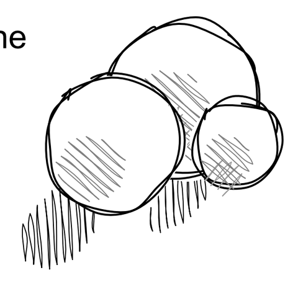
Computational Models of Fairness for Tomorrow's Web

Seán Healy
ADAPT Centre
Dublin City University

Tim Berners-Lee,
writing on the 35th anniversary of the world wide web:

> There are two clear, connected issues to address. The first is the extent of power concentration, which contradicts the decentralised spirit I originally envisioned [for the web]. This has segmented the web, with a fight to keep users hooked on one platform, to optimise profit through the passive observation of content […]. Compounding this issue is the second, the personal data market that has exploited people's time and data.



Joy Buolamwini (in "Unmasking AI"):

> As tempting as it may be, we cannot use AI to sidestep the hard work of organizing society so that where you are born, the resources of your community, and the labels placed upon you are not the primary determinants of your destiny.

Abstract

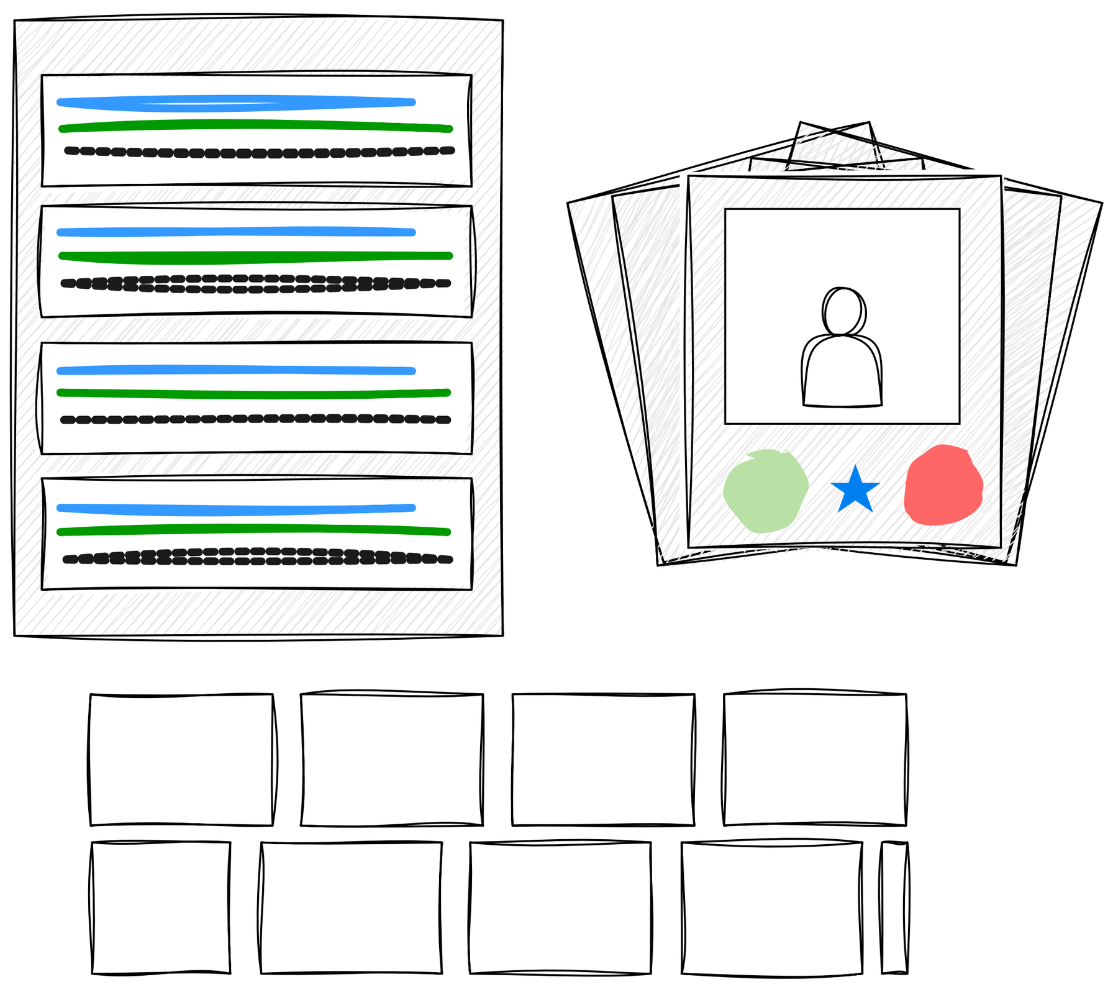
The web is the world's digital society, and there is now strong EU funding to organise this digital society in a way that is more equitable to stakeholders on both sides of the screen: users and businesses on the web.

