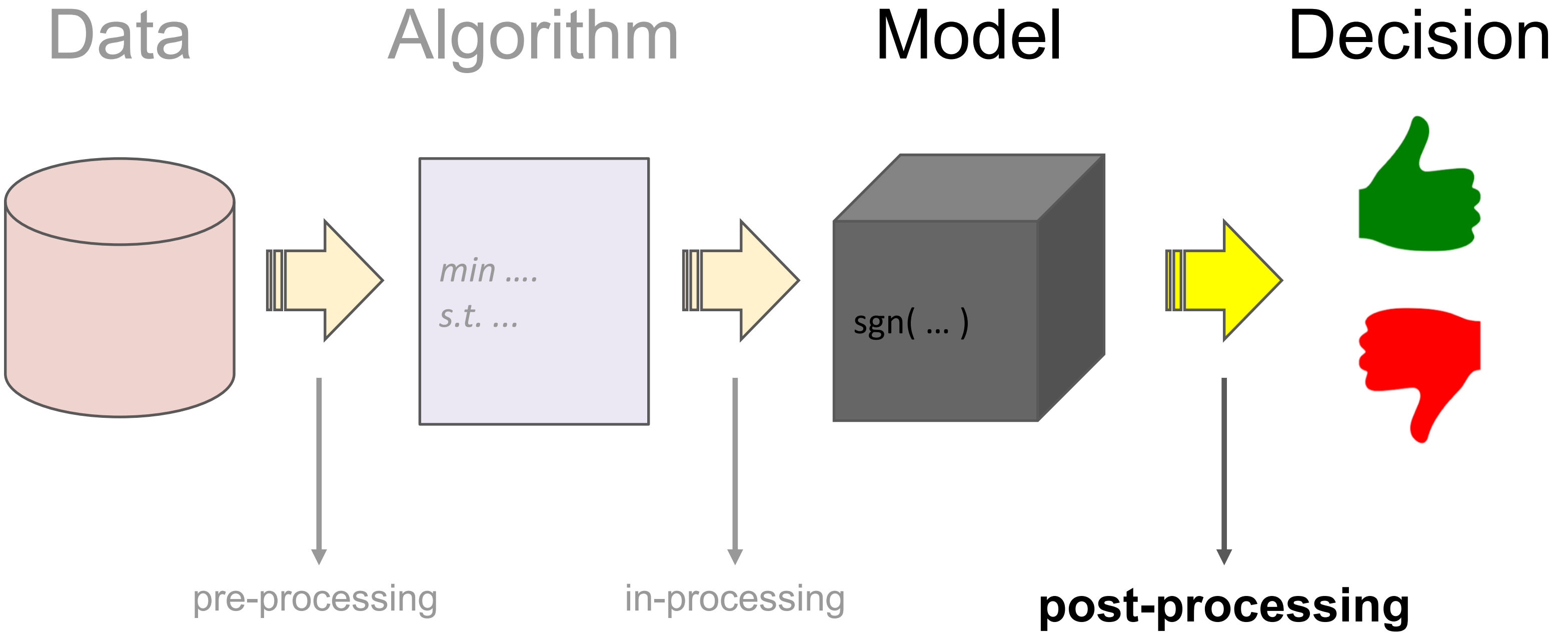
1. Can algorithms discriminate?
2. Algorithmic fairness in IR
3. Measuring fairness in rankings
4. ➡ Creating fair rankings
5. Transparency in ranking

Fairness: (pre,post,in)-processing



Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2125-2126). ACM.

Post-processing methods



Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2125-2126). ACM.

Single protected attribute

Rank separately protected P and nonprotected N

Determine the *minimum number* of protected elements required at every ranking position using *p*, *α*

For every position

```
    **If** *enough* protected elements**:** pick next from best of P, N
        **else**: pick next from P
```

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R. (2017). FA*IR: A fair top-k ranking algorithm. In Proc. of the ACM on Conference on Information and Knowledge Management (pp. 1569-1578). ACM.

Multiple protected attribs (Celis et al.)

$$\arg\max_{x \in R_{m,n}} \sum_{i \in [m], j \in [n]} W_{ij} x_{ij} \qquad \text{s.t.} \quad L_{k\ell} \leq \sum_{1 \leq j \leq k} \sum_{i \in P_\ell} x_{ij} \leq U_{k\ell} \qquad \forall\ \ell \in [p], k \in [n]$$

x_{ij} is whether we place item *i* in position *j*

$R_{m,n}$ is the constraint that each item goes in one position only

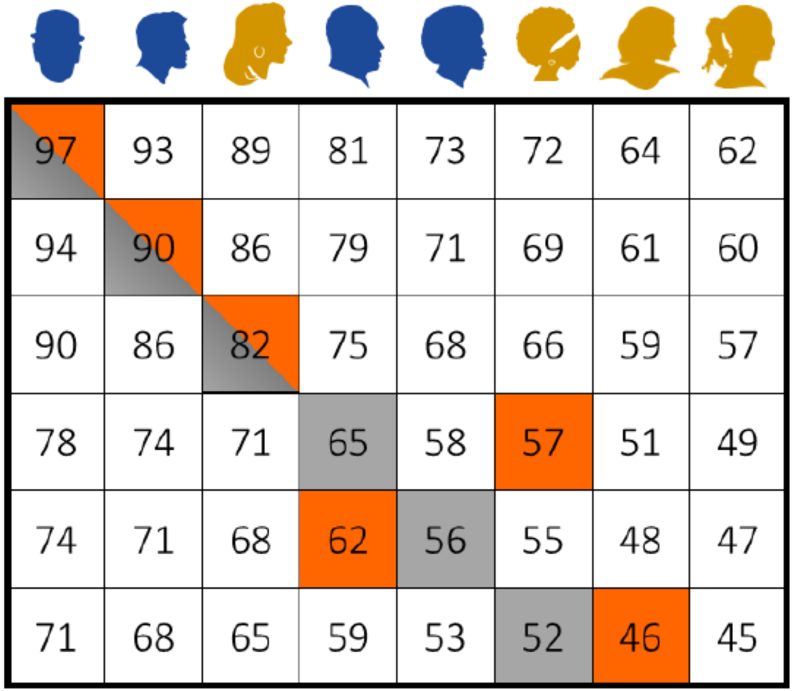
W_{ij} is the utility of placing in position *i* the item *j* (non-decr.)

U_{kl} is the **given** max. number of items of class *l* up to pos *k*

Celis, L. E., Straszak, D., & Vishnoi, N. K. (2018). Ranking with fairness constraints. In *Proc. of 45th International Colloquium on Automata, Languages, and Programming (pp. 28:1-28:15).*

Example (Celis et al.)

