# FairER: Entity Resolution with Fairness Constraints

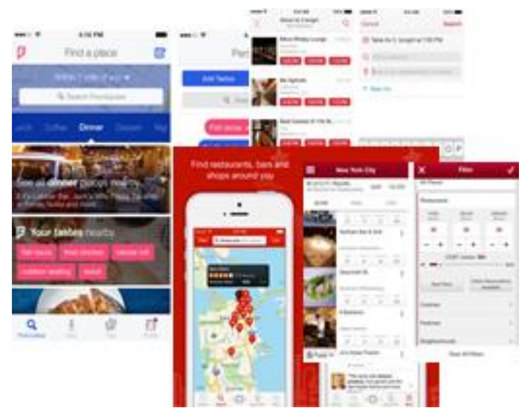

**Vasilis Efthymiou[1], Kostas Stefanidis[2], Evaggelia Pitoura[3], Vassilis Christophides[4]**

[1]FORTH-ICS, Greece [2]Tampere University, Finland [3]University of Ioannina, Greece [4]ENSEA, ETIS, Cergy, France


## Motivation

Data-driven decision-making AI systems…

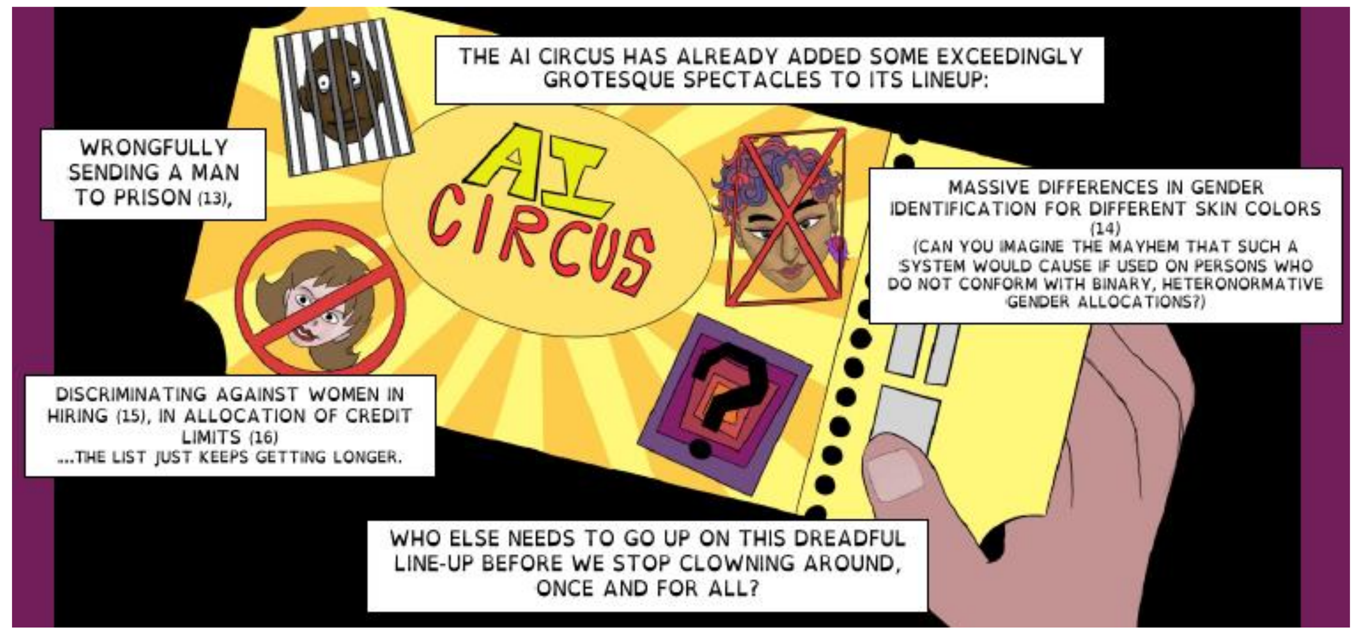
*Where to eat?*



*Who to date?*

*What to read, watch, buy..?*

…have some serious fairness issues




Falaah Arif Khan and Julia Stoyanovich. "Mirror, Mirror". Data, Responsibly Comics, Volume 1 (2020) https://dataresponsibly.github.io/comics/vol1/mirror_en.pdf

