Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison

Amifa Raj

BOISE STATE UNIVERSITY

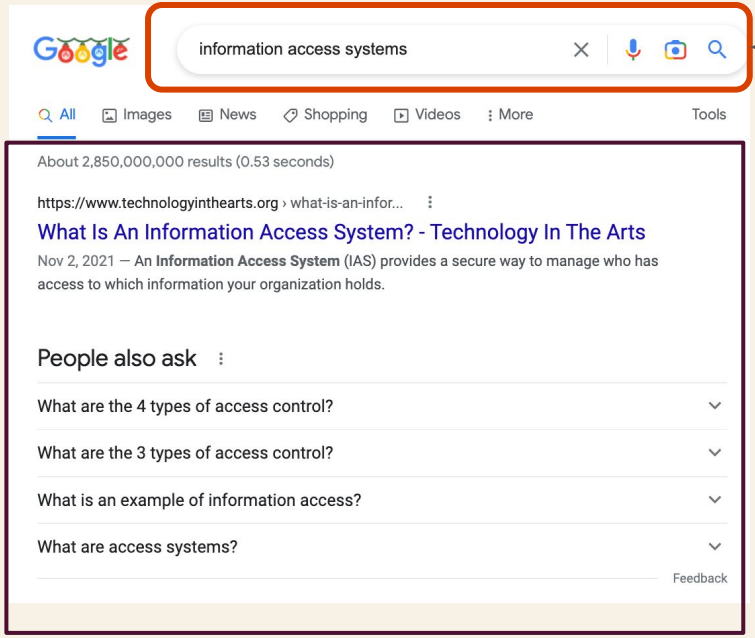
People and Information Research Team

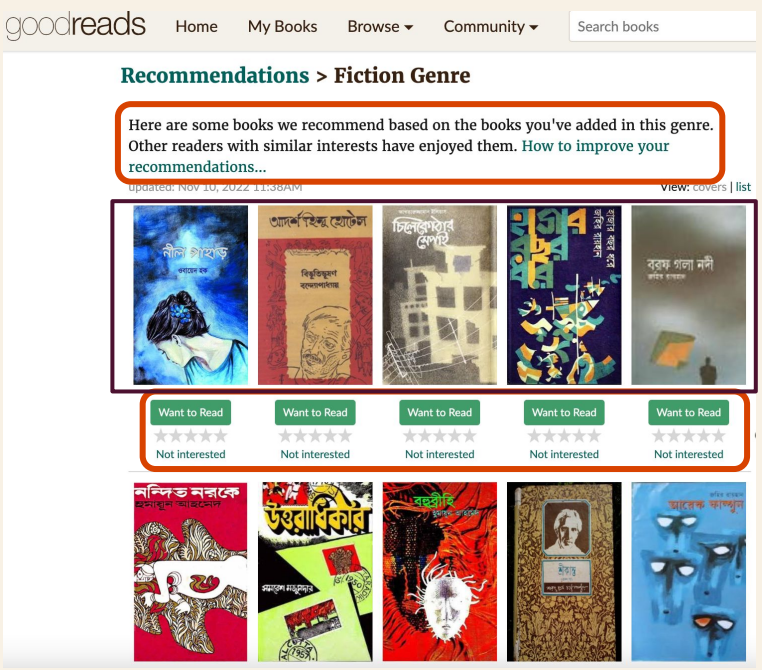
NSF

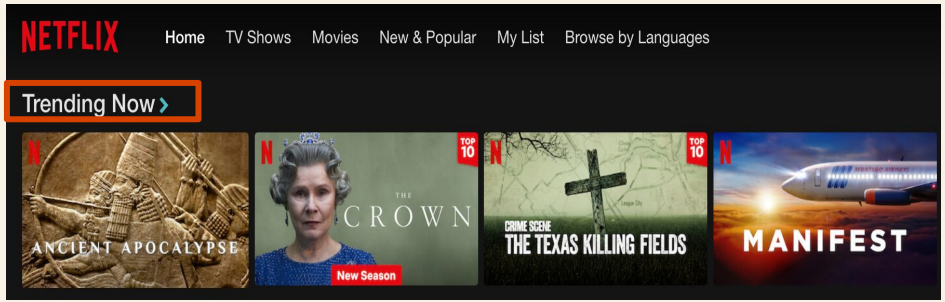
Grant No. IIS 17-51278

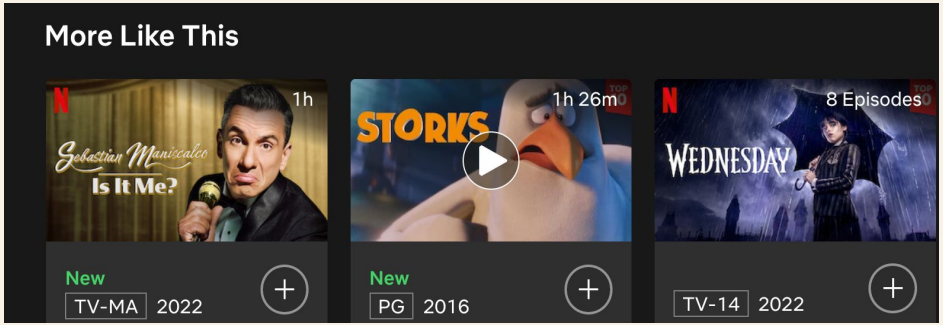

2022

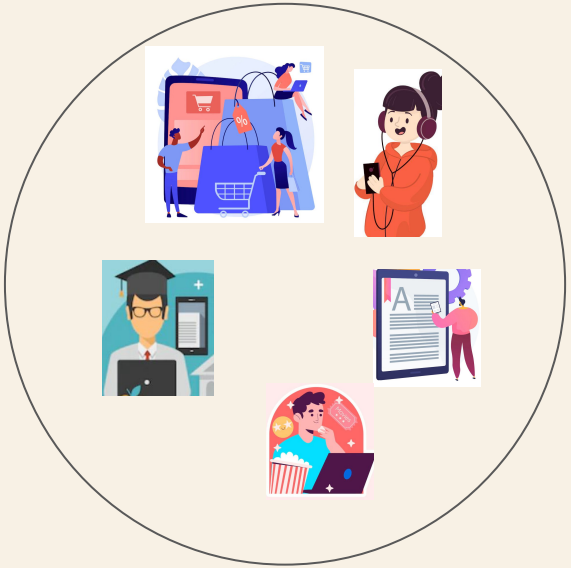
Raj, A., & Ekstrand, M. D. (2022, July). Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison. In Proceedings of the 45th International ACM SIGIR Conference on
DOI 10.1145/3477495.3532018


Google
information access systems
query
All
Images
News
Shopping
Videos
More
Tools
About 2,850,000,000 results (0.53 seconds)
https://www.technologyinthearts.org › what-is-an-infor...
What Is An Information Access System? - Technology In The Arts
Nov 2, 2021 — An Information Access System (IAS) provides a secure way to manage who has access to which information your organization holds.
People also ask
What are the 4 types of access control?
What are the 3 types of access control?
What is an example of information access?
What are access systems?
Feedback


goodreads
Home
My Books
Browse
Community
Search books
Recommendations > Fiction Genre
Here are some books we recommend based on the books you've added in this genre. Other readers with similar interests have enjoyed them. How to improve your recommendations...
Want to Read
Not interested

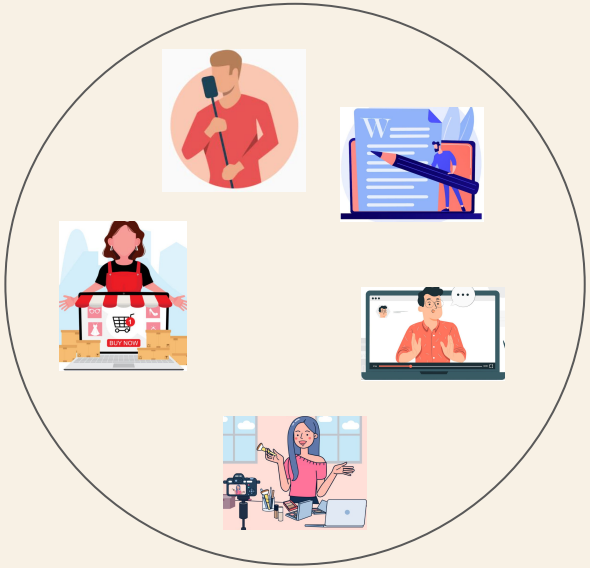

NETFLIX
Home
TV Shows
Movies
New & Popular
My List
Browse by Languages
Trending Now
ANCIENT APOCALYPSE
THE CROWN
New Season
CRIME SCENE
THE TEXAS KILLING FIELDS
MANIFEST