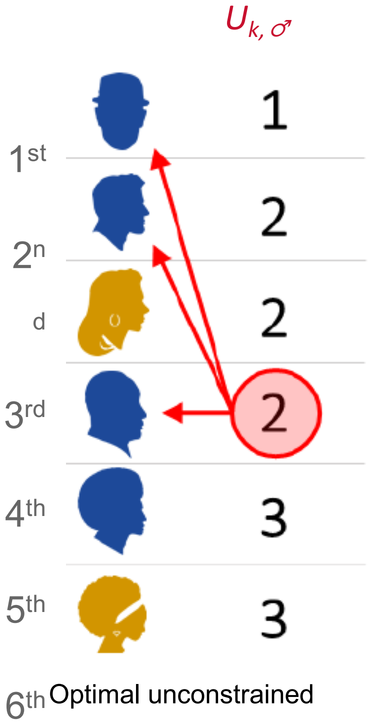
W_{ij}



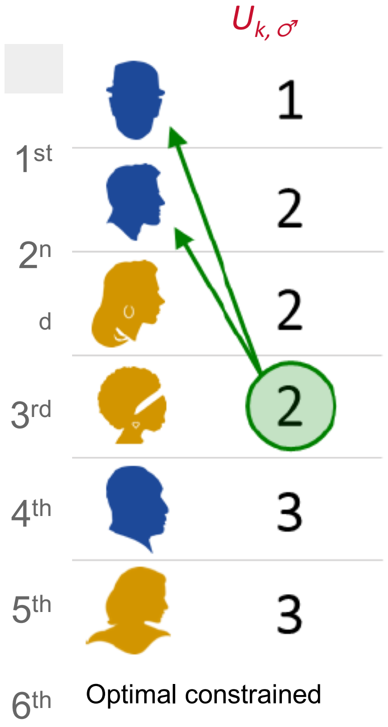
1st	97	93	89	81	73	72	64	62
2n	94	90	86	79	71	69	61	60
d	90	86	82	75	68	66	59	57
3rd	78	74	71	65	58	57	51	49
4th	74	71	68	62	56	55	48	47
5th	71	68	65	59	53	52	46	45

6th

▨ Optimal unconstrained 🟧 Optimal constrained



Optimal unconstrained



Optimal constrained

Celis, L. E., Straszak, D., & Vishnoi, N. K. (2018). Ranking with fairness constraints. In *Proc. of 45th International Colloquium on Automata, Languages, and Programming (pp. 28:1-28:15).*

Results in Celis et al.

Let Δ = max. number of constrained attributes of an element

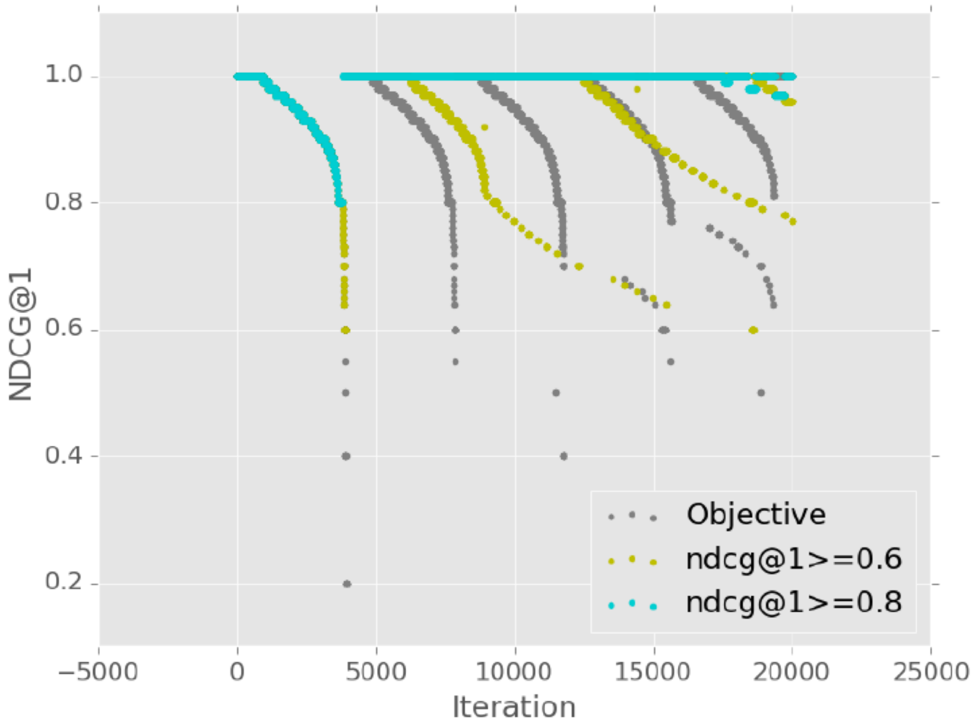
If $\Delta = 1$: solvable in polynomial time

If $\Delta > 1$: approximately solvable in polynomial time
using an LP relaxation, violates constraints
by at most a $(\Delta+2)$ factor

Celis, L. E., Straszak, D., & Vishnoi, N. K. (2018). Ranking with fairness constraints. In *Proc. of 45th International Colloquium on Automata, Languages, and Programming (pp. 28:1-28:15).*

Amortized fairness

Change elements at top positions to ensure enough exposure is given to different groups



Biega, A. J., Gummadi, K. P., & Weikum, G. (2018). Equity of Attention: Amortizing Individual Fairness in Rankings. Proc. of SIGIR.

Singh and Joachims

Probabilistic ranking ***P***

$P_{i,j}$ is probability of placing document *i* in position *j*

$$\mathrm{U}(\mathbf{P}|q) = \sum_{d_i \in \mathcal{D}} \sum_{j=1}^{N} \mathbf{P}_{i,j}\, u(d_i|q)\, \mathbf{v}_j$$

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

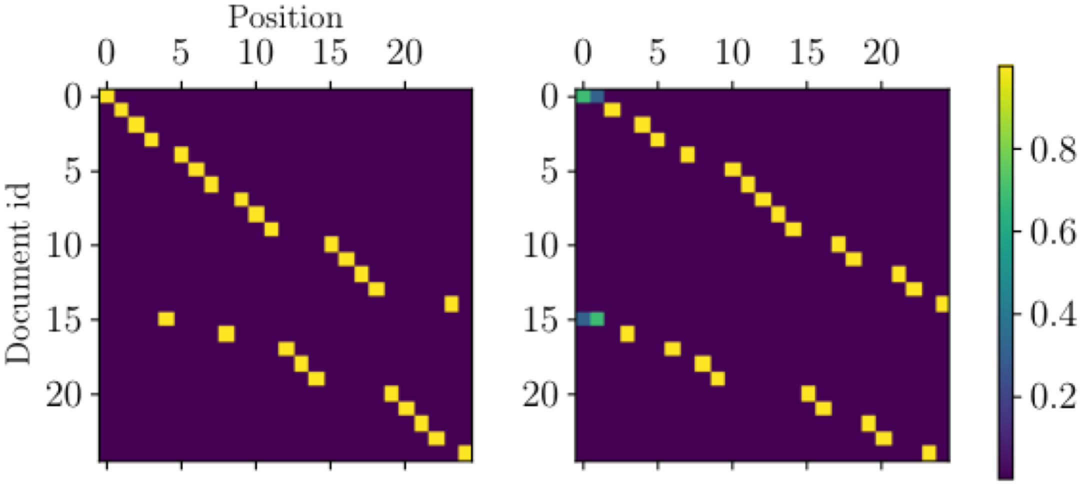
Maximize utility and reduce DTR and DIR

(utility-normalized exposure or predicted click-through rates)

Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2219-2228). ACM.

Singh and Joachims (cont.)