## Entity Resolution

Identify data records from different sources that refer to the same real-world entities (matches)

**E**

| id | Name | Location | Employer | Rep | Sex |
|---|---|---|---|---|---|
| $e_1$ | Danny Barber | LA | UCLA | 600 | M |
| $e_2$ | Susan Doe | Texas | UT Austin | 7,000 | F |
| $e_3$ | Peter Simons | NY | NYU | 4 | M |
| $e_4$ | M. Anderson | Denmark | Aarhus Univ. | 8 | M |
| $e_5$ | Julia Rondo | France | CNRS, Paris | 460 | F |
| $e_6$ | J. Parker | California | UC Berkeley | 381 | M |

**E'**

| id | Full-name | Affiliation | h-index | Sex |
|---|---|---|---|---|
| $e'_1$ | Doe, S. | UT | 14 | F |
| $e'_2$ | J. Parker | UCSC | 5 | M |
| $e'_3$ | Simons, Pete | NYU | 11 | M |
| $e'_4$ | M. Anderson | Aarhus | 15652 | M |
| $e'_5$ | J. Rondo | CNRS | 4653 | F |
| $e'_6$ | Juliana Rondo | CNRS | 25 | F |

## Traditional (fairness-agnostic) ER

- Goal: Discover as many *matches* as possible
- Matching decision process for $(e,e')$: Relies on a scoring function $s$ assessing the similarity of the attribute values and names used to describe $e$ and $e'$
  - scoring function types: heuristics, aggl. clustering, supervised

*Definition 2.1 (Fairness-agnostic ER).* Given a set of candidate matches $C \subseteq E \times E'$ and a scoring function $s : E \times E' \to \mathbb{R}$, produce a subset $R \subseteq C$ of matches that maximizes the cumulative scores:

$$R = \underset{R^* \subseteq C}{\operatorname{argmax}} \sum_{(e_i, e'_j) \in R^*} s(e_i, e'_j).$$

- Matching pairs ranked according to their scores
- Additional constraints may be imposed to $R$
  - e.g., 1-to-1 constraint (clean-clean ER)
- Issue: ignores the qualitative features of the results

### Fairness-aware ER

- Retrieved results: not only the most likely matches, but also satisfy some *fairness constraint*
- *Fairness in ER decisions*: equal decision measures that allow to examine the allocation of benefits and harms across groups by looking at the decision alone
  - group-based fairness: protected vs non-protected
    - all groups should receive similar treatment, i.e., have similar chances to be resolved
- Ranked group fairness: a fairness constraint should be satisfied when considering the results within a given rank

*Definition 2.2 (Fairness-aware ER).* Given a set of candidate matches $C \subseteq E \times E'$, a scoring function $s : E \times E' \to \mathbb{R}$, and a fairness criterion $F$, produce a ranking of matches $R \subseteq C$ that for any given rank position $k$, maximizes the cumulative scores:

$$R = \underset{R^* \subseteq C}{\operatorname{argmax}} \sum_{(e_i, e'_j) \in R^*} s(e_i, e'_j)$$

$$\text{s.t. } R[k] \text{ satisfies } F,$$

where $R[k]$ are the $k$ first results of $R$.

## FairER

Fairness criterion F: $\left| \frac{|R_p|}{k} - \frac{|R_n|}{k} \right| = \varepsilon^*,$

where $\frac{|R_p|}{k}$ (resp. $\frac{|R_p|}{k}$) is the ratio of protected (resp. non-protected) group members in the first $k$ results and $\varepsilon^*$is the smallest possible ratio for a given $k$