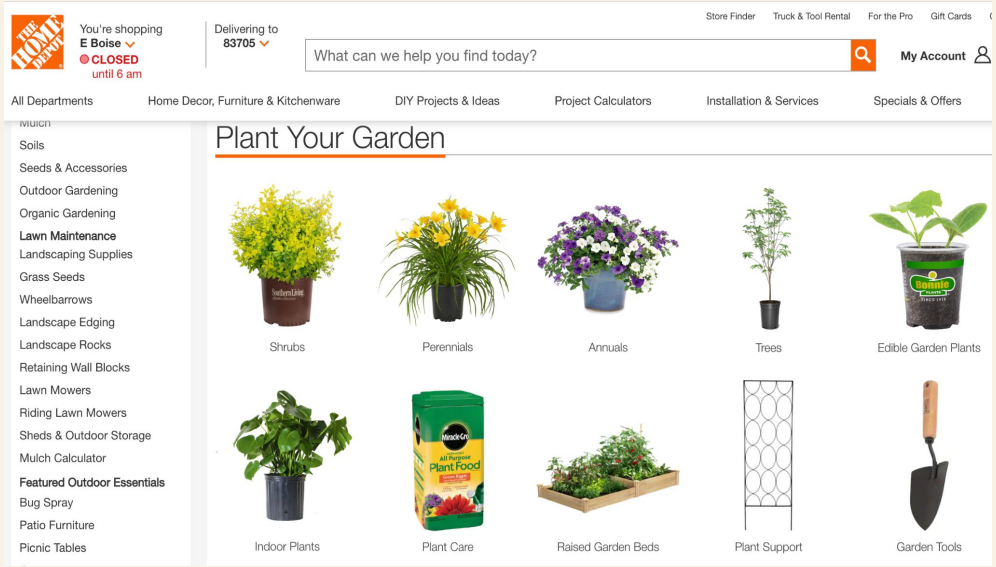

More Like This
Sebastian Maniscalco
Is It Me?
1h
New
TV-MA
2022
STORKS
1h 26m
New
PG
2016
WEDNESDAY
8 Episodes
TV-14
2022

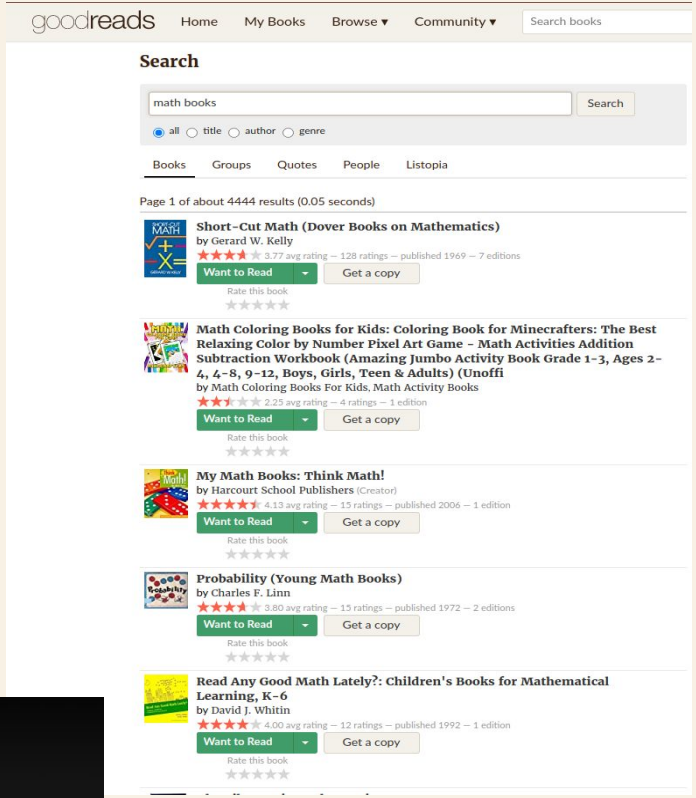


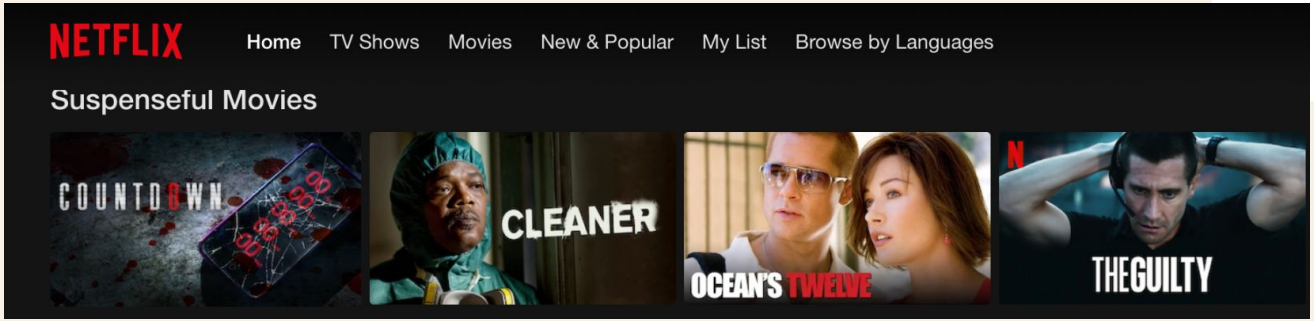
Provider

Consumer




You're shopping
E Boise
CLOSED
until 6 am
Delivering to
83705
What can we help you find today?
Store Finder
Truck & Tool Rental
For the Pro
Gift Cards
My Account
All Departments
Home Decor, Furniture & Kitchenware
DIY Projects & Ideas
Project Calculators
Installation & Services
Specials & Offers
Soils
Seeds & Accessories
Outdoor Gardening
Organic Gardening
Lawn Maintenance
Landscaping Supplies
Grass Seeds
Wheelbarrows
Landscape Edging
Landscape Rocks
Retaining Wall Blocks
Lawn Mowers
Riding Lawn Mowers
Sheds & Outdoor Storage
Mulch Calculator
Featured Outdoor Essentials
Bug Spray
Patio Furniture
Picnic Tables
Plant Your Garden
Shrubs
Perennials
Annuals
Trees
Edible Garden Plants
Indoor Plants
Plant Care
Raised Garden Beds
Plant Support
Garden Tools


goodreads
Home
My Books
Browse
Community
Search books
Search
math books
Search
all
title
author
genre
Books
Groups
Quotes
People
Listopia
Page 1 of about 4444 results (0.05 seconds)
Short-Cut Math (Dover Books on Mathematics)
by Gerard W. Kelly
3.77 avg rating — 128 ratings — published 1969 — 7 editions
Want to Read
Get a copy
Rate this book
Math Coloring Books for Kids: Coloring Book for Minecrafters: The Best Relaxing Color by Number Pixel Art Game - Math Activities Addition Subtraction Workbook (Amazing Jumbo Activity Book Grade 1-3, Ages 2-4, 4-8, 9-12, Boys, Girls, Teen & Adults) (Unoffi
by Math Coloring Books For Kids, Math Activity Books
2.25 avg rating — 4 ratings — 1 edition
Want to Read
Get a copy
Rate this book
My Math Books: Think Math!
by Harcourt School Publishers (Creator)
4.13 avg rating — 15 ratings — published 2006 — 1 edition
Want to Read
Get a copy
Rate this book
Probability (Young Math Books)
by Charles F. Linn
3.80 avg rating — 15 ratings — published 1972 — 2 editions
Want to Read
Get a copy
Rate this book
Read Any Good Math Lately?: Children's Books for Mathematical Learning, K-6
by David J. Whitin
4.00 avg rating — 12 ratings — published 1992 — 1 edition
Want to Read
Get a copy
Rate this book


NETFLIX
Home
TV Shows
Movies
New & Popular
My List
Browse by Languages
Suspenseful Movies
COUNTDOWN
CLEANER
OCEAN'S TWELVE
THE GUILTY