Experimental results: (a) unconstrained and (b) fair ranking

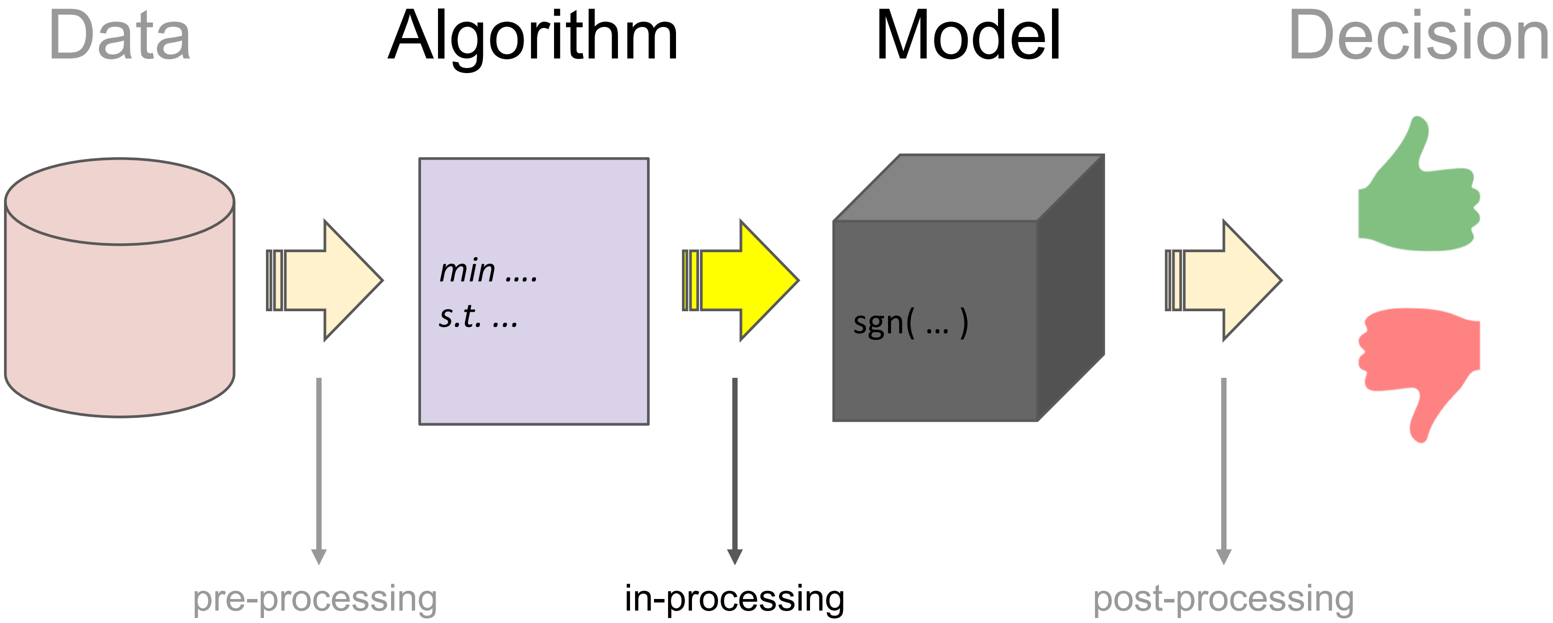


(a) DCG=5.2027

(b) DCG=5.1360

Singh, A., & Joachims, T. (2018). Fairness of Exposure in Rankings. In Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2219-2228). ACM.

In-processing methods



Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2125-2126). ACM.

Listwise LTR method

Optimize LTR with a combination of two losses:

- *L* = loss due to difference between ranking predictions and training elements
- *U* = loss due to expected different exposure

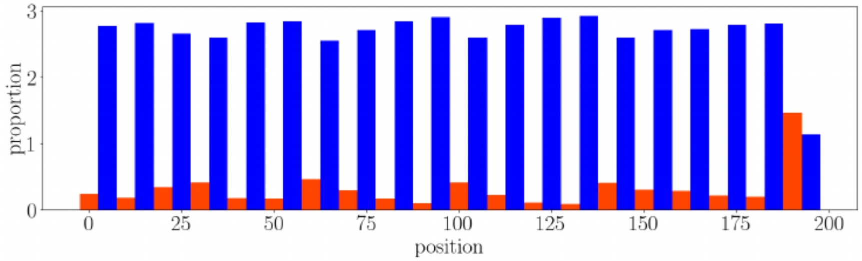
$$L_{DELTR}\left(y^{(q)}, \hat{y}^{(q)}\right) = L\left(y^{(q)}, \hat{y}^{(q)}\right) + \gamma U\left(\hat{y}^{(q)}\right)$$

$$U(\hat{y}^{(q)}) = \max\left(0, \mathsf{Exposure}(G_0|P_{\hat{y}^{(q)}}) - \mathsf{Exposure}(G_1|P_{\hat{y}^{(q)}})\right)^2$$

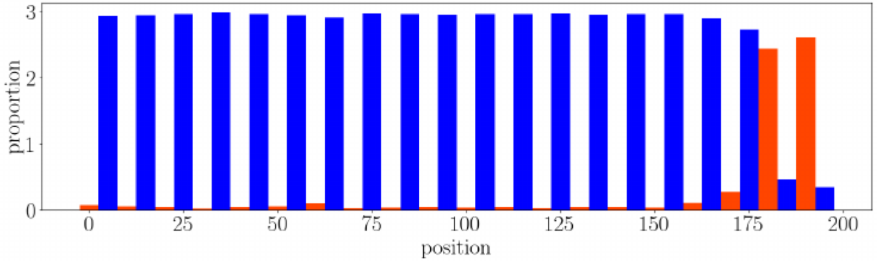
Zehlike, M., and Castillo, C. (2020). Reducing Disparate Exposure in Ranking: A Learning To Rank Approach. WWW 2020 (Short) 2018 pre-print.

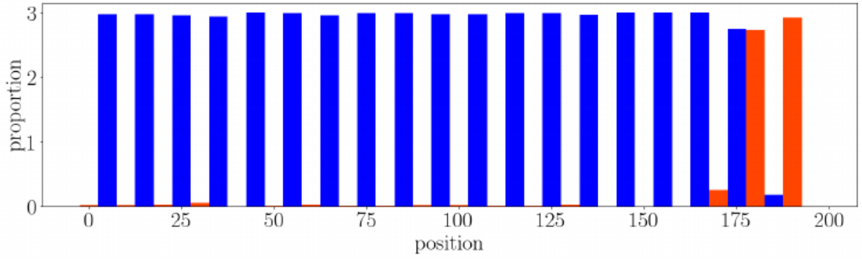
DELTR: W3C Corpus (TREC Expert)

"Color-blind"

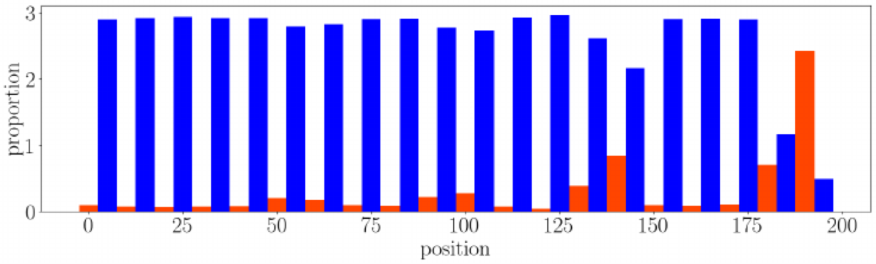


DELTR (small gamma)





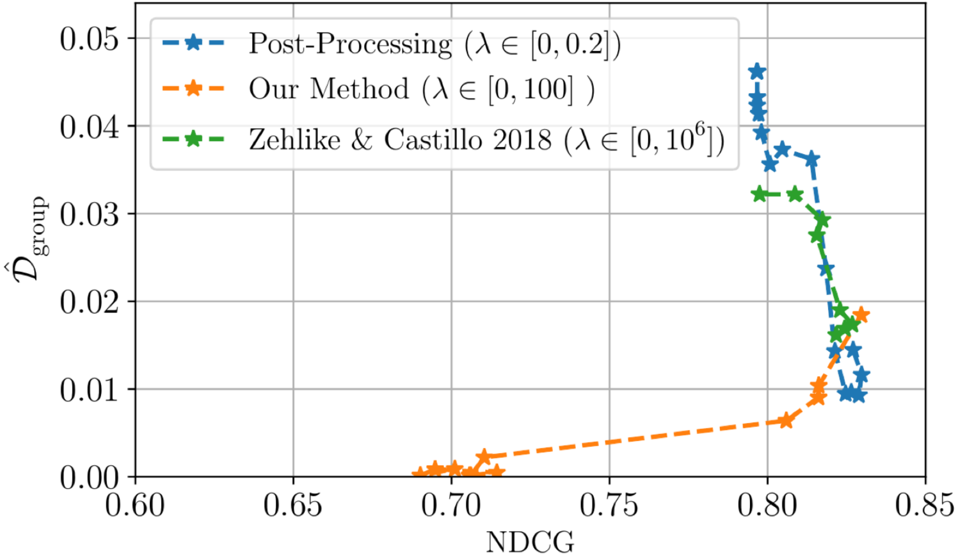
Learning to Rank



DELTR (large gamma)

Optimizing NDCG, ...