Two notable fields within computer science are currently addressing "ethical issues" within web search:

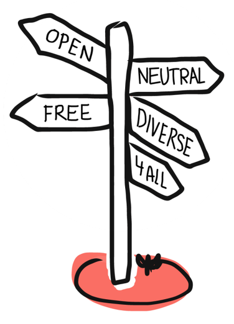
The academic field of "fair ranking" looks at lists and grids of items within web platforms, and asks the following research questions:

- "Do these displays exhibit societal biases in the ways they display data?" (these are termed "pre-existing biases" in the literature)
- "Is the display populated with results in a way that proliferates some new bias? (this is referred to as "emergent bias")
- "Is the display biased by design? (e.g. too few result slots) (this is often discussed as "technical bias")

Fair ranking researchers focus less on whether their proposed solutions work within the paradigm of a decentralised web, working more on ensuring that protected groups and individuals are treated fairly within a list of results on a webpage. Proposed solutions may require inordinate computing resources (as is the case with the current state-of-the-art for something called *individually fair re-ranking*).



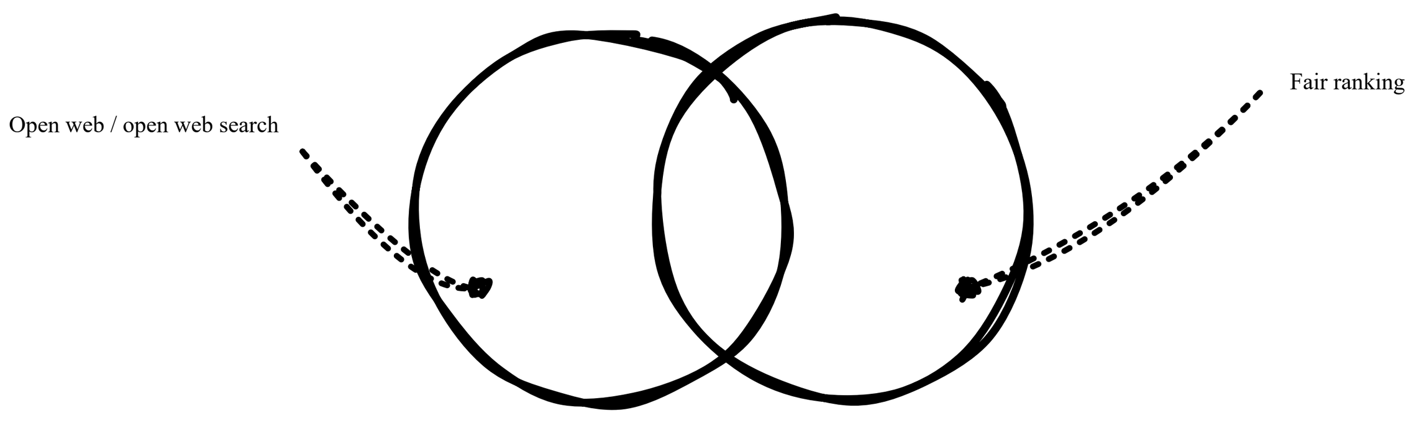
Simultaneous to this "fair ranking" research drive, another field known as "open (web) search" aims to address transparency and accountability issues in the current search landscape.



- The aim of this academic field is to uproot big tech solutionism from the task of managing the world's web search and recommendation needs.
- A lot of work in this field aims to establish shared protocols and technical standards for distributing the task of web search across a federation of participating search servers (funded by the EU).

Researchers working on *open search* problems place their research focus on the accountability and transparency problems inherent in centralised search solutions, focusing less on equal treatment of groups and individuals within webpage displays.