TIME

SUBSC

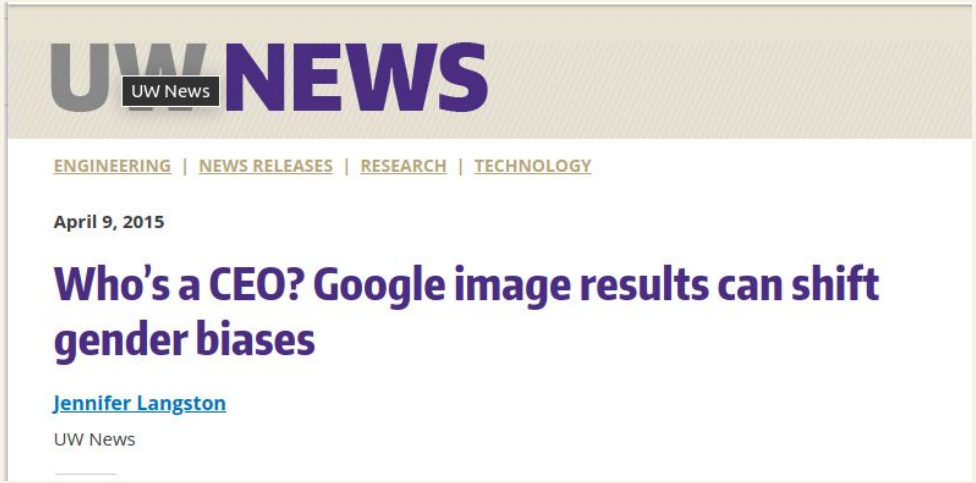
IDEAS • TECH BIAS

Google Has a Striking History of Bias Against Black Girls

BY SAFIYA NOBLE MARCH 26, 2018 4:30 PM EDT

IDEAS

Dr. Safiya U. Noble is the author of Algorithms of Oppression: How Search Engines Reinforce Racism and is an assistant professor of communication at the University of Southern California, Annenberg School of Communication & Journalism. She is a partner in Stratelligence and co-founder of the Information Ethics & Equity Institute.


UW NEWS

UW News

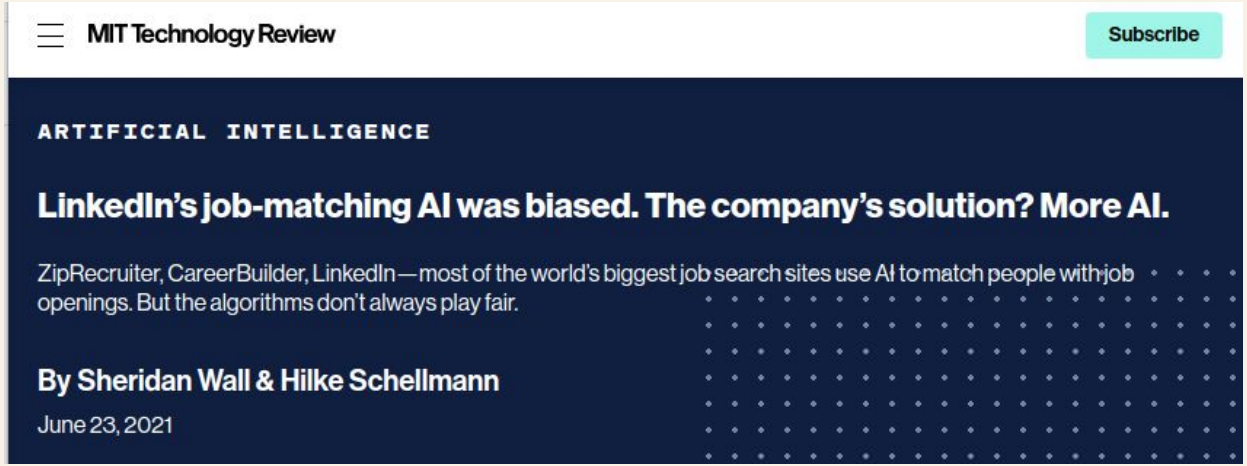
ENGINEERING | NEWS RELEASES | RESEARCH | TECHNOLOGY

April 9, 2015

Who's a CEO? Google image results can shift gender biases

Jennifer Langston

UW News


MIT Technology Review

Subscribe

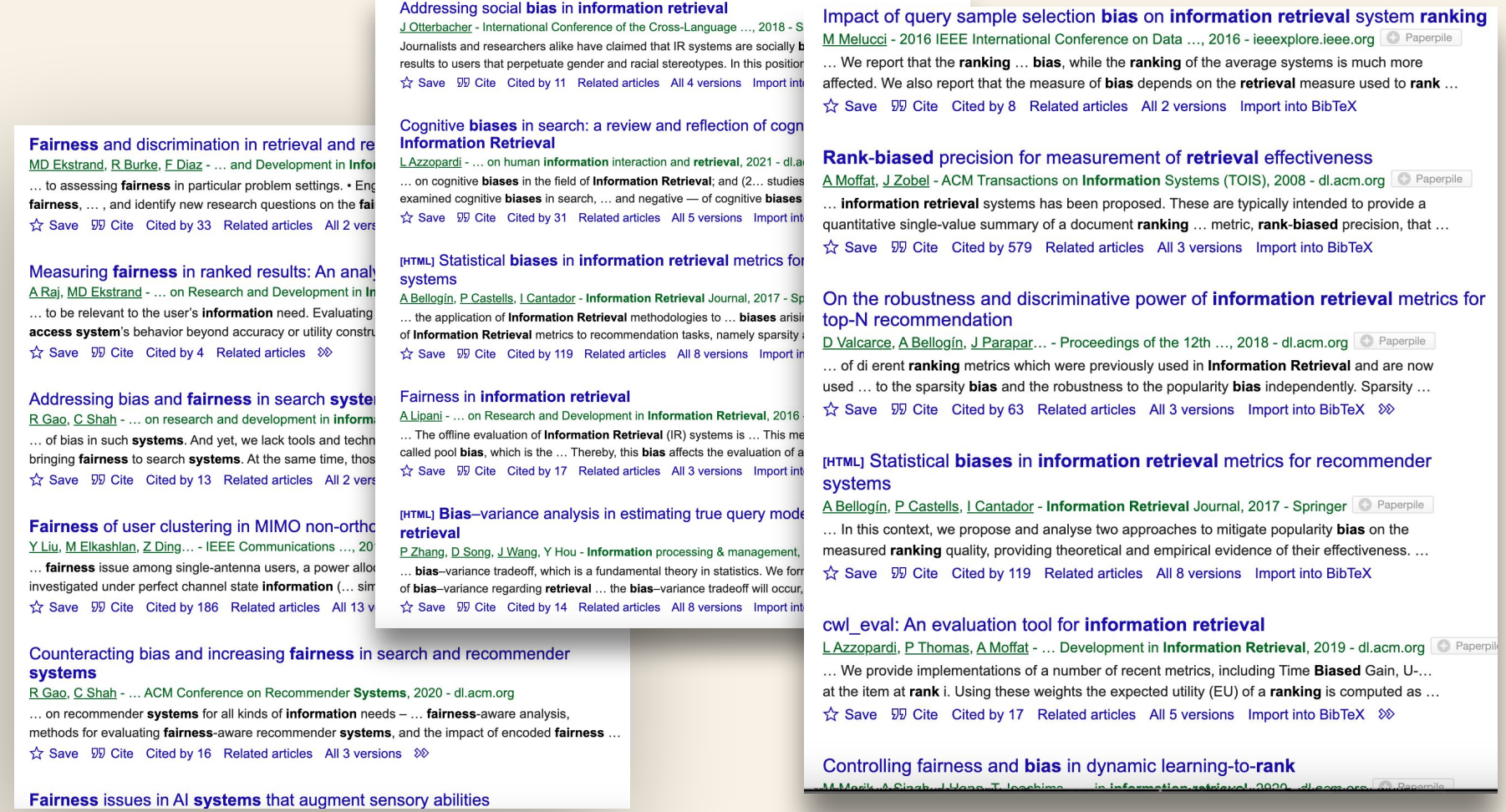
ARTIFICIAL INTELLIGENCE

LinkedIn's job-matching AI was biased. The company's solution? More AI.

ZipRecruiter, CareerBuilder, LinkedIn—most of the world's biggest job search sites use AI to match people with job openings. But the algorithms don't always play fair.

