What is fairness?

fairness Add to list Share

/ˈfɛrnɪs/ /ˈfɛnɪs/ IPA guide

Fairness is the quality of making judgments that are free from discrimination. Judges, umpires, and teachers should all strive to practice *fairness*.

CRITEO

What is fairness?

fairnes

/ˈfɛrnɪs/

Fairness is the
from discrimin
strive to practi

discrimination Add to list Share

/dəˈskrɪməˌneɪʃən/ /dɪskrɪmɪˈneɪʃən/ IPA guide

Other forms: **discriminations**

Discrimination means distinguishing between two or more people or things. Usually, *discrimination* means an unfair system that treats one group of people worse than another.

Discrimination isn't always a bad thing. You have to use discrimination to tell a good restaurant from a bad one, or a good idea from a bad one. However, this word is very tarnished by its main use these days: as a word for unfair treatment of people, based on racism, sexism, homophobia, or another prejudice. There is a huge history of discrimination in the United States.

CRITEO

Translations of Fairness

The following table provides the translation of fairness, as stated in Article 5, in each of the 23 languages with an official translation of the GDPR.

Centre for Information Policy Leadership
HUNTON ANDREWS KURTH

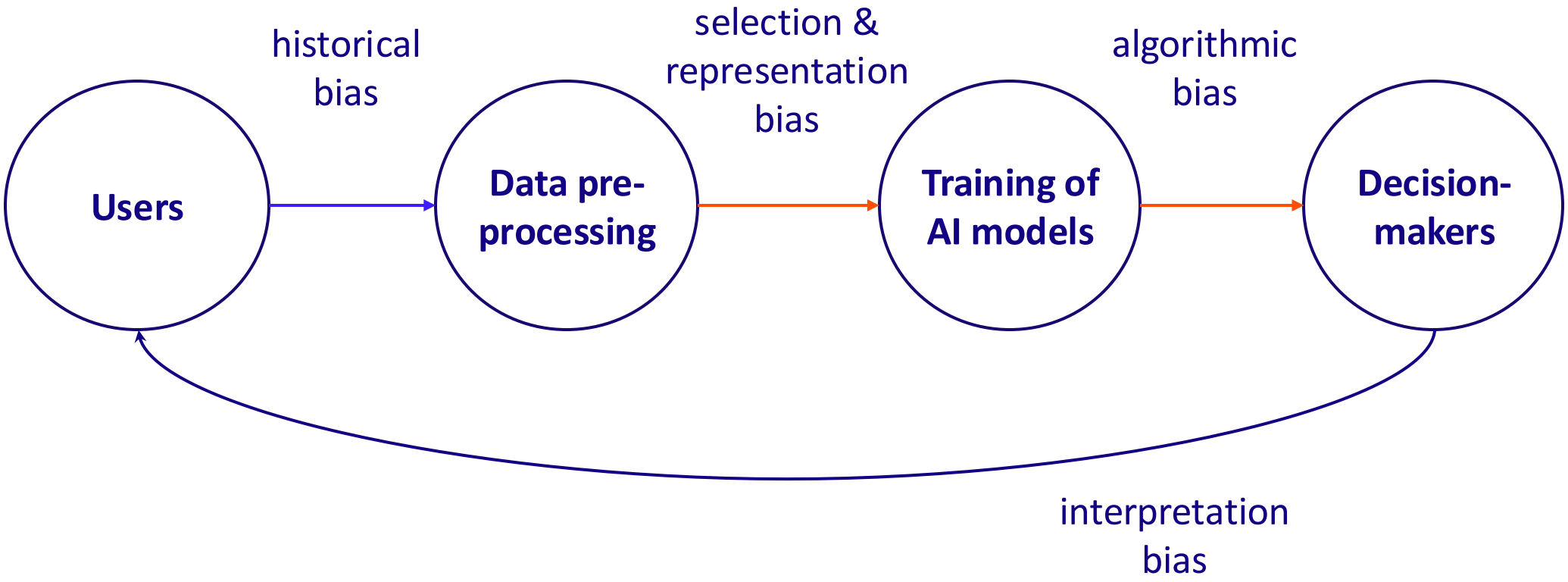
Language	Article 5 wording	Google Translation
Bulgarian	добросъвестност	Good faith
Croatian	poštenosti	Honesty
Czech	korektnost	Correctness, propriety
Danish	rimelighed	Reasonably
Dutch	behoorlijkheid	Goodness
English	fairness	Fairness
Estonian	õiglus	Justice, justness, fairness, equity, righteousness
Finnish	kohtuullisuus	Equity, fairness
French	loyauté	Loyalty, trustworthiness, fidelity
Gaelic	cothroime	Fairness
German	Verarbeitung nach Treu und Glauben	Good faith processing
Greek	αντικειμενικότητα	Objectivity

Language	Article 5 wording	Google Translation
Hungarian	tisztességes eljárás	Due process, fair play
Italian	correttezza	Correctness, fairness, propriety, honesty
Latvian	godprātība	Integrity, honesty, good faith
Lithuanian	sąžiningumo	Fairness, honesty, integrity, good faith
Maltese	ġustizzja	Justice, fairness
Polish	rzetelność	Reliability, dependability, honesty, rectitude, squareness
Romanian	echitate	Fairness, equity, justice, uprightness
Slovak	spravodlivosť	Justice, justness, equity, rectitude, uprightness, narrow way, virtuousness
Slovenian	pravičnost	Justice
Spanish	lealtad	Loyalty, allegiance, devotion
Swedish	korrekthet	Correctness, propriety

Source: CIPL 2020. *Artificial Intelligence and Data Protection: Delivering Sustainable AI Accountability* in Practice