Singh and Joachims [NeurIPS 2019] present a more general framework that can optimize NDCG as well as individual and group fairness metrics



Singh, A. and Joachims, T. (2019). Policy learning for fairness in ranking. *Proc. NeurIPS.*

Learning from clicks

Clicks are biased towards top results, learning to rank needs to take this into account, e.g.:

Inverse propensity weighting
[Wang et al. SIGIR 2016, Joachims et al. WSDM 2017]

Learning propensity weights and unbiased ranker
[Ai et al. SIGIR 2018]

Learning from top-k observations [Oosterhuis 2020]

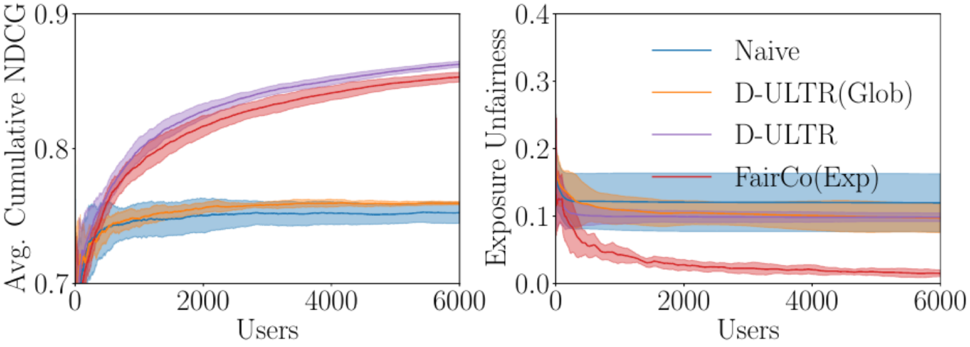
Controlling unfairness in LTR

FairCo adds a factor to correct unfairness to a LTR objective:

max(D(Gi, G)) .... where D(Gi, G) is either ...

Disparate exposure or

Disparate treatment (utility-normalized disparate exposure)

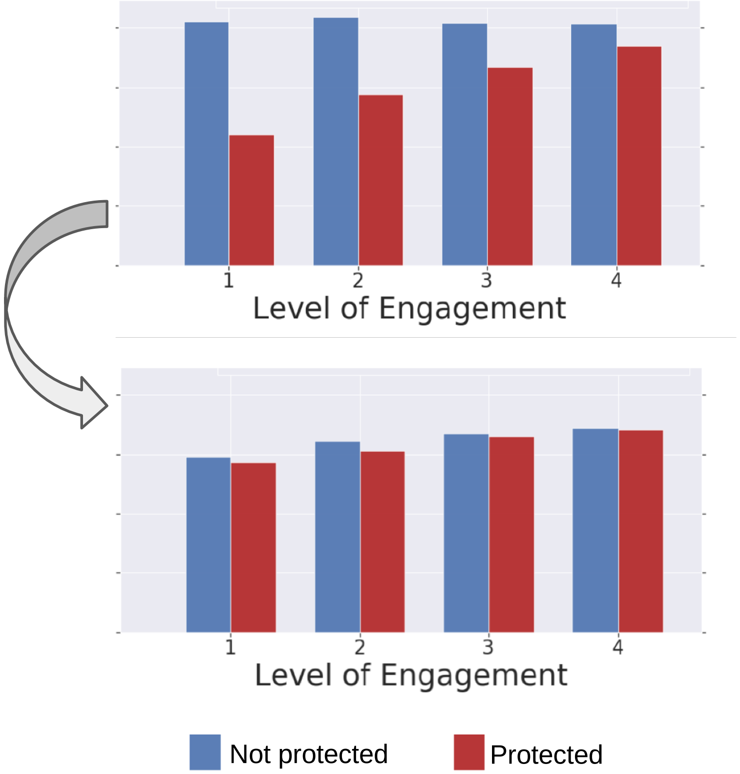


Morik, M., Singh, A., Hong, J., & Joachims, T. (2020). Controlling Fairness and Bias in Dynamic Learning-to-Rank. *Proc. SIGIR*

Other pairwise LTR methods

Inter-group pairwise fairness measures "success" rate:

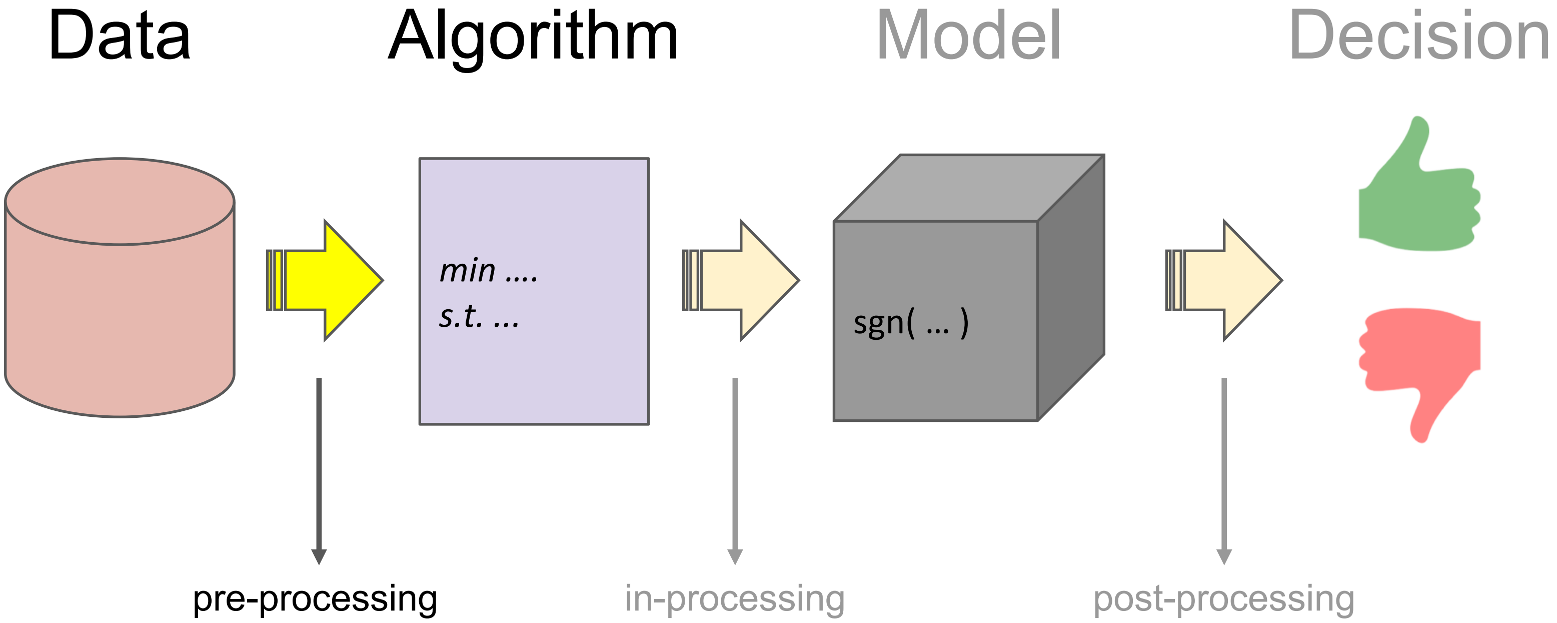
- u, v are **relevant**,
- u, v are **equally engaging**,
- u, v belong to **different** groups,
- u is ranked **above** v,
- u is clicked, v is not clicked