By Sheridan Wall & Hilke Schellmann

June 23, 2021



Fairness and discrimination in retrieval and re
MD Ekstrand, R Burke, F Diaz - … and Development in **Infor**
… to assessing **fairness** in particular problem settings. • Eng
fairness, … , and identify new research questions on the **fai**
☆ Save 99 Cite Cited by 33 Related articles All 2 vers

Measuring **fairness** in ranked results: An analy
A Raj, MD Ekstrand - … on Research and Development in **In**
… to be relevant to the user's **information** need. Evaluating
access system's behavior beyond accuracy or utility constru
☆ Save 99 Cite Cited by 4 Related articles »

Addressing bias and **fairness** in search **syste**
R Gao, C Shah - … on research and development in **inform**
… of bias in such **systems**. And yet, we lack tools and techn
bringing **fairness** to search **systems**. At the same time, thos
☆ Save 99 Cite Cited by 13 Related articles All 2 vers

Fairness of user clustering in MIMO non-ortho
Y Liu, M Elkashlan, Z Ding… - IEEE Communications …, 20
… **fairness** issue among single-antenna users, a power allo
investigated under perfect channel state **information** (… sim
☆ Save 99 Cite Cited by 186 Related articles All 13 v

Counteracting bias and increasing **fairness** in search and recommender **systems**
R Gao, C Shah - … ACM Conference on Recommender **Systems**, 2020 - dl.acm.org
… on recommender **systems** for all kinds of **information** needs – … **fairness**-aware analysis,
methods for evaluating **fairness**-aware recommender **systems**, and the impact of encoded **fairness** …
☆ Save 99 Cite Cited by 16 Related articles All 3 versions »

Fairness issues in AI **systems** that augment sensory abilities

Addressing social **bias** in **information retrieval**
J Otterbacher - International Conference of the Cross-Language …, 2018 - S
Journalists and researchers alike have claimed that IR systems are socially **b**
results to users that perpetuate gender and racial stereotypes. In this positio
☆ Save 99 Cite Cited by 11 Related articles All 4 versions Import int

Cognitive **biases** in search: a review and reflection of cogn
Information Retrieval
L Azzopardi - … on human **information** interaction and **retrieval**, 2021 - dl.a
… on cognitive **biases** in the field of **Information Retrieval**; and (2… studies
examined cognitive **biases** in search, … and negative — of cognitive **biases**
☆ Save 99 Cite Cited by 31 Related articles All 5 versions Import int

[HTML] Statistical **biases** in **information retrieval** metrics for
systems
A Bellogín, P Castells, I Cantador - **Information Retrieval** Journal, 2017 - Sp
… the application of **Information Retrieval** methodologies to … **biases** arisi
of **Information Retrieval** metrics to recommendation tasks, namely sparsity
☆ Save 99 Cite Cited by 119 Related articles All 8 versions Import in

Fairness in **information retrieval**
A Lipani - … on Research and Development in **Information Retrieval**, 2016
… The offline evaluation of **Information Retrieval** (IR) systems is … This me
called pool **bias**, which is the … Thereby, this **bias** affects the evaluation of a
☆ Save 99 Cite Cited by 17 Related articles All 3 versions Import int

[HTML] **Bias**–variance analysis in estimating true query mode
retrieval
P Zhang, D Song, J Wang, Y Hou - **Information** processing & management,
… **bias**–variance tradeoff, which is a fundamental theory in statistics. We forr
of **bias**–variance regarding **retrieval** … the **bias**–variance tradeoff will occur,
☆ Save 99 Cite Cited by 14 Related articles All 8 versions Import int

Impact of query sample selection **bias** on **information retrieval** system **ranking**
M Melucci - 2016 IEEE International Conference on Data …, 2016 - ieeexplore.ieee.org Paperpile
… We report that the **ranking** … **bias**, while the **ranking** of the average systems is much more
affected. We also report that the measure of **bias** depends on the **retrieval** measure used to **rank** …
☆ Save 99 Cite Cited by 8 Related articles All 2 versions Import into BibTeX

Rank-biased precision for measurement of **retrieval** effectiveness
A Moffat, J Zobel - ACM Transactions on **Information** Systems (TOIS), 2008 - dl.acm.org Paperpile
… **information retrieval** systems has been proposed. These are typically intended to provide a
quantitative single-value summary of a document **ranking** … metric, **rank-biased** precision, that …
☆ Save 99 Cite Cited by 579 Related articles All 3 versions Import into BibTeX

On the robustness and discriminative power of **information retrieval** metrics for top-N recommendation
D Valcarce, A Bellogín, J Parapar… - Proceedings of the 12th …, 2018 - dl.acm.org Paperpile
… of di erent **ranking** metrics which were previously used in **Information Retrieval** and are now
used … to the sparsity **bias** and the robustness to the popularity **bias** independently. Sparsity …
☆ Save 99 Cite Cited by 63 Related articles All 3 versions Import into BibTeX »

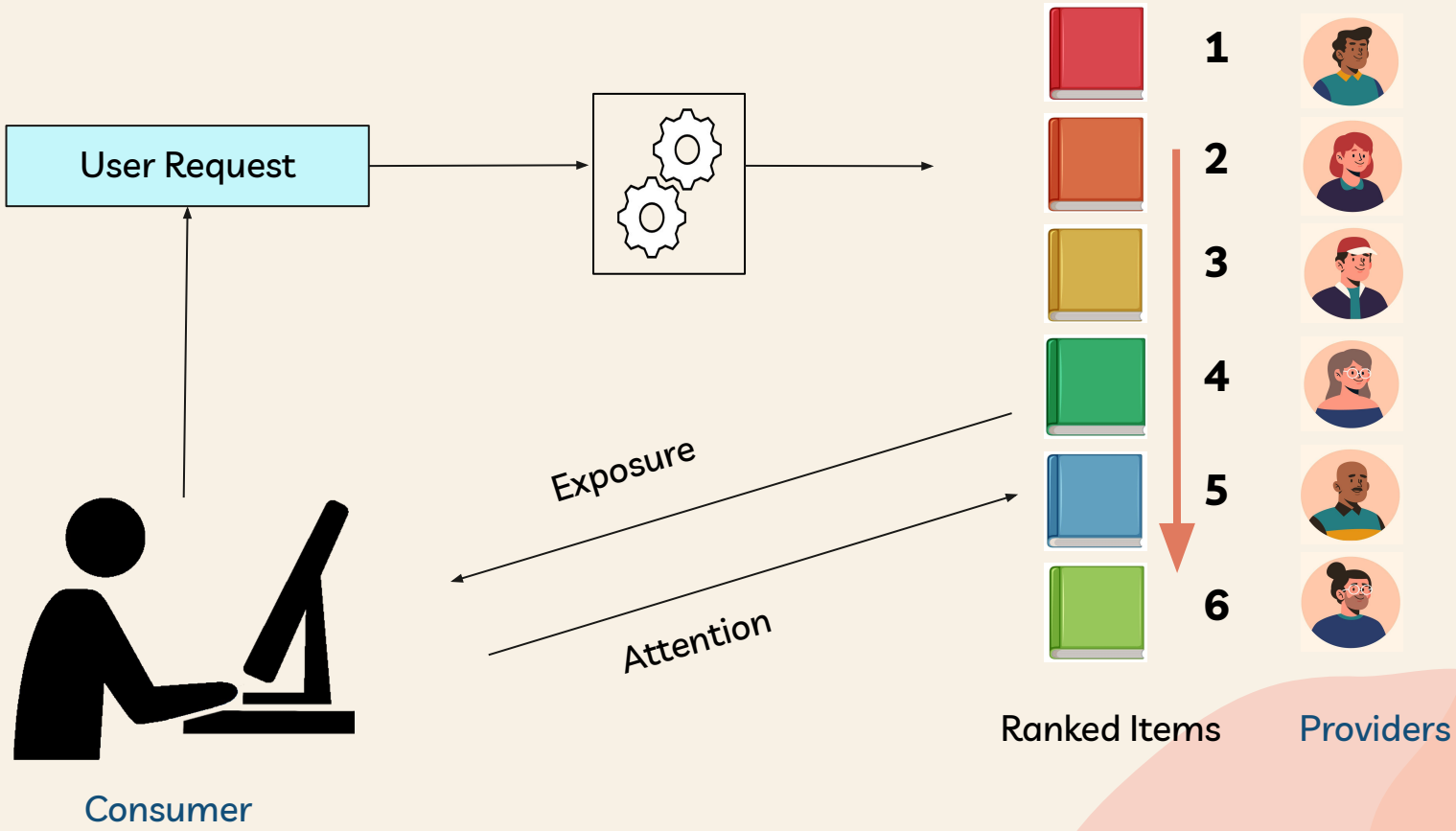
[HTML] Statistical **biases** in **information retrieval** metrics for recommender systems
A Bellogín, P Castells, I Cantador - **Information Retrieval** Journal, 2017 - Springer Paperpile
… In this context, we propose and analyse two approaches to mitigate popularity **bias** on the
measured **ranking** quality, providing theoretical and empirical evidence of their effectiveness. …
☆ Save 99 Cite Cited by 119 Related articles All 8 versions Import into BibTeX

cwl_eval: An evaluation tool for **information retrieval**
L Azzopardi, P Thomas, A Moffat - … Development in **Information Retrieval**, 2019 - dl.acm.org Paperpil
… We provide implementations of a number of recent metrics, including Time **Biased** Gain, U-…
at the item at **rank** i. Using these weights the expected utility (EU) of a **ranking** is computed as …
☆ Save 99 Cite Cited by 17 Related articles All 5 versions Import into BibTeX »