CRITEO

Unfairness in AI models



CRITEO

We need

Accountability

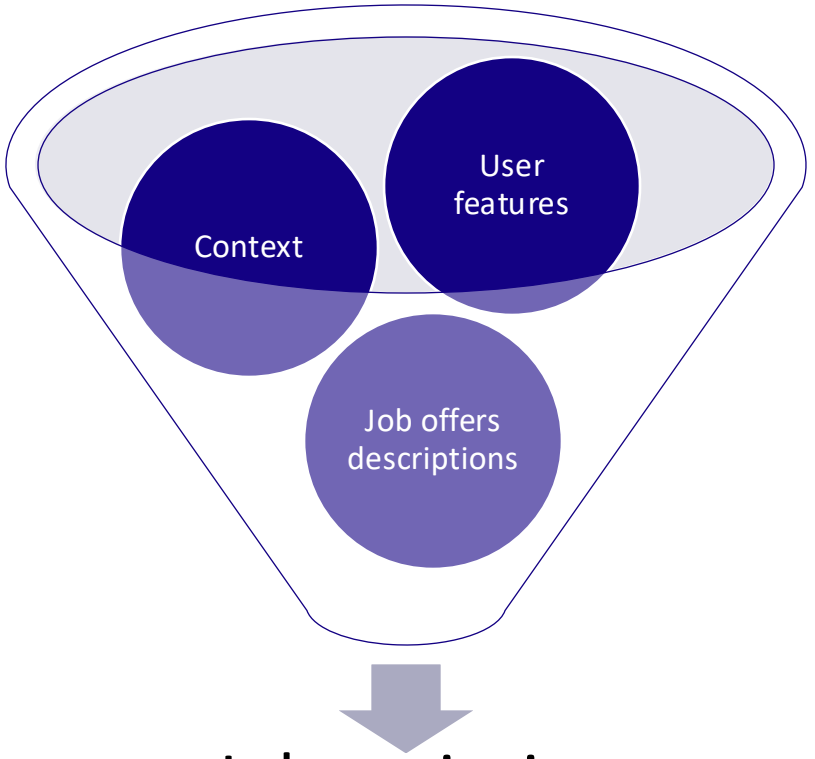
Transparency

Research

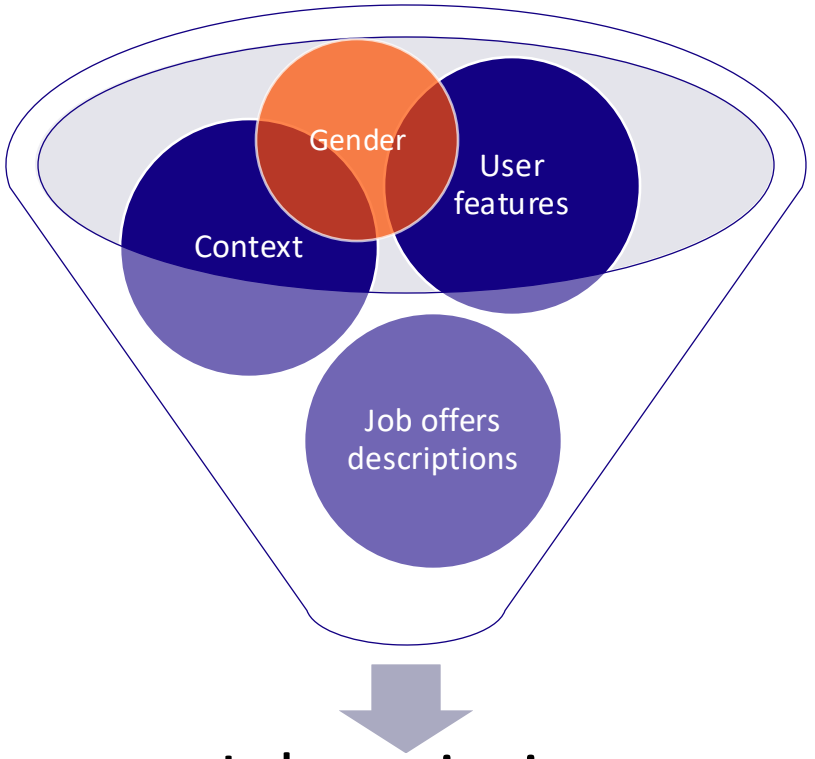
CRITEO

Once upon a time you browse on the Internet… And you see a **job advertising**. Which one will it be?

CRITEO

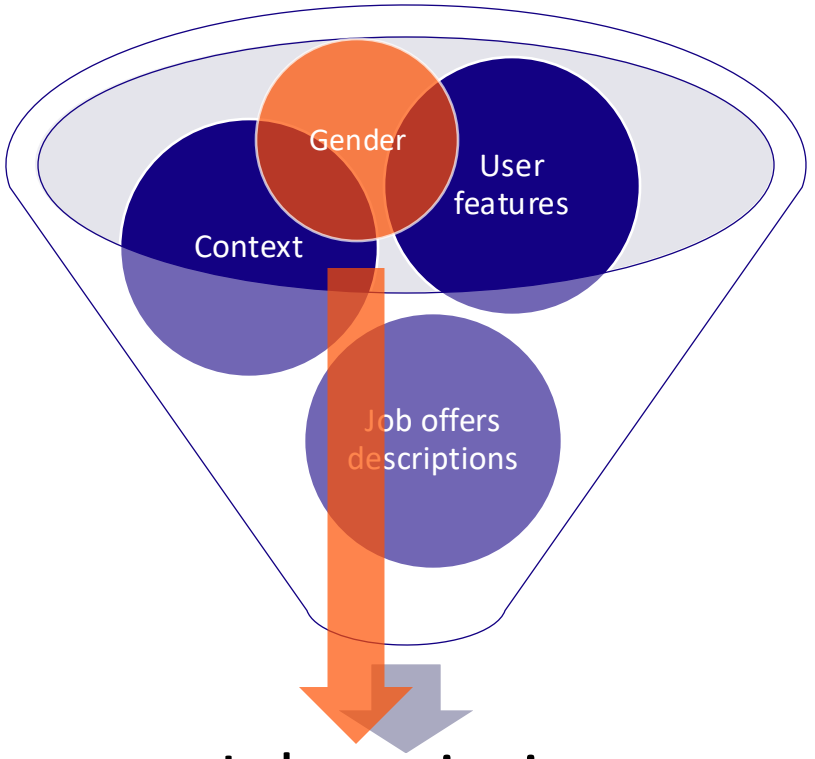


Job seniority
(Team lead / Office assistant)