A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, C. Goodrow (2019). Fairness in Recommendation Ranking through Pairwise Comparisons. Proc. of KDD

** Protected is "sub-group" in the paper*

Pre-processing methods

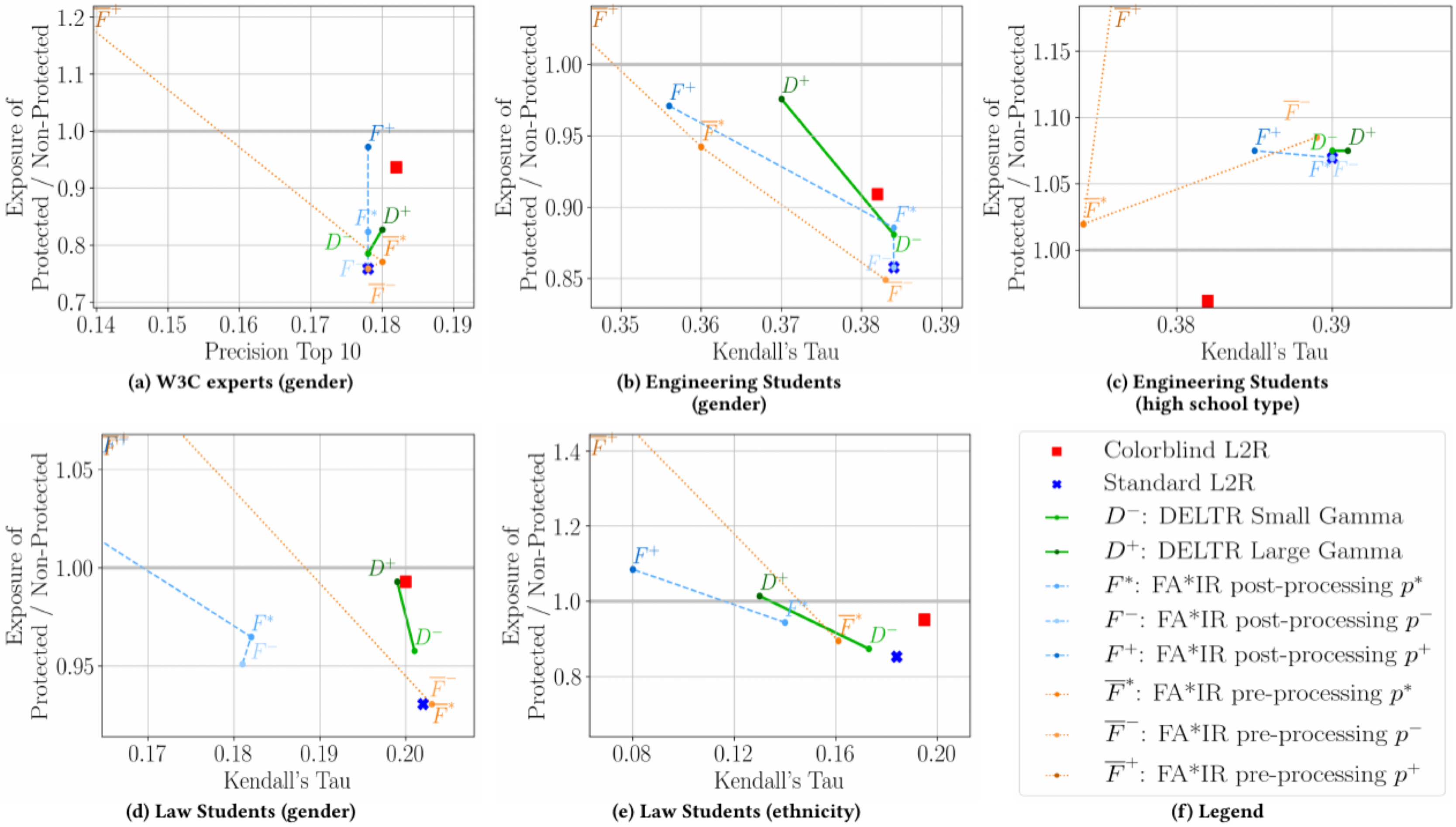


Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2125-2126). ACM.

Simple pre-processing of training data

1. Before training a LTR system
    - Ensure rankings given as input satisfy a fair ranking condition
2. Train the LTR as usual
3. Profit?

Zehlike, M., and Castillo, C. (2018). Reducing Disparate Exposure in Ranking: A Learning To Rank Approach. Preprint arXiv:1805.08716.



(a) W3C experts (gender)

(b) Engineering Students (gender)

(c) Engineering Students (high school type)

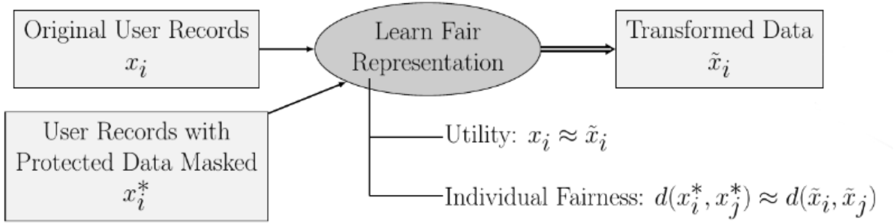
(d) Law Students (gender)

(e) Law Students (ethnicity)

(f) Legend

Zehlike, M., and Castillo, C. (2018). Reducing Disparate Exposure in Ranking: A Learning To Rank Approach. Preprint arXiv:1805.08716.

(Individually) fair representations



Original User Records x_i

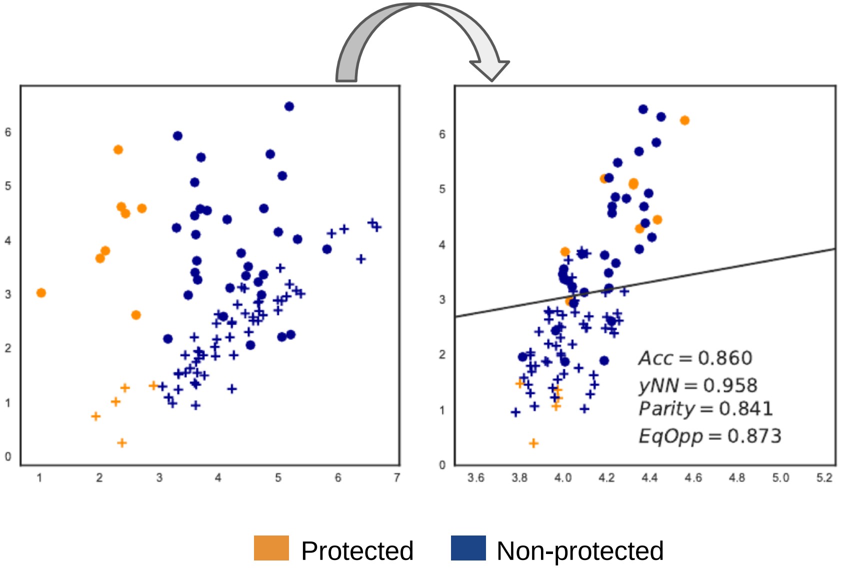
User Records with Protected Data Masked x_i^*

Learn Fair Representation

Transformed Data $\tilde{x}_i$

Utility: $x_i \approx \tilde{x}_i$

Individual Fairness: $d(x_i^*, x_j^*) \approx d(\tilde{x}_i, \tilde{x}_j)$



Input data is transformed to reduce the extent to which the distance between items is affected by protected attributes

Lahoti, P., Gummadi, K. P., & Weikum, G. (2019). iFair: Learning individually fair data representations for algorithmic decision making. In Proc. ICDE. IEEE.