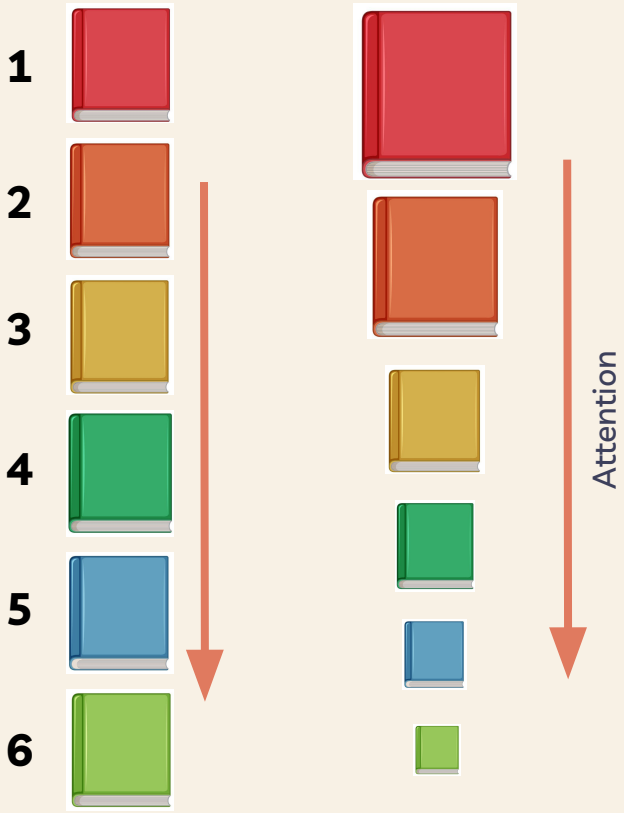
Controlling fairness and **bias** in dynamic learning-to-**rank**

Bias

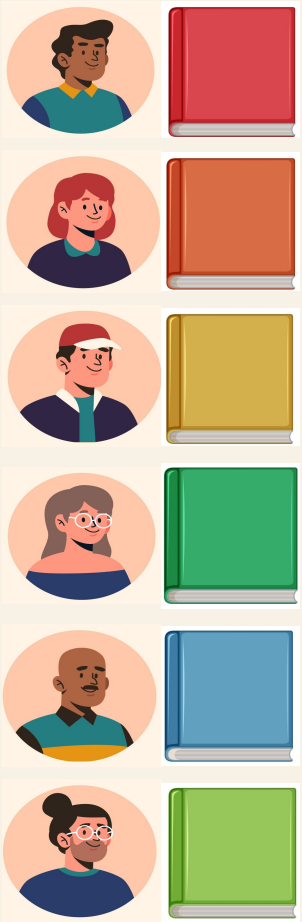
- Difficult to define
- Domain dependent
- Systematic and unfair discrimination against certain **individual** or **group** entities by denying opportunity and assigning unfair outcomes
- Group (Sensitive Attributes) and Individual Fairness


User Request
Exposure
Attention
1
2
3
4
5
6
Consumer
Ranked Items
Providers

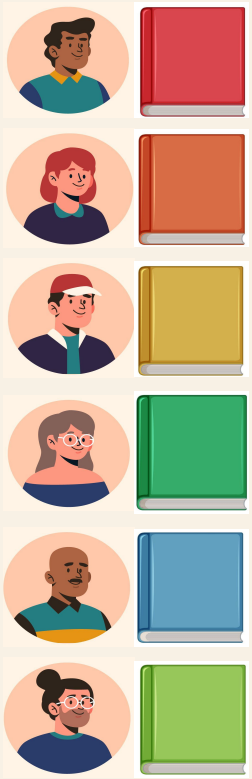
Disparate Exposure



Ranked List



Fairness Positioning



Provider Fairness

Geography

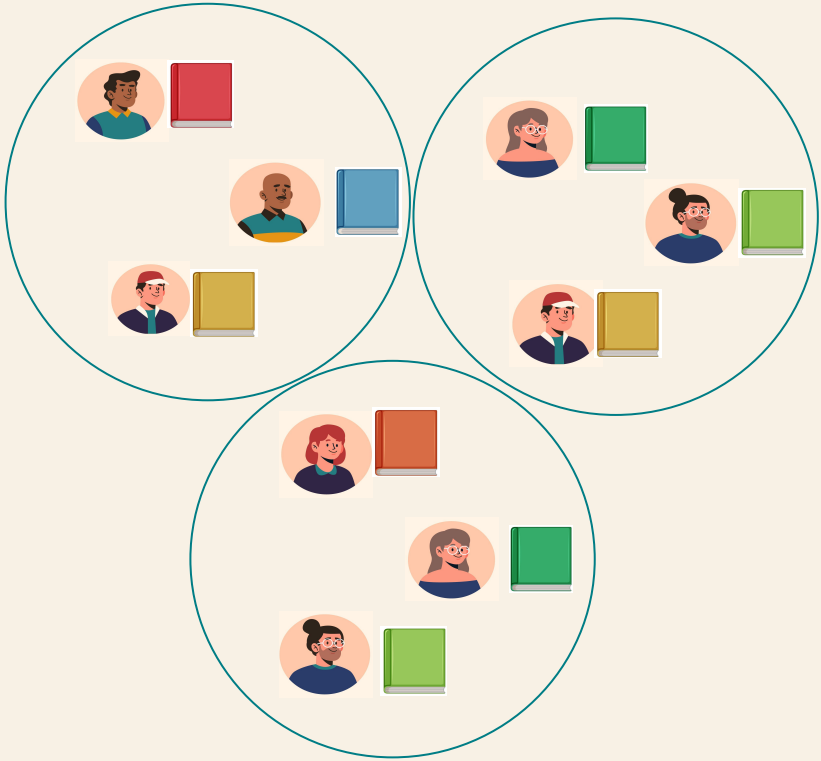
Age

Race

Gender

Economic status

Sensitive Attributes



Group Fairness

Fair Ranking Metrics

PreFΔ
(Yang et. al.; SSDBM ‘17):

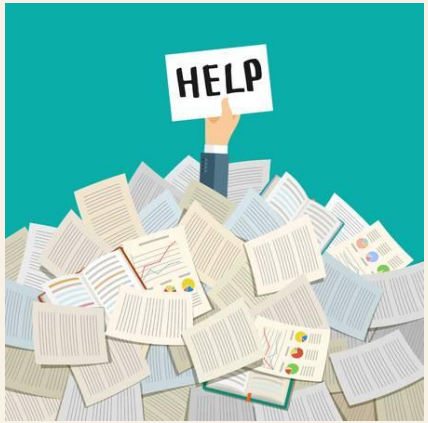
AWRF
(Sapienzynski et. al.; WWW’19)

EEL, EED, EER
Diaz et.al.; CIKM’20)

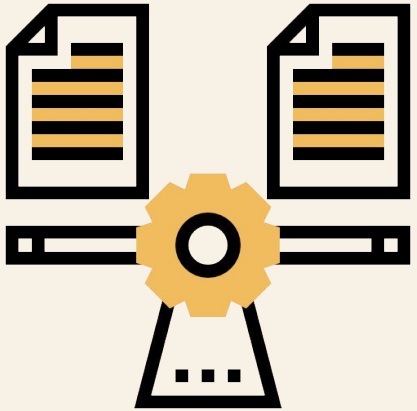
FAIR
(Zehlike et.al.; CIKM’17)

IAA
(Biega et. al.; SIGIR’18)

DP, EUR, RUR
(Singh et.al.; KDD’18)



Several Fair Ranking Metrics



No Comparative and Comprehensive Analysis

Why is the Problem a Problem?