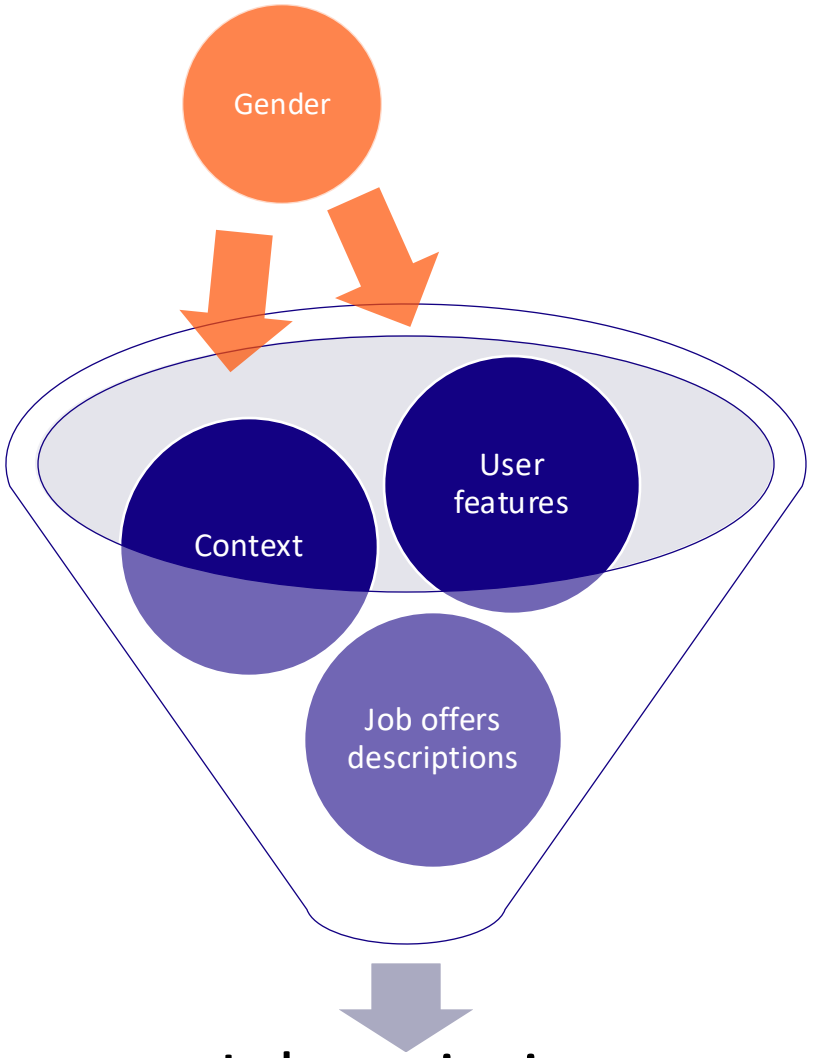
Job seniority
(Team lead / Office assistant)

CRITEO



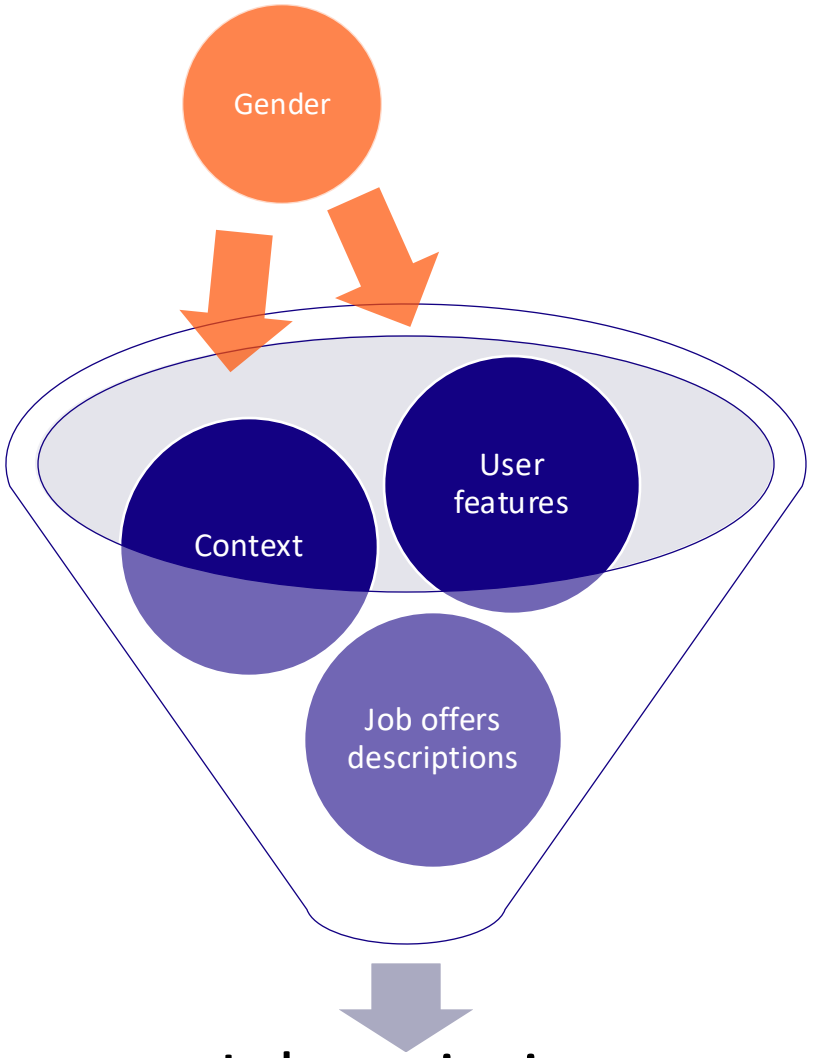
Job seniority
(Team lead / Office assistant)

CRITEO



Job seniority
(Team lead / Office assistant)