### Example

desc. score

| Traditional ER | R[4] | |
|---|---|---|
| $e_3$-$e'_3$ | $e_3$-$e'_3$ | 3 male |
| $e_4$-$e'_4$ | $e_4$-$e'_4$ | |
| $e_6$-$e'_2$ | $e_6$-$e'_2$ | |
| $e_5$-$e'_6$ | $e_5$-$e'_6$ | 1 female |
| ~~$e_5$-$e'_5$~~ | | |
| ~~$e_5$-$e'_1$~~ | | |
| $e_2$-$e'_1$ | | |

desc. score

| FairER $Q_p$ | FairER $Q_n$ | R[4] | |
|---|---|---|---|
| $e_5$-$e'_6$ | $e_3$-$e'_3$ | $e_5$-$e'_6$ | female |
| ~~$e_5$-$e'_5$~~ | $e_4$-$e'_4$ | $e_3$-$e'_3$ | male |
| ~~$e_5$-$e'_1$~~ | $e_6$-$e'_2$ | $e_2$-$e'_1$ | female |
| $e_2$-$e'_1$ | $e_1$-$e'_2$ | $e_4$-$e'_4$ | male |
| $e_2$-$e'_5$ | ~~$e_3$-$e'_2$~~ | | |

**Proposition**: FairER is a 1-1/e approximation to the problem of fairness-aware ER, for F defined as $||R\_p|/k-|R\_n|/k|=\varepsilon^*$

## Experimental Evaluation

**Baselines**: Color-blind (DeepMatcher, 1-to-1 constraint), FA*IR (fair ranking algorithm)

**Datasets**: All 7 datasets available from DeepMatcher

Accuracy@k

| Method | D1 | D2 | D3 | D4 | D5 | D6 | D7 |
|---|---|---|---|---|---|---|---|
| Accuracy@5 | | | | | | | |
| Color-blind | **1** | **1** | **1** | **1** | **1** | 0.8 | 0 |
| Fa*ir | **1** | **1** | **1** | **1** | **1** | **1** | 0 |
| FairER | **1** | **1** | **1** | **1** | **1** | 0.8 | **0.4** |
| Accuracy@10 | | | | | | | |
| Color-blind | **0.9** | **1** | **1** | **1** | **1** | 0.7 | **0.4** |
| Fa*ir | **0.9** | **1** | **1** | **1** | **1** | 0.6 | **0.4** |
| FairER | **0.9** | **1** | 0.8 | **1** | **1** | **0.8** | **0.4** |
| Accuracy@15 | | | | | | | |
| Color-blind | **0.73** | **1** | **1** | **1** | 0.93 | **0.73** | **0.6** |
| Fa*ir | **0.73** | **1** | 0.93 | **1** | 0.93 | 0.71 | **0.6** |
| FairER | 0.66 | **1** | 0.66 | **1** | **1** | **0.73** | 0.33 |

Bias@k = $\frac{|R_p|}{k} - \frac{|R_n|}{k}$

fav. non-pr. no bias fav. prot.

0

| Method | D1 | D2 | D3 | D4 | D5 | D6 | D7 |
|---|---|---|---|---|---|---|---|
| $k = 5$ ($\epsilon* = 0.2$) | | | | | | | |
| Color-blind | -0.6 | -0.6 | -1 | -1 | -1 | -1 | -0.6 |
| Fa*ir | -0.6 | -0.6 | -0.6 | -0.6 | -0.6 | -0.6 | -0.6 |
| FairER | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** |
| $k = 10$ ($\epsilon* = 0$) | | | | | | | |
| Color-blind | -0.6 | -0.4 | -1 | -0.6 | -0.8 | -0.8 | -0.8 |
| Fa*ir | -0.6 | -0.4 | -0.6 | -0.6 | -0.6 | -0.6 | -0.6 |
| FairER | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| $k = 15$ ($\epsilon* = 0.07$) | | | | | | | |
| Color-blind | -0.6 | -0.33 | -0.73 | -0.6 | -0.87 | -0.87 | -0.87 |
| Fa*ir | -0.47 | -0.33 | -0.47 | -0.47 | -0.47 | -0.57 | -0.47 |
| FairER | **0.07** | **0.07** | **0.07** | **0.07** | **0.07** | **0.07** | **0.07** |

https://github.com/vefthym/fairER