Finding suitable Metrics

Differences among the Metrics

Implementation in Real-world IAS dataset

Research Questions

RQ1. What are the conceptual differences among the fair ranking metrics?

RQ2. What is needed to apply these metrics to real IAS?

RQ3. What are the design decisions and parameters involved, and how sensitive are the resulting metrics to those decisions?

RQ4. What are the empirical differences in how these metrics assess the relative fairness of different recommendation algorithms or retrieval runs?

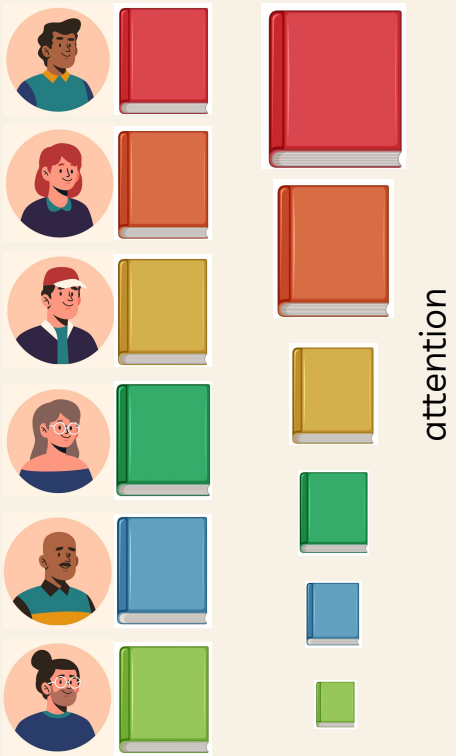
Research Tasks

Conceptual Analysis of Fair Ranking Metrics

Implementing Fair Ranking Metrics in Real-World IAS Datasets

Sensitivity Analysis

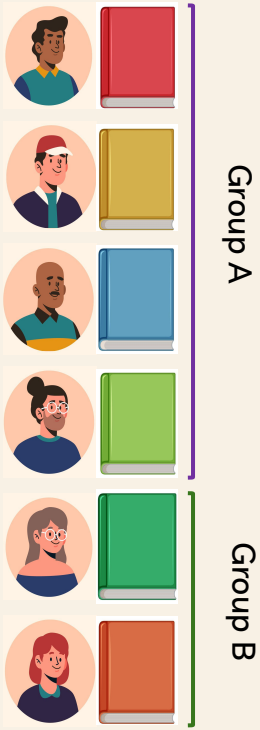
Metrics Design Decomposition



Fairness Goal

What does it mean to be fair?

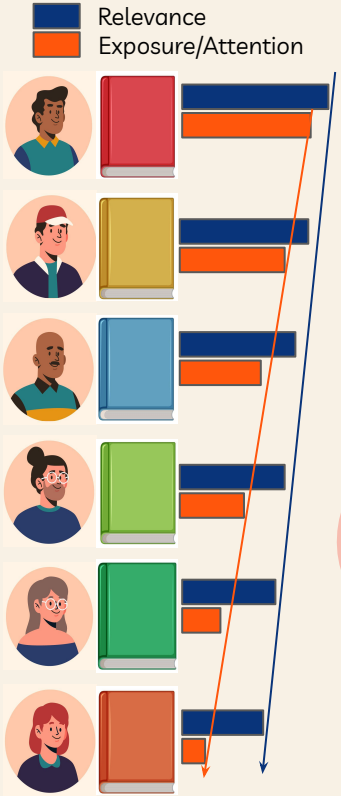
Fairness Goal



PreFΔ, FAIR, AWRF, DP, EED

Item position should not be affected by membership

Statistical Parity

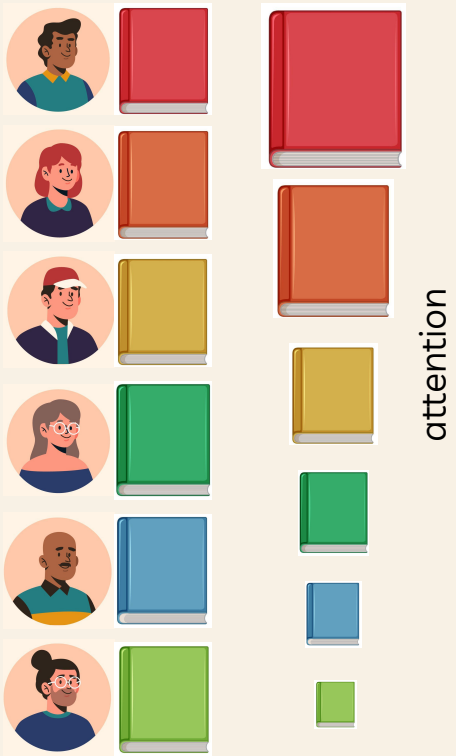


IAA, EUR, RUR, EEL, EER

Exposure/attention should be proportional to relevance

Equal Opportunity

Metrics Design Decomposition



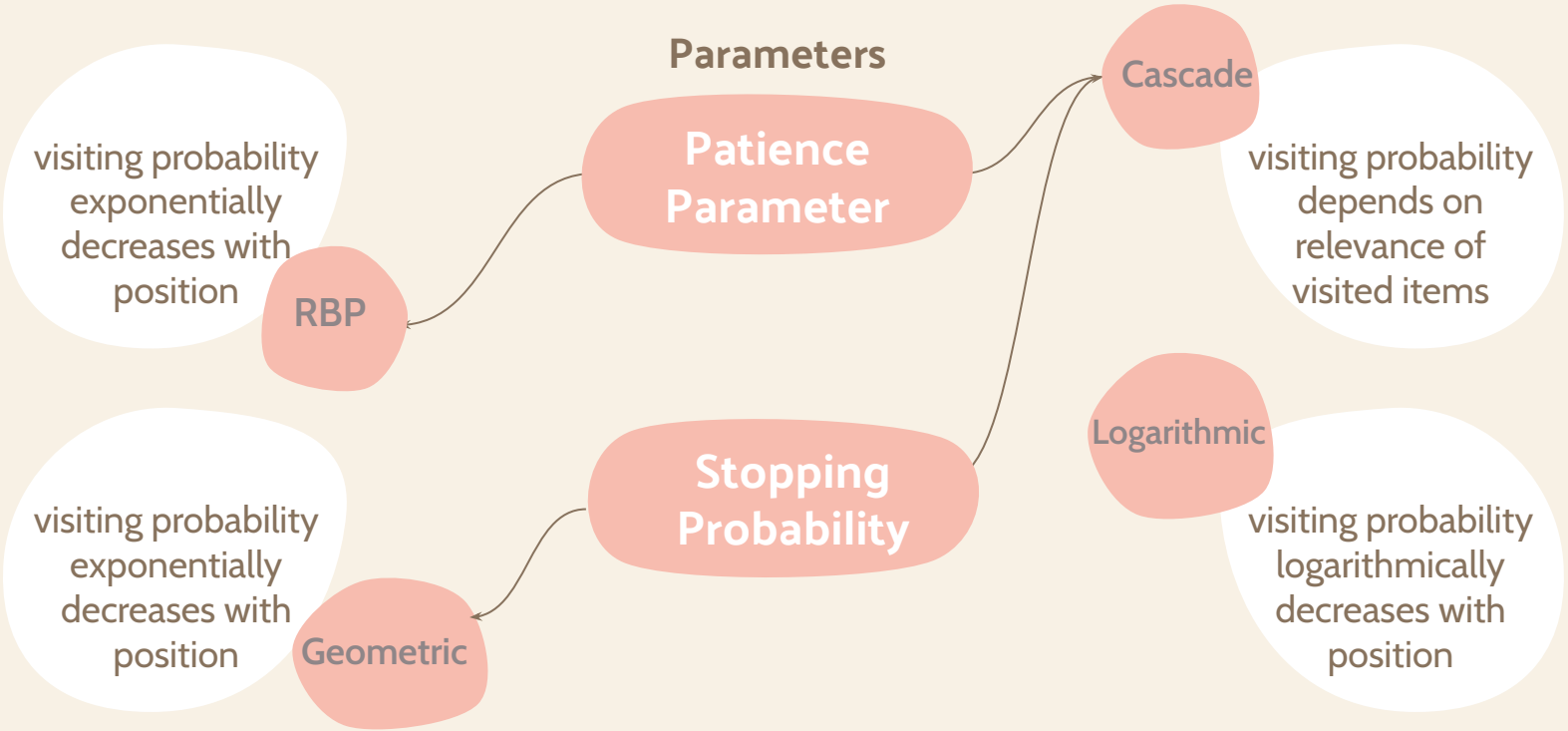
Fairness Goal

What does it mean to be fair?