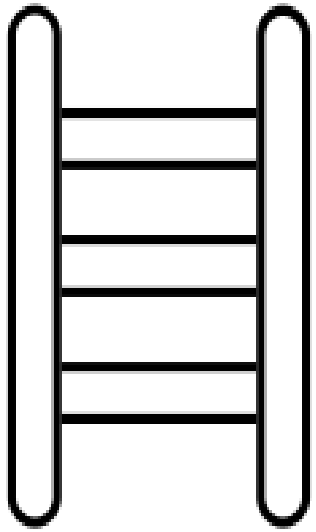
CRITEO



Job seniority
(Team lead / Office assistant)

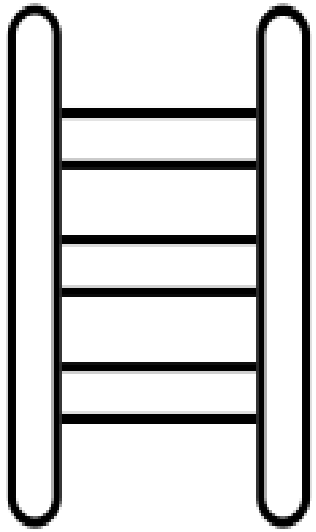
CRITEO

Fairness ladder



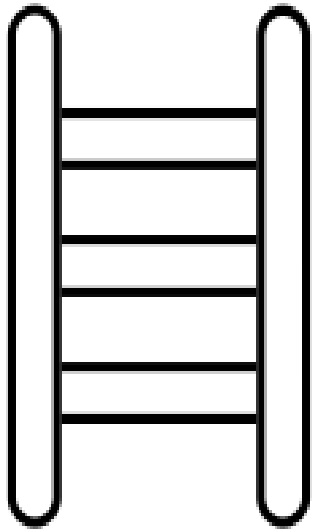
CRITEO

Fairness ladder



Level 1 = fairness through unawareness
(e.g. do not directly use gender)

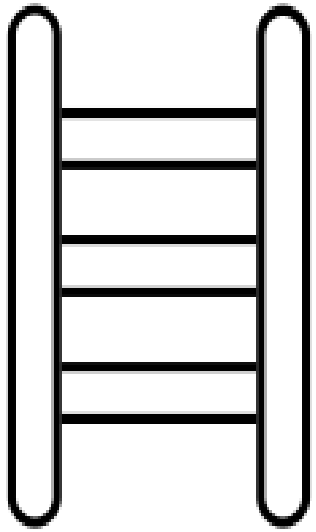
Fairness ladder



Level 1 = fairness through unawareness
(e.g. do not directly use gender)

Level 2 = fairness in prediction
(e.g. bias reduction methods)

Fairness ladder



Level 1 = fairness through unawareness
(e.g. do not directly use gender)

Level 2 = fairness in prediction
(e.g. bias reduction methods)

Level 3 = correct external bias
(e.g. positive discrimination)

CRITEO

EU AI Act

Annex III

High-Risk AI Systems Referred (6)3

In each of the areas listed under points 1-8, the AI systems specifically mentioned under each letter are considered to be high-risk AI systems pursuant to Article 6(3):

1. Biometrics:

   a. Remote biometric identification systems.

2. Critical infrastructure:

...

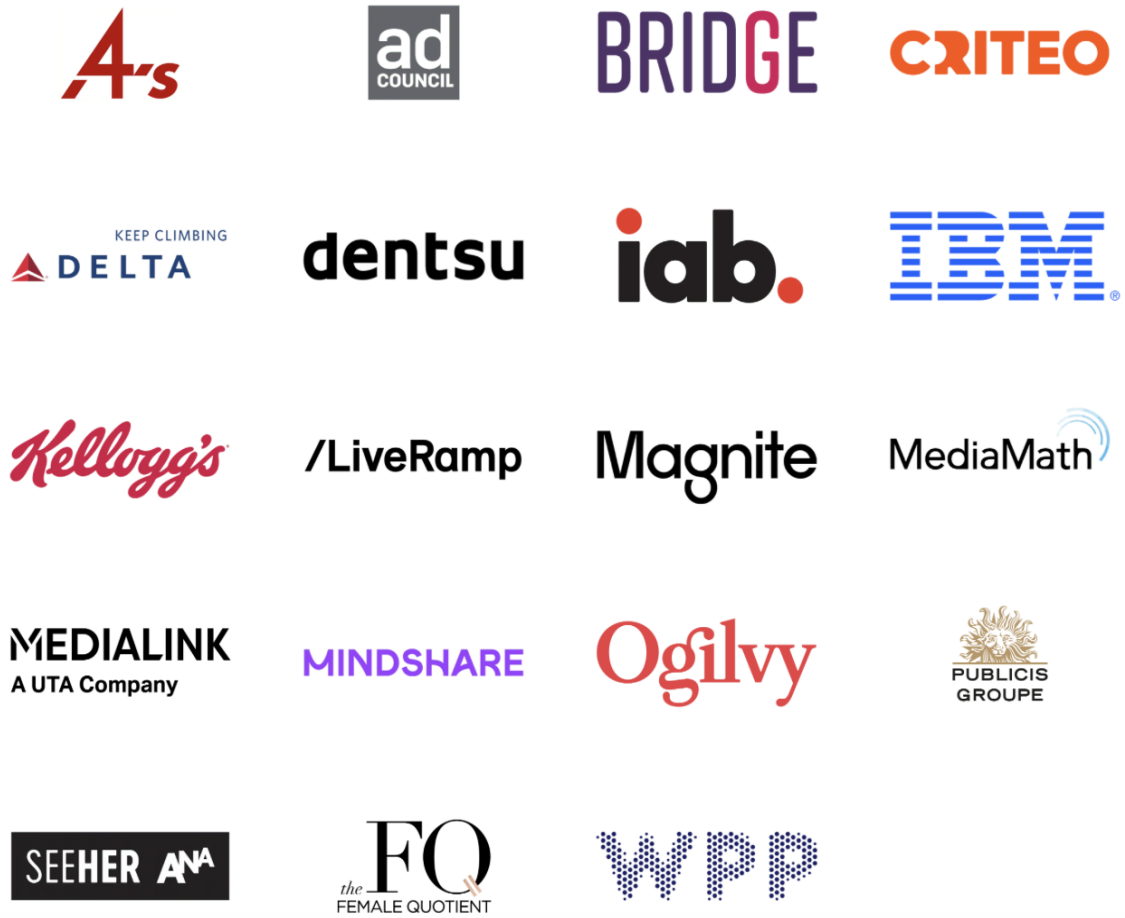
4. Employment, workers management and access to self-employment:

   a. AI systems intended to be used for recruitment or selection of natural persons, notably to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates;

   b. AI intended to be used to make decisions on promotion and termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics and to monitor and evaluate performance and behaviour of persons in such relationships.