Contents

1. Can algorithms discriminate?
2. Algorithmic fairness in IR
3. Measuring fairness in rankings
4. Creating fair rankings
5. Transparency in ranking

Transparency: why and how?

Why:

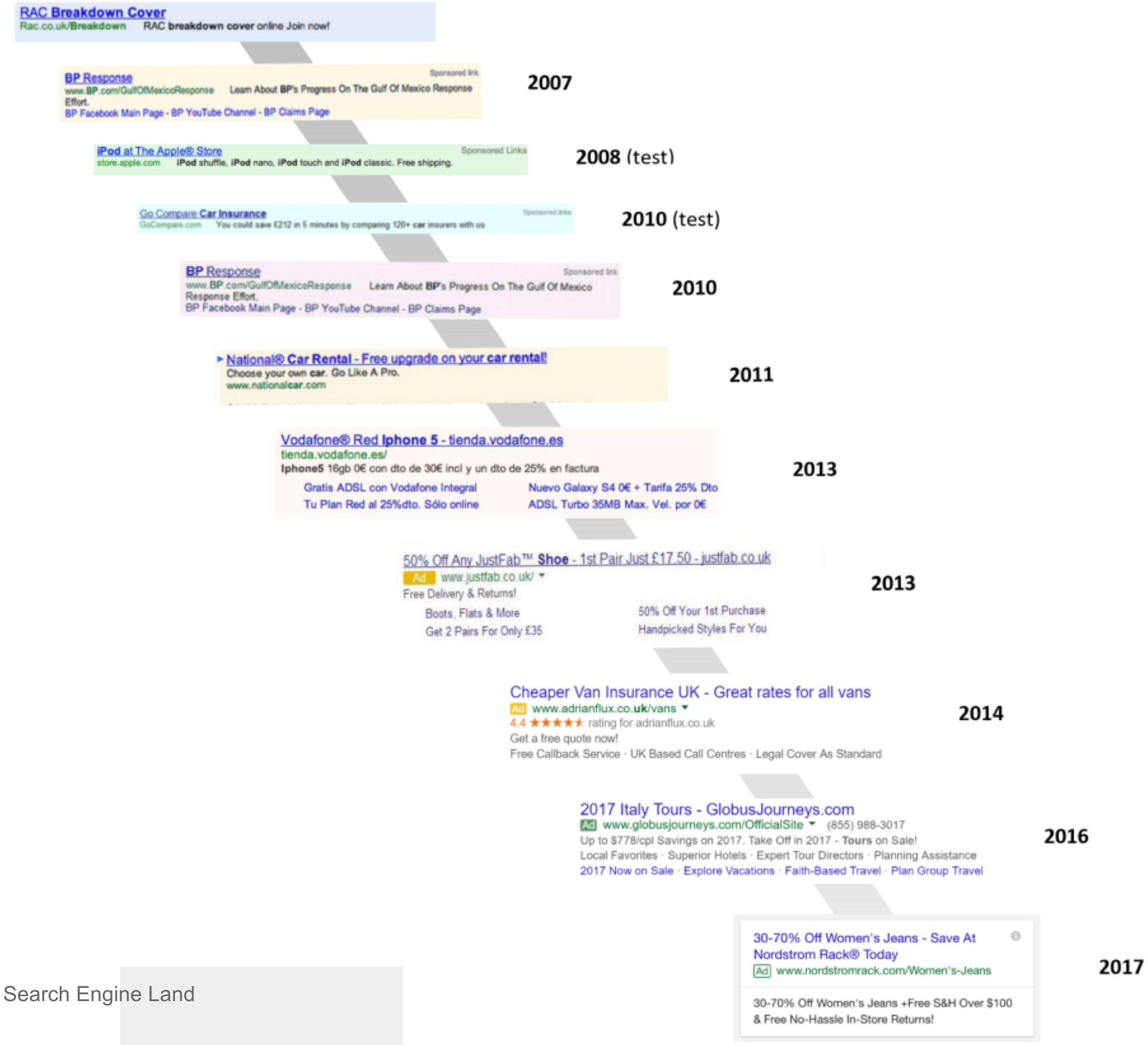
- Being able to **test** (remember we disregarded animosity)
- Supporting **ethics** compliance
- Ensuring implementation reflects **objectives**
- Making **trade-offs** visible

How:

- Explanations tend to be **contrastive**: why P and not Q?
- Explanations should empower users to **challenge** rankings

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. Preprint arXiv:1702.08608.
Miller, T., Howe, P., & Sonenberg, L. (2017). Explainable AI: Beware of inmates running the asylum. In IJCAI-17 Workshop on Explainable AI (XAI) (Vol. 36).

Advertising transparency is increasingly "transparent" (!)



RAC Breakdown Cover
Rac.co.uk/Breakdown RAC breakdown cover online Join now!

BP Response Sponsored link
www.BP.com/GulfOfMexicoResponse Learn About BP's Progress On The Gulf Of Mexico Response Effort.
BP Facebook Main Page - BP YouTube Channel - BP Claims Page

2007

iPod at The Apple® Store Sponsored Links
store.apple.com iPod shuffle, iPod nano, iPod touch and iPod classic. Free shipping.

2008 (test)

Go Compare Car Insurance Sponsored links
GoCompare.com You could save £212 in 5 minutes by comparing 120+ car insurers with us

2010 (test)

BP Response Sponsored link
www.BP.com/GulfOfMexicoResponse Learn About BP's Progress On The Gulf Of Mexico Response Effort.
BP Facebook Main Page - BP YouTube Channel - BP Claims Page

2010

National® Car Rental - Free upgrade on your car rental!
Choose your own car. Go Like A Pro.
www.nationalcar.com

2011

Vodafone® Red Iphone 5 - tienda.vodafone.es
tienda.vodafone.es/
Iphone5 16gb 0€ con dto de 30€ incl y un dto de 25% en factura
Gratis ADSL con Vodafone Integral Nuevo Galaxy S4 0€ + Tarifa 25% Dto
Tu Plan Red al 25%dto. Sólo online ADSL Turbo 35MB Max. Vel. por 0€

2013

50% Off Any JustFab™ Shoe - 1st Pair Just £17.50 - justfab.co.uk
Ad www.justfab.co.uk/
Free Delivery & Returns!
Boots, Flats & More 50% Off Your 1st Purchase
Get 2 Pairs For Only £35 Handpicked Styles For You

2013

Cheaper Van Insurance UK - Great rates for all vans
Ad www.adrianflux.co.uk/vans
4.4 ★★★★★ rating for adrianflux.co.uk
Get a free quote now!
Free Callback Service · UK Based Call Centres · Legal Cover As Standard