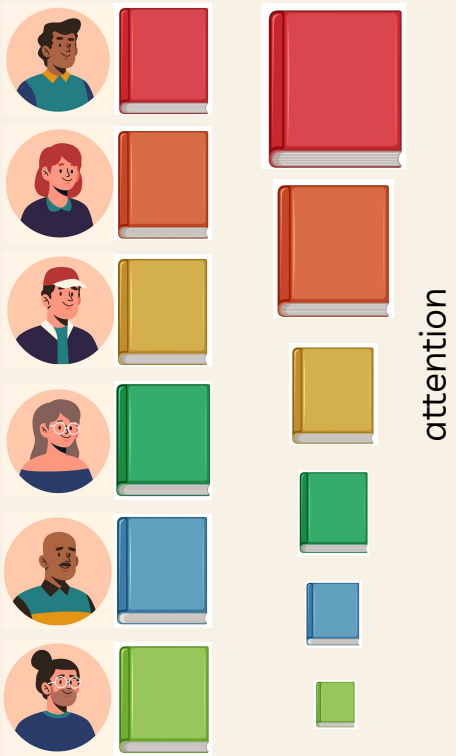
Browsing Model

How to measure position weight?

Browsing Models



Metrics Design Decomposition



Fairness Goal

What does it mean to be fair?

Browsing Model

How to measure position weight?

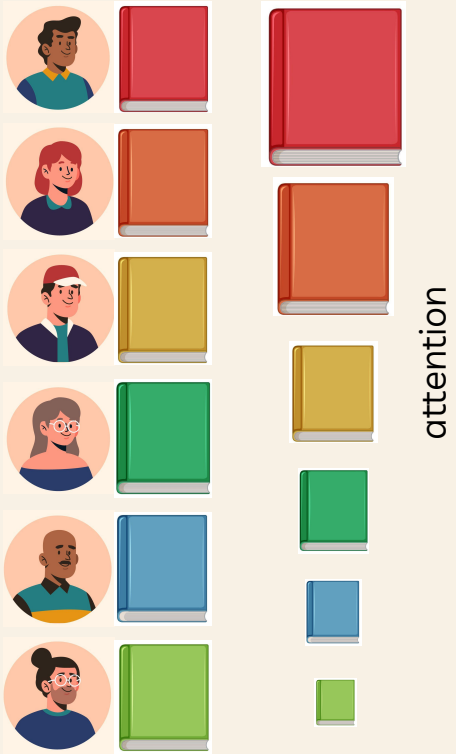
Target Exposure

Compare system exposure with what?

Target Exposure

- Population estimator
  - From full ranking
  - Configured
- Ideal exposure based on relevance
- Estimated utility (Predicted relevance)

Metrics Design Decomposition



Fairness Goal

What does it mean to be fair?

Browsing Model

How to measure position weight?

Target Exposure

Compare system exposure with what?

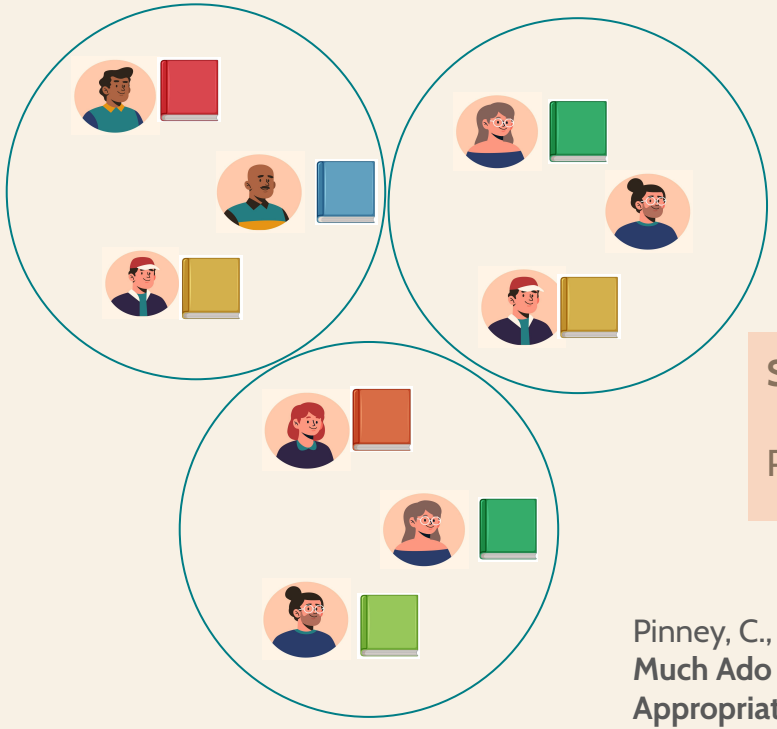
Relevance

How to incorporate relevance?

Group Membership

Does it allow multinomial and soft group association?

Group Membership



Multinomial Protected Attributes

Non-Binary Groups, such as gender

Soft Group Association

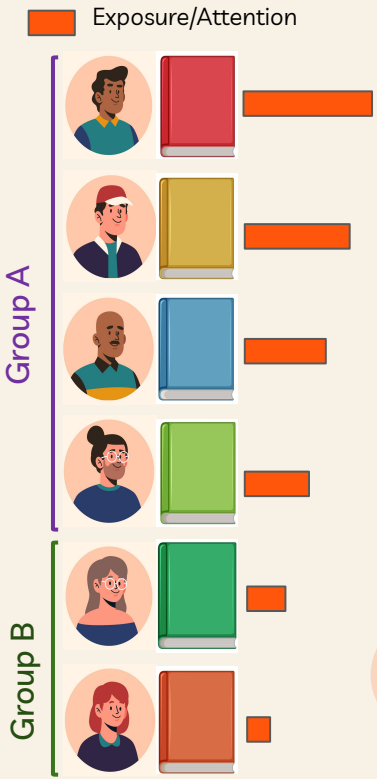
Partial or mixed group membership such as race

Pinney, C., Raj, A., Hanna, A., & Ekstrand, M. D. (2023)
Much Ado About Gender: Current Practices and Future Recommendations for Appropriate Gender-Aware Information Access.
To appear in CHIIR 2023 proceedings.
https://arxiv.org/abs/2301.04780

Summary of Fair Ranking Metrics

Metric(s)	Goal	Weighting	Relevance	Binomial?
PreFd	Each prefix representative of whole ranking	✗	✗	Dep on d
FAIR	Each prefix matches target distribution	✗	✗	✓
AWRF	Weighted representation matches population	Geometric	✗	✗
DP	Exposure equal across groups	Logarithmic	✗	✓
EUR	Exposure proportional to relevance	Logarithmic	✓	✓
RUR	Discounted gain proportional to relevance	Logarithmic	✓	✓
IAA	Exposure proportional to predicted relevance	Geometric	Predicted	✗
EEL, EER	Exposure matches ideal (from relevance)	Cascade, Geom	✓	✗
EED	Exposure well-distributed	Cascade, Geom	✗	✗

Statistical Parity



AWRF

(Sapienzynski et. al, WWW'19)

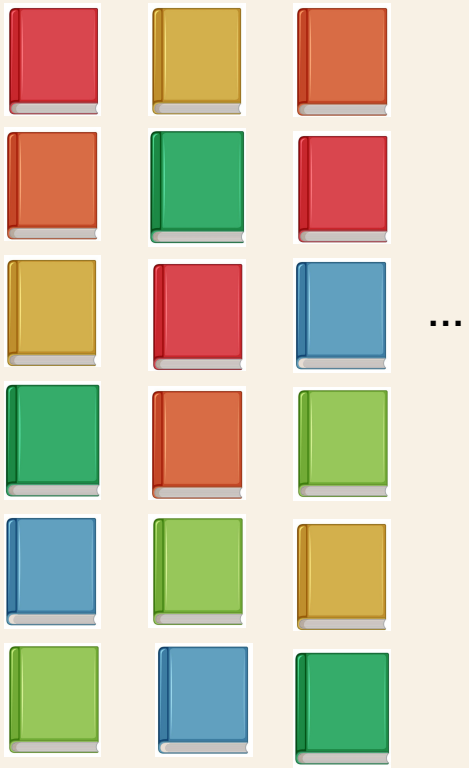
Expected cumulative exposure(Group B x position weight) >=p

Target distribution is the group distribution in entire ranked list (true demographics)

- no relevance information
- geometric attention decay
- non-binary group membership
- uses a target distribution to compare

PreFΔ (Yang et. al, SSDBM'17) and FAIR (Zehlike et. al, CIKM'17) differ in measuring position weight and allowing multinomial groups.