https://www.euaiact.com/annex/3

Organizations that have committed to the Pledge



https://www.ibm.com/watson-advertising/news/advertising-fairness

Motivation

- Ethics: existing historical **bias** might be **amplified**
- Pragmatics: possible **legal charges** for AI models creators and users
- Economics: service **quality decreases**
- We want to **investigate** and **anticipate**

How to investigate possible bias?

CRITEO

Choose

the protected attribute
(e.g. gender)

Study

where the bias appear

Test

bias mitigation techniques with proper metrics

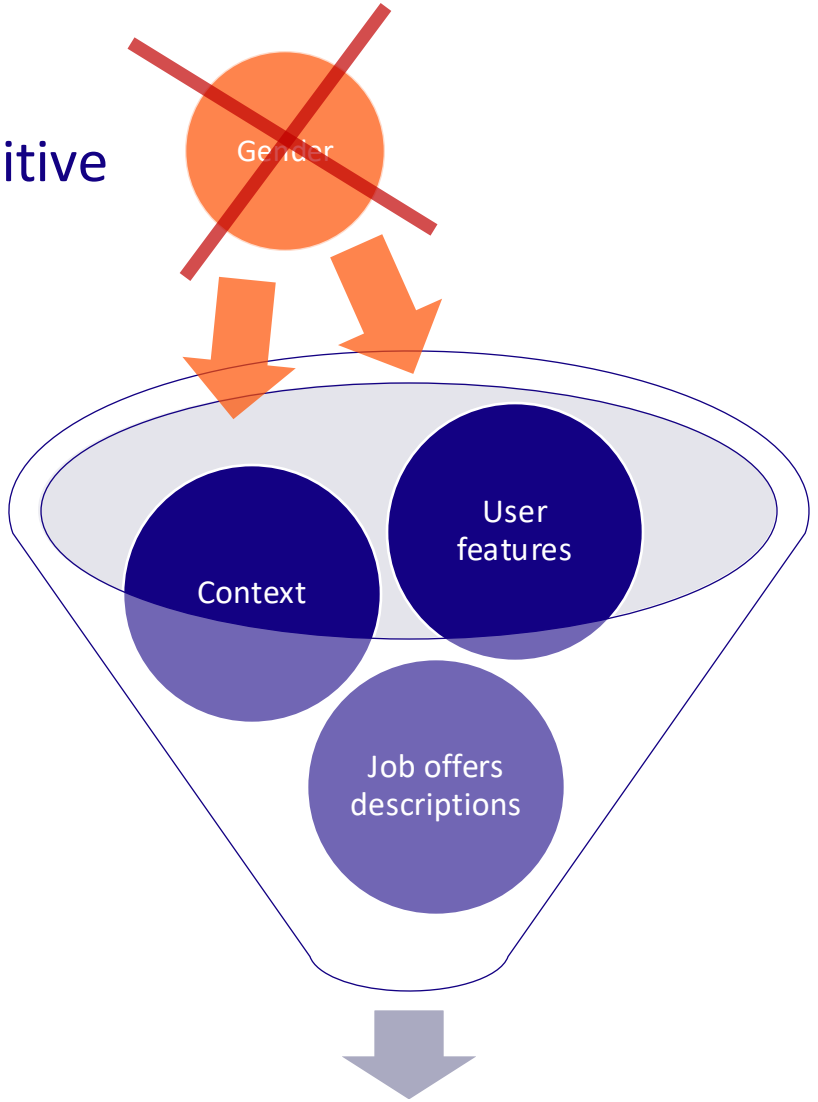
Challenges for Criteo

Complex system

Bias-utility trade-off