2014

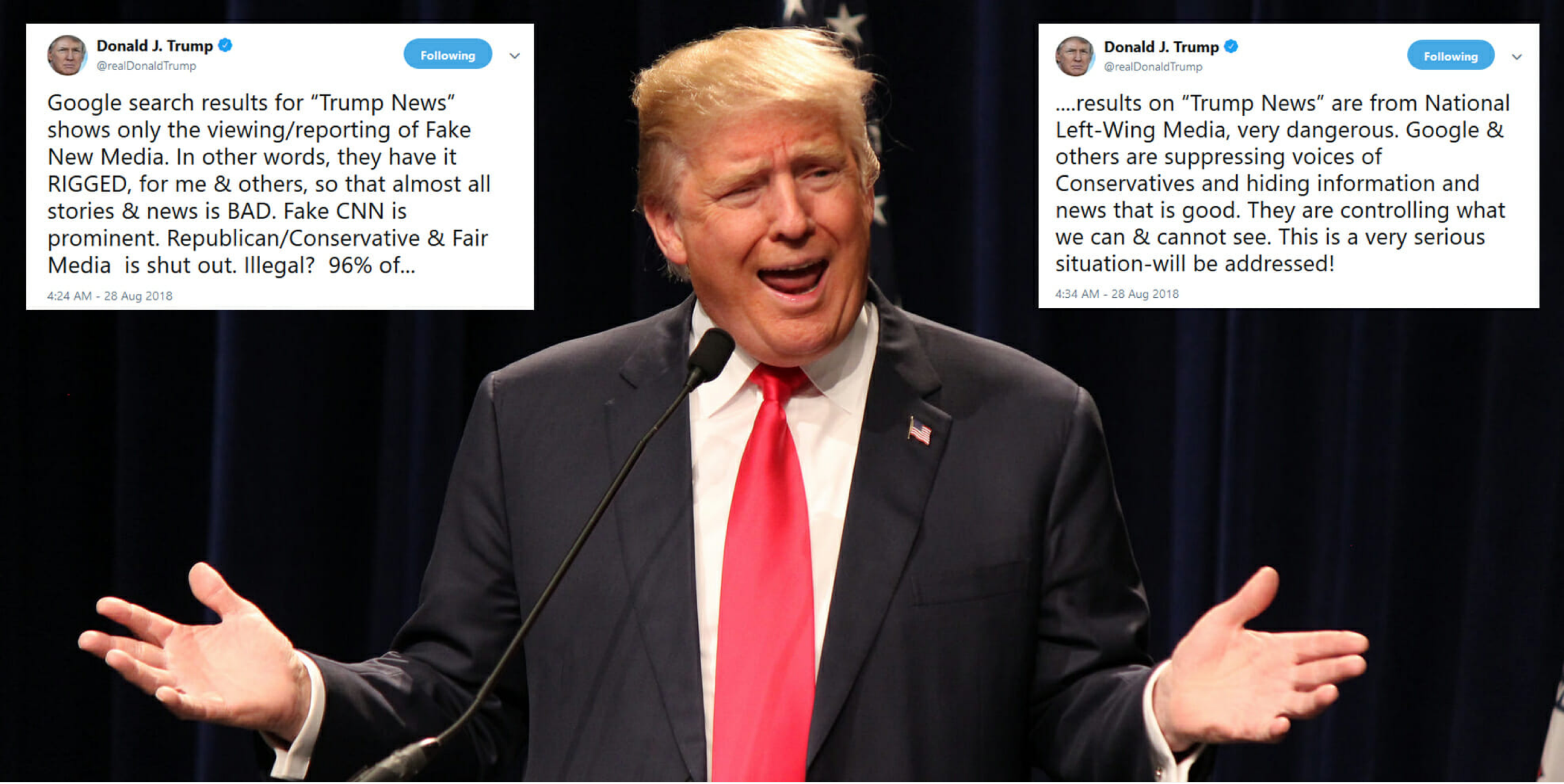
2017 Italy Tours - GlobusJourneys.com
Ad www.globusjourneys.com/OfficialSite (855) 988-3017
Up to $778/cpl Savings on 2017. Take Off in 2017 - Tours on Sale!
Local Favorites · Superior Hotels · Expert Tour Directors · Planning Assistance
2017 Now on Sale · Explore Vacations · Faith-Based Travel · Plan Group Travel

2016

30-70% Off Women's Jeans - Save At Nordstrom Rack® Today
Ad www.nordstromrack.com/Women's-Jeans
30-70% Off Women's Jeans +Free S&H Over $100 & Free No-Hassle In-Store Returns!

2017

Marvin, G. (2017): A visual history of Google ad labeling in search results. Search Engine Land



Donald J. Trump
@realDonaldTrump

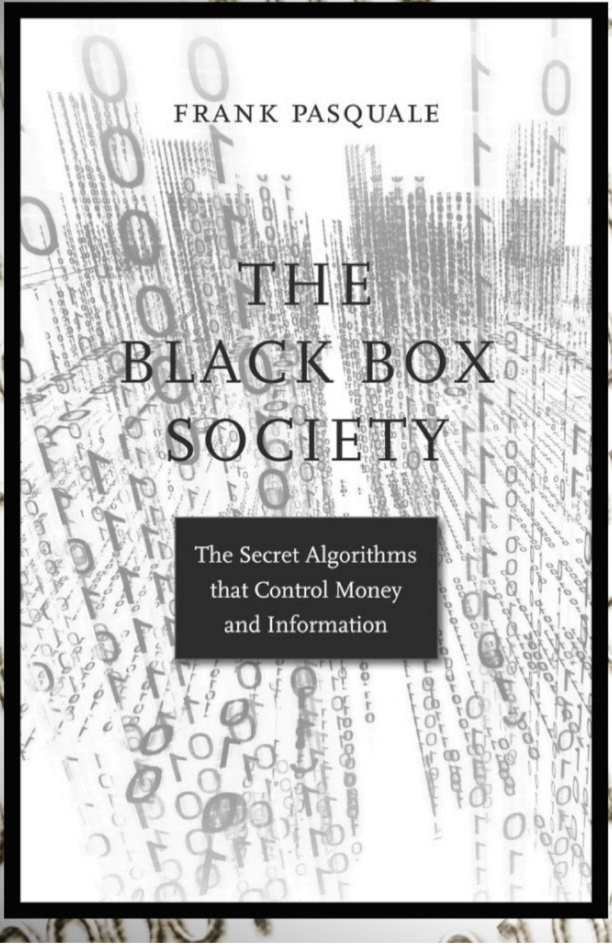
Google search results for "Trump News" shows only the viewing/reporting of Fake New Media. In other words, they have it RIGGED, for me & others, so that almost all stories & news is BAD. Fake CNN is prominent. Republican/Conservative & Fair Media is shut out. Illegal? 96% of...

4:24 AM - 28 Aug 2018

Donald J. Trump
@realDonaldTrump

....results on "Trump News" are from National Left-Wing Media, very dangerous. Google & others are suppressing voices of Conservatives and hiding information and news that is good. They are controlling what we can & cannot see. This is a very serious situation-will be addressed!

4:34 AM - 28 Aug 2018



FRANK PASQUALE

THE BLACK BOX SOCIETY

The Secret Algorithms that Control Money and Information

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Transparency in algorithmic rankings