Sequences of Ranking



...

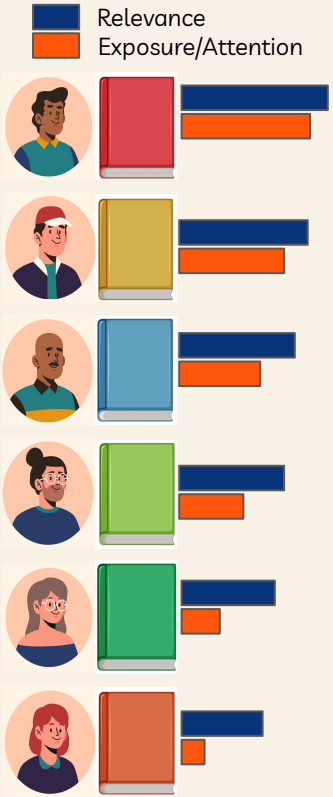
Statistical Priority

DP, EED

Equal Opportunity

IAA, EUR, RUR, EER, EEL

Equal Opportunity



EE*
(Diaz et. al, CIKM'20)

EEL(Expected Exposure Loss): ||target-system||2

EER (Expected Exposure Relevance): Exposure-relevance distribution

- stochastic ranking
- rbp & cascade attention decay
- non-binary group membership

IAA (Biega et. al, SIGIR'18) differs in weighting strategy, group membership, and relevance

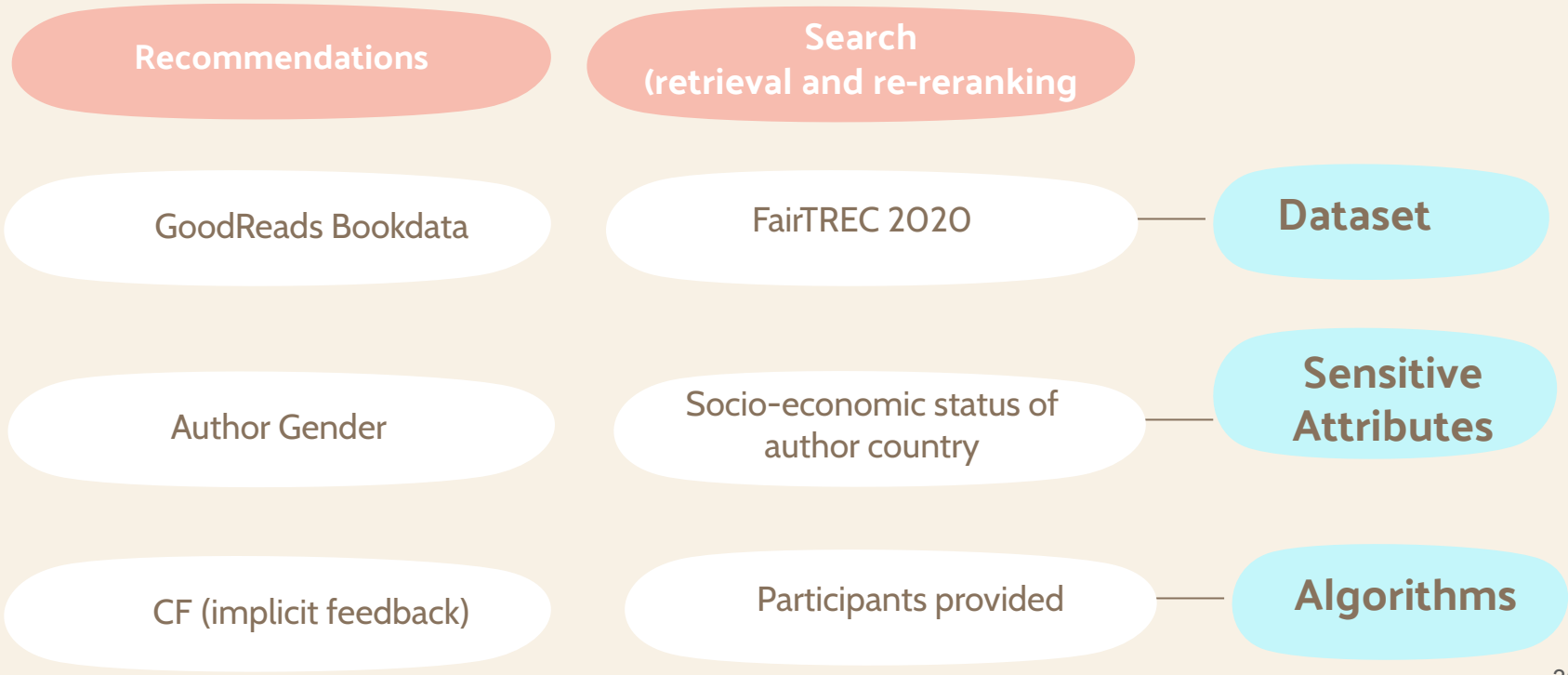
EUR, RUR (Singh et. al, SIGKDD'18) differs in weighting strategy and group membership

Task 1 Findings

Task 1: Conceptual Analysis Fair Ranking Metrics

- Metrics are conceptually similar with common components like relevance, browsing model, aggregation, target exposure
- Metrics differs in their design choices and fairness assumption
- Metrics with same goal can have different design choices

Implementing the Metrics



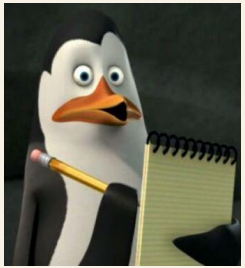
https://fair-trec.github.io/

Challenges in Implementation

Missing Relevance Information

IAA, EE*, DP, EUR, RUR

AWRF, IAA, DP, EUR, RUR, EE*



Parameter Setting

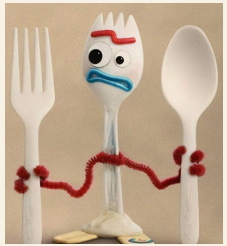
Missing Group Label

All the metrics

Extreme Imbalance

- **PreFΔ** and **RUR:** suffer from missing data (sparsity) problem
- Reformulated ratio-based metric to smoothed log ratio

PreFΔ, FAIR, IAA, DP, EUR, RUR



Soft Group Association
Non-binary groups

Task 2 Findings

Task 2: Implementing Fair Ranking Metrics in Real-World Datasets

- Missing data, missing relevance information, ranked list size are crucial/delicate factors in implementing metrics.
- Metrics with similar fairness goals differ in their ease of implementations

Sensitivity Analysis

Ranked-list size

- No effect on metrics for FairTREC
- Ratio-based metrics and FAIR showed sensitivity

Weighting Strategy

- Default parameters
- EEL and logRUR showed high sensitivity

Parameter Settings

- Almost all metrics showed sensitivity
- logRUR is extremely sensitive

Task 3 Findings

Task 3: Sensitivity Analysis

- Metrics differ in their sensitivity towards external factors.
- High sensitivity towards design choices add complexity in the usability of metrics

Key Findings

Defining metrics in unified framework

- Metrics are surprisingly similar

Implement the metrics in same experimental setup

- Missing data, missing relevance information, ranked list size are crucial/delicate factors in implementing metrics.

Sensitivity Analysis

- Metrics differ in their sensitivity towards external factors.

Recommendations

	Allow multinomial protected attributes	Allow soft group association	Sensitivity towards design choices
Single-list metrics FAIR, AWRF	AWRF	AWRF	AWRF
Demographic Parity in Sequence DP, EED	EED	EED	EED
Equal Opportunity in Sequence EUR, RUR, IAA, EER, EEL	EER, EEL	EER, EEL	EER, EEL, IAA

Research Directions

- Simulation study to understand the impact of crucial factors in metric implementation.
- Incorporating various browsing models
- Missing label
- Missing or sparse relevance
- Ambiguous or multiple group association
- Robust, explainable, and efficient metric design

THANK You!

Credits: This presentation template was created **by Slidesgo**, including icons **by Flaticon**, and infographics & images **by Freepik**