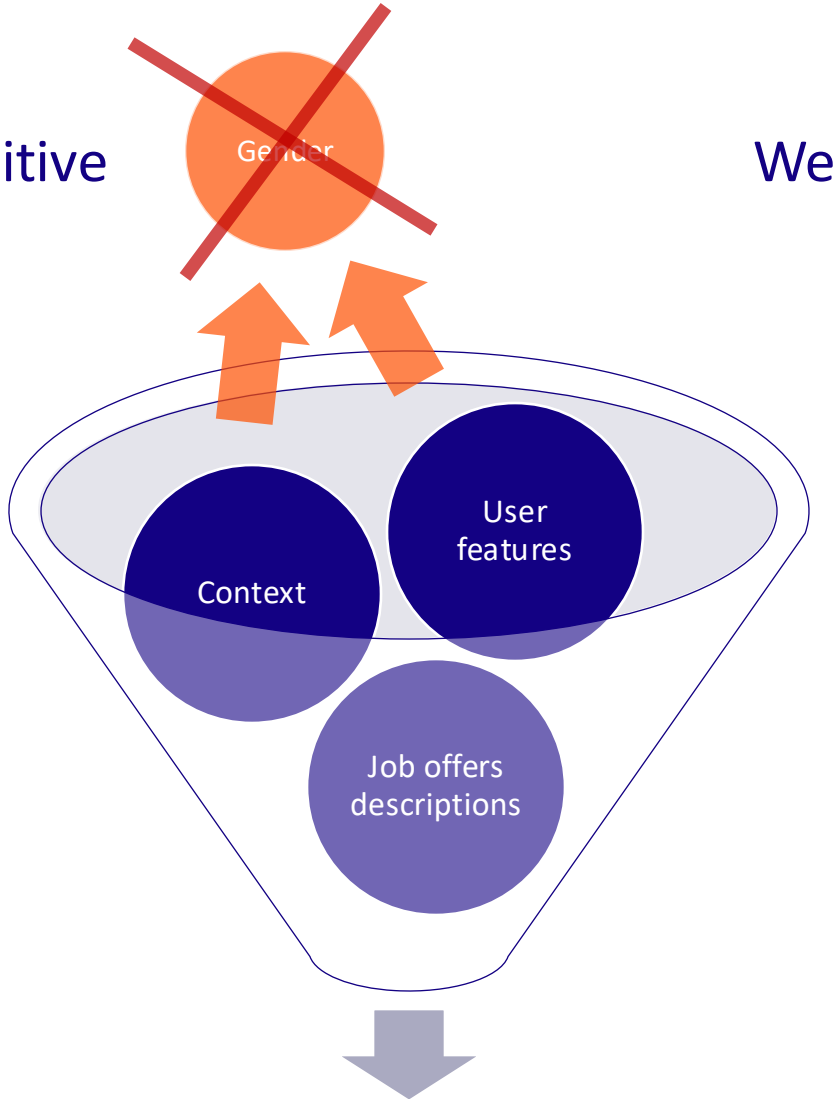
Protected attributes

We **do not have access** to sensitive information



CRITEO

We **do not have access** to sensitive information



We need an **approximation** for investigation

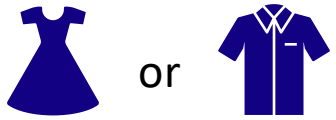
CRITEO

Gender “proxy”

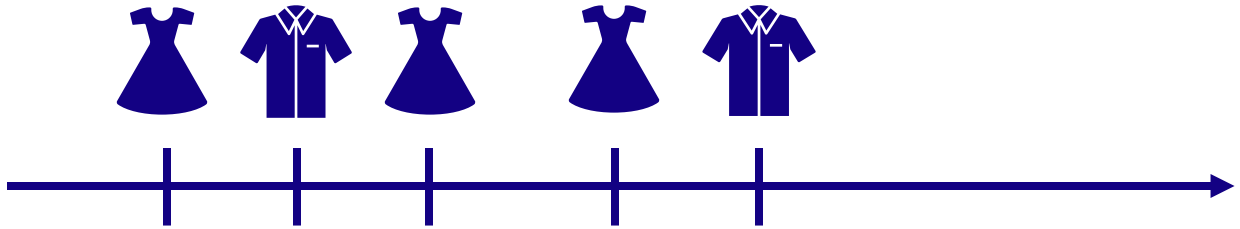
Product has a gender: or

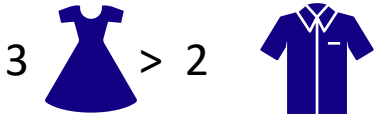
CRITEO

Gender “proxy”

Product has a gender:  or

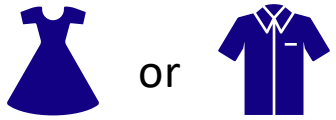
User timeline with products:



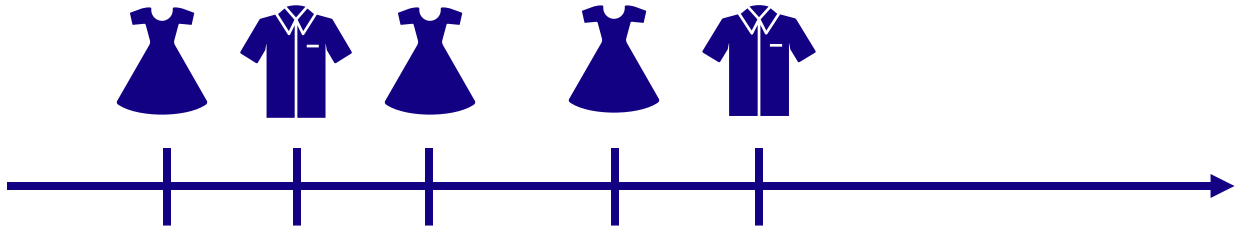
Female 

CRITEO

Gender “proxy”

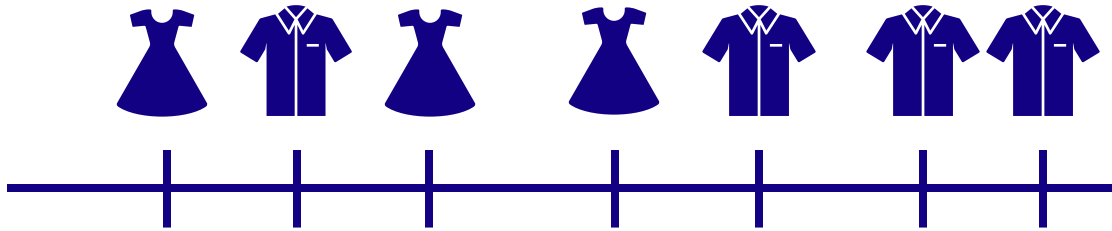
Product has a gender:  or

User timeline with products:



Female

3 > 2

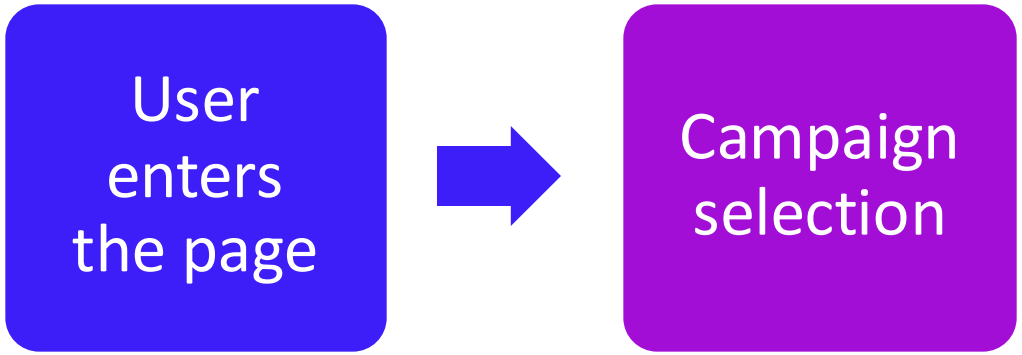


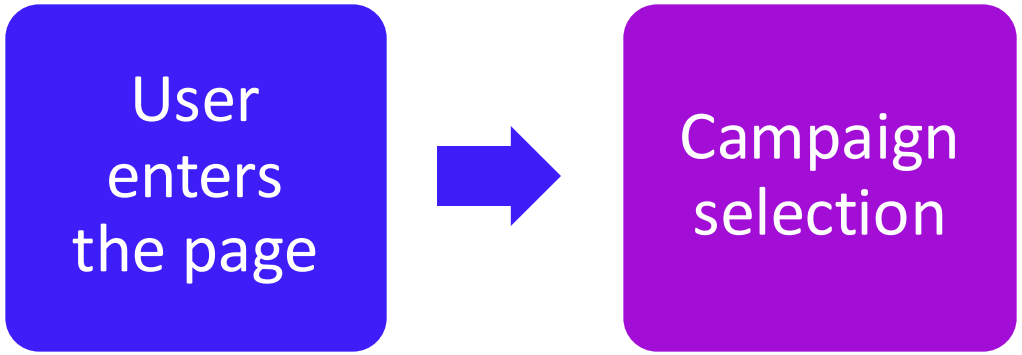
Male

3 < 4

CRITEO

How does bias appear with respect to gender “proxy”?

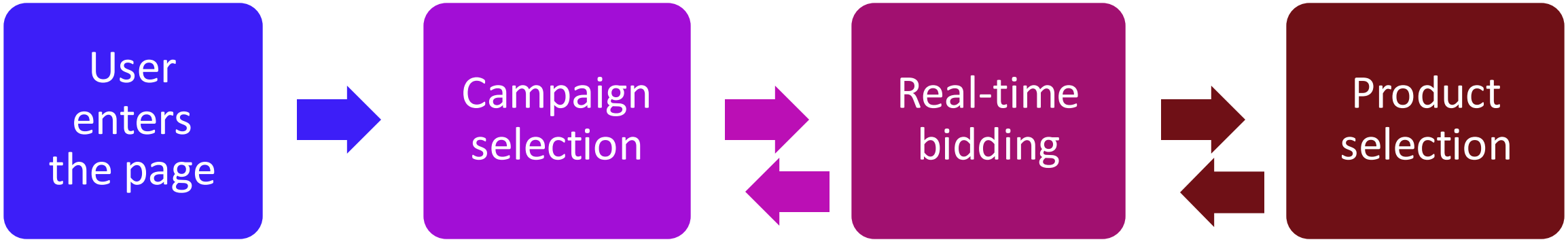

User enters the page
Campaign selection



Selection bias

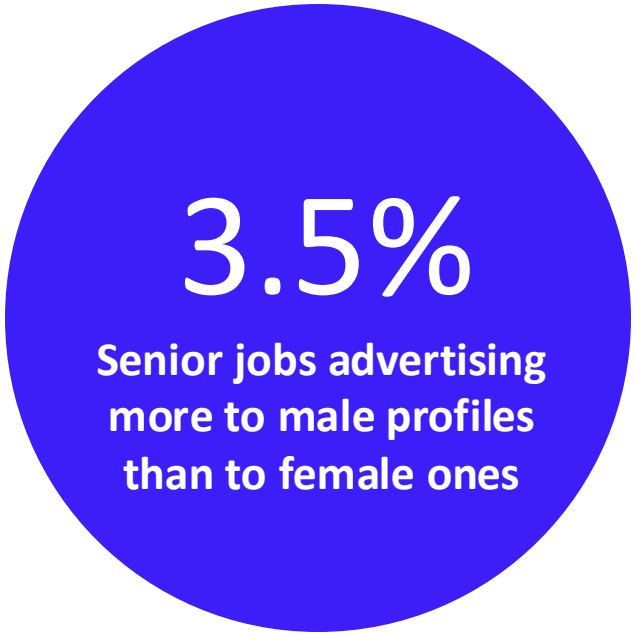
population	advertising	job campaign	ratio
female	53.6%	39.2%	0.73
male	46.4%	60.7%	1.31

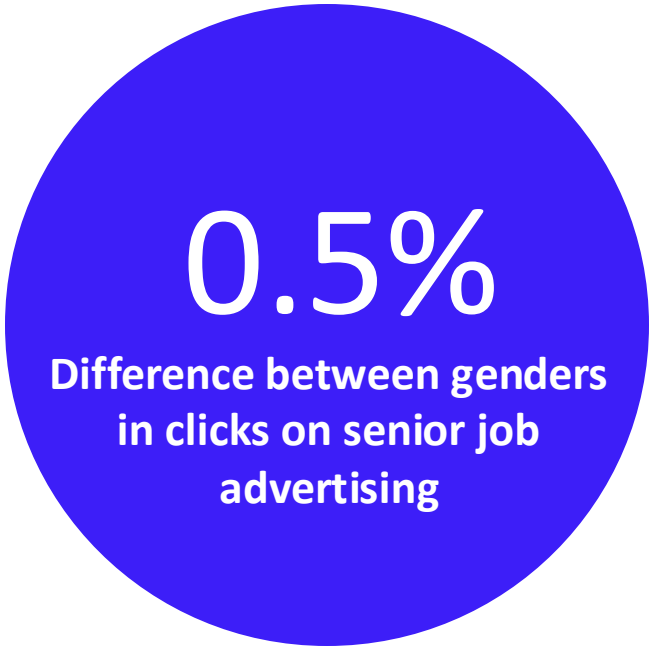
- Women profiles are a prized demographics.
- Competition among advertisers.
- The feedback loop mechanism may amplify bias.