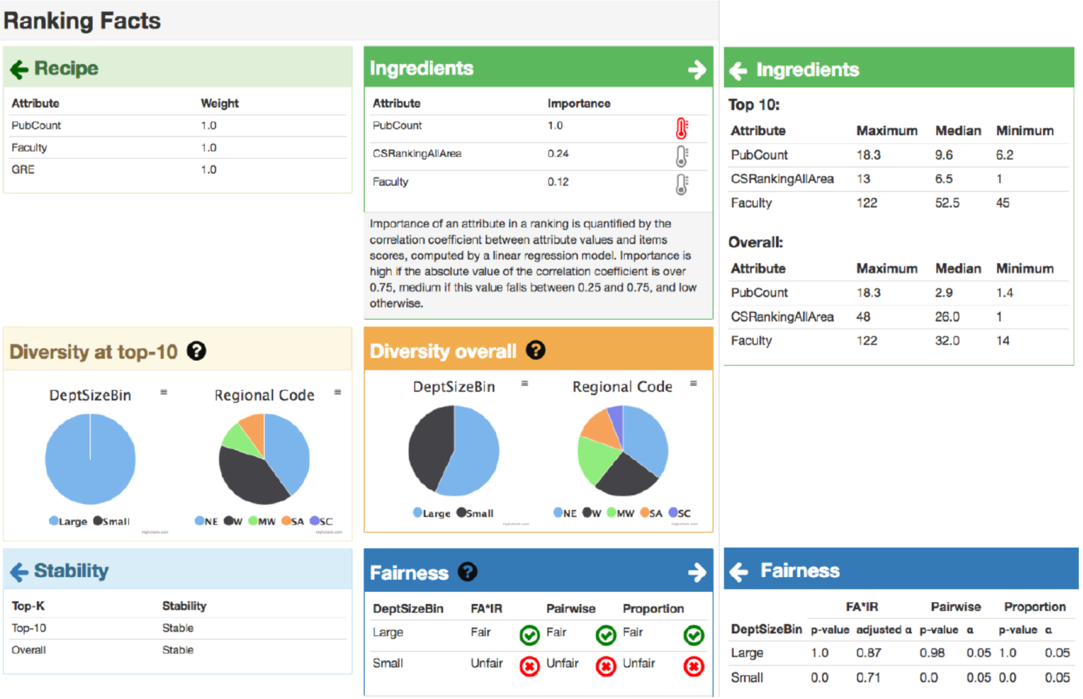
"Broadcast television can be monitored by anyone … **If the nightly television news does not cover a protest, the lack of coverage is evident** … However, **there is no transparency in algorithmic filtering**: how is one to know whether Facebook is showing [news about a protest] to everyone else but him or her, whether there is just no interest in the topic, or whether it is the algorithmic feedback cycle that is depressing the updates in favor of a more algorithm-friendly topic …?"

That Matter
ZEYNEP TUFEKCI
Author, "Twitter and Tear Gas: The Power and Fragility of Networked Protest"

Tufekci, Z. (2017). Twitter and tear gas: The power and fragility of networked protest. Yale University Press.

Nutritional labels for rankings

Provide transparency about ranking factors, composition of the list, and fairness tests



Ranking Facts

Recipe

Attribute	Weight
PubCount	1.0
Faculty	1.0
GRE	1.0

Ingredients

Attribute	Importance
PubCount	1.0
CSRankingAllArea	0.24
Faculty	0.12

Importance of an attribute in a ranking is quantified by the correlation coefficient between attribute values and items scores, computed by a linear regression model. Importance is high if the absolute value of the correlation coefficient is over 0.75, medium if this value falls between 0.25 and 0.75, and low otherwise.

Ingredients

Top 10:

Attribute	Maximum	Median	Minimum
PubCount	18.3	9.6	6.2
CSRankingAllArea	13	6.5	1
Faculty	122	52.5	45

Overall:

Attribute	Maximum	Median	Minimum
PubCount	18.3	2.9	1.4
CSRankingAllArea	48	26.0	1
Faculty	122	32.0	14

Diversity at top-10 ?

Diversity overall ?

Stability

Top-K	Stability
Top-10	Stable
Overall	Stable

Fairness ?

DeptSizeBin	FA*IR		Pairwise		Proportion	
Large	Fair	✓	Fair	✓	Fair	✓
Small	Unfair	✗	Unfair	✗	Unfair	✗

Fairness

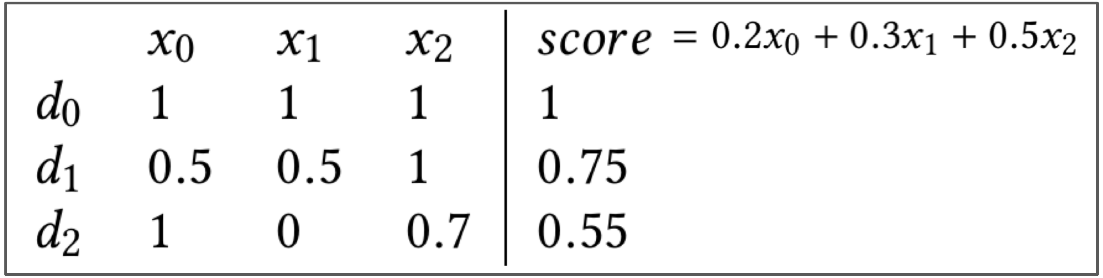
	FA*IR		Pairwise		Proportion	
DeptSizeBin	p-value	adjusted α	p-value	α	p-value	α
Large	1.0	0.87	0.98	0.05	1.0	0.05
Small	0.0	0.71	0.0	0.05	0.0	0.05

Example ranking labels for a ranking of computer science departments ▶

Yang, K., Stoyanovich, J., Asudeh, A., Howe, B., Jagadish, H. V., & Miklau, G. (2018). A Nutritional Label for Rankings. In Proc. SIGMOD (pp. 1773-1776). ACM.

Perturbation-based method

Suppose the score is a linear function of features, and documents are ranked by decreasing score ▶



	x_0	x_1	x_2	$score = 0.2x_0 + 0.3x_1 + 0.5x_2$
d_0	1	1	1	1
d_1	0.5	0.5	1	0.75
d_2	1	0	0.7	0.55

Feature x_2 has the highest weight but even if it were 0.6 for d_0 (lower than any other), document d_0 still would be at the top

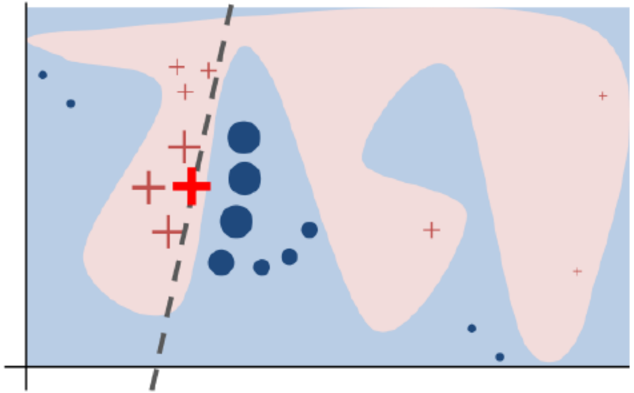
In contrast, changing feature x_1 to 0 would change the ranking, hence x_1 is a better explanation

ter Hoeve, M., Schuth, A., Odijk, D., & de Rijke, M. (2018). Faithfully Explaining Rankings in a News Recommender System. arXiv preprint arXiv:1805.05447.

Replace with explainable model

Model introspection approaches explain what a particular model is doing, *model agnostic* approaches do not

A classical idea in model interpretability is to mimic a black-box model with a different model that uses a simpler logic but generates a similar output [LIME Ribeiro et al. KDD 2016]



Singh, J., & Anand, A. (2020). Model agnostic interpretability of rankers via intent modelling. *Proc. FAT**

Transparency can help us researchers

Transparency helps us avoid (at least) two pitfalls:

- **Sneaking positive/affirmative action** without a consensus or where it is not welcome
- **Certifying an algorithm that is part of an unfair system** or is used in conditions of unfairness

Barocas, S. (2017). What is the problem to which fair machine learning is the solution? AI Now Experts Workshop on Bias and Inclusion
Keyes, O., Hutson, J., & Durbin, M. (2019). A Mulching Proposal. arXiv preprint arXiv:1908.06166.

Conclusions

Take-home messages

Fairness in IR/RecSys is less studied than in ML/DM

Sometimes it requires solving an exciting algorithmic puzzle,
but often it does not

Paraphrasing Solon Barocas:
«What is the problem to which fair ranking is the solution?»

Different solutions address different problems
(remove discrimination ≠ provide equal opportunity)

See also

Fairness in Ranking: A Survey NEW (March 2021)

by M. Zehlike, K. Yang, J. Stoyanovich

Fair Information Access tutorial at SIGIR/RecSys/...

by M. Ekstrand, F. Diaz, and R. Burke

FAccT Conference

Happened March 3rd-10th, 2021

FAccT 2021
Toronto, Canada