	not clicked		clicked		
	non-senior	senior	non-senior	senior	all
female	189982	342221	1274	2636	536113
male	166394	366140	917	2662	536113
all	356376	708361	2191	5298	1072226

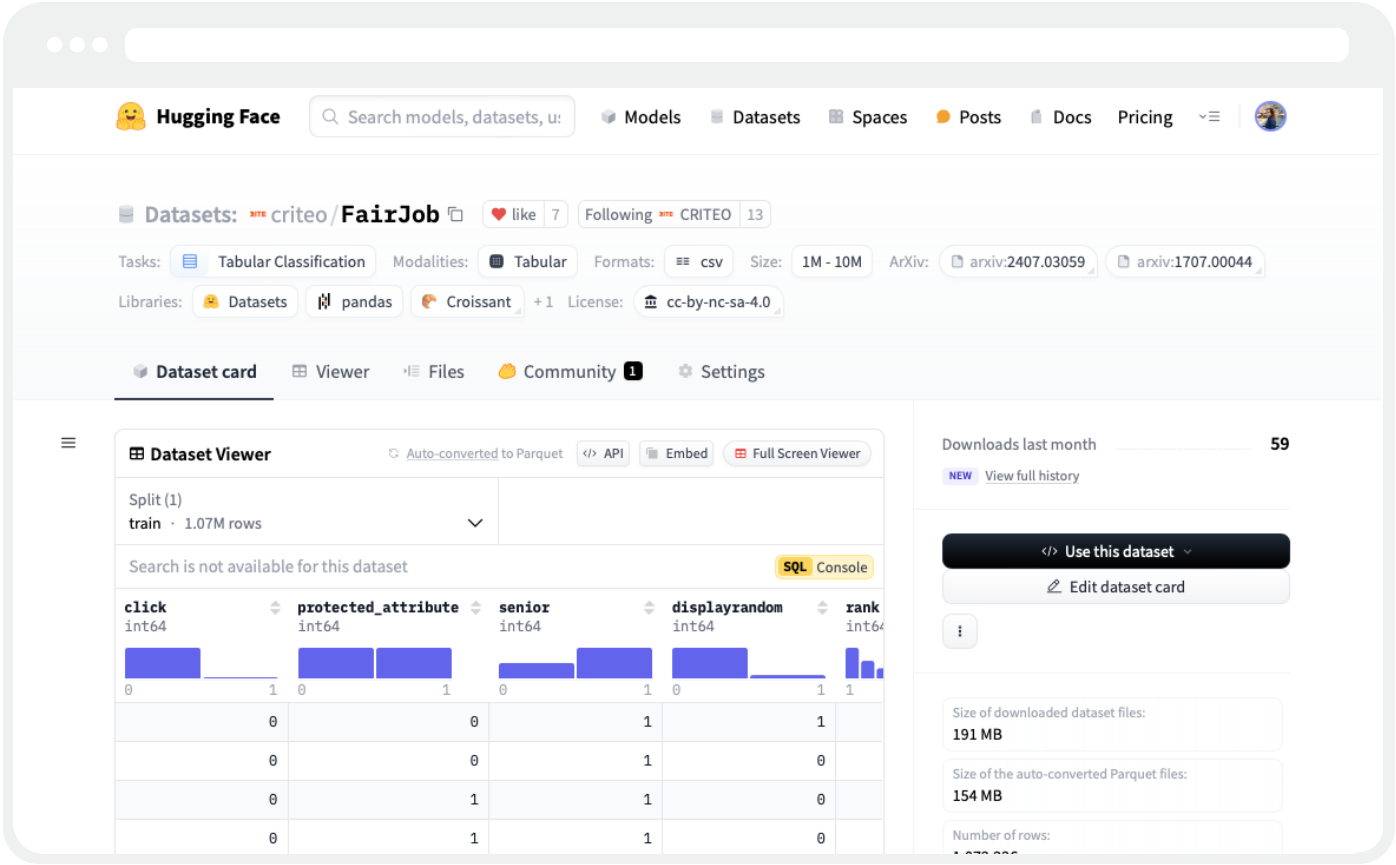
20%
Job campaign targeted more male profiles than female ones


3.5%
Senior jobs advertising more to male profiles than to female ones


0.5%
Difference between genders in clicks on senior job advertising

FairJob dataset on 🤗

- The **first fairness-aware dataset** for advertising job offers
- More than **900 downloads** since June 2024
- Adapted **utility metric** to take into account selection bias
- **Baseline** experiments on bias-utility trade-off



CRITEO

Our current solutions

Complex system

Analyze step by step

Bias-utility trade-off

Release dataset
Benchmarking

Protected attributes

Proxy

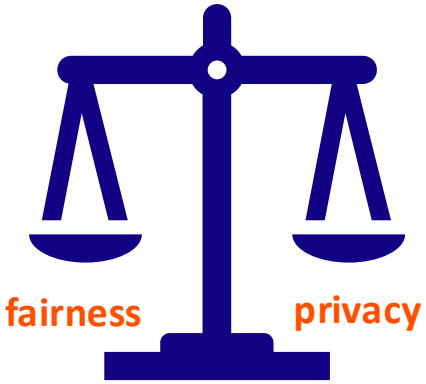
Conclusions

Benchmarking

CRITEO

Conclusions

Benchmarking

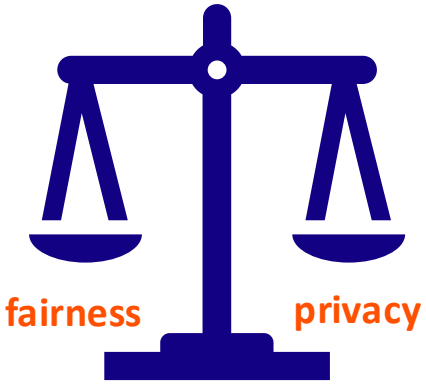


Protected attribute access

CRITEO

Conclusions

Benchmarking



Protected attribute access

Fairness definition

CRITEO

CRITEO

Thank you!