These research fields don't always see eye-to-eye; in one paper, listed on the openwebsearch.eu website, the authors highlight how a popular technique in fair re-ranking (FairCo), when put to use for statistically validating the fairness of an election news feed, can easily be manipulated by the service provider to feign 'equal exposure' of left and right-leaning candidates'; the authors call for a 'more holistic' approach to fairness issues on the web.



My research sits at the intersection of these two fields. Rather than discarding work from either of these research efforts, I ask the following research question:

RQ Can fair re-ranking techniques even work in open web search settings, or does this field pre-suppose a centralised view of the web?

Choosing Appropriate Research Methodologies

Experimental Computer Science (ECS)

- Useful methodology for answering questions like:
  - "How much better/fairer" does method X perform, as compared to method Y?"
  - "Do users feel better about system X over system Y?"
- This side of computer science is closer to the natural sciences in terms of research methodology.
- The primary scientific method used in ECS is *inference*.
- Evaluation tends to happen via statistical significance testing, AVOVA, etc.

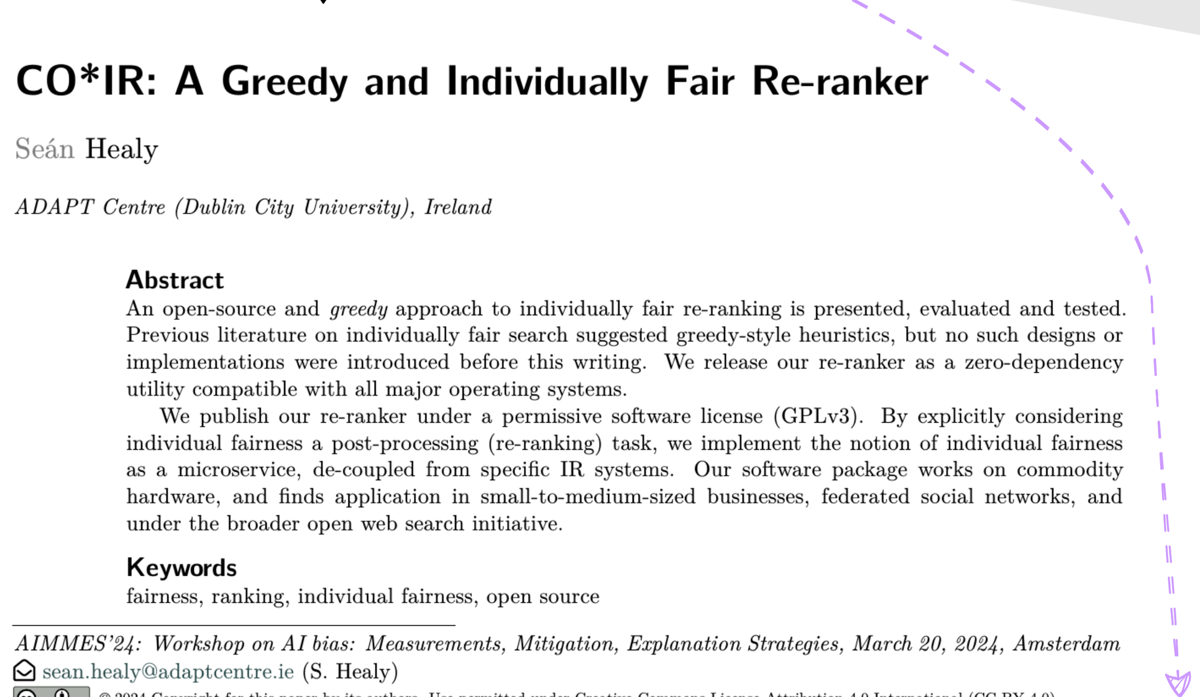
Theoretical Computer Science (TCS)

- This is a useful methodology for answering questions like:
  - "Is X even possible?" (in theory)
- This side of computer science overlaps heavily with mathematics and formal linguistics (occasionally publishing in philosophical journals).
- The primary scientific method used in TCS is *deduction*.
- Evaluation tends to involve mathematical proofs, algorithmic analysis, and formal verification methods.
- Proposed solutions in TCS are abstract, not targetting any particular application area, e.g. sorting algorithm research.

Corpus Linguistics / Critical Thinking

- Useful approach for highlighting broad issues or patterns in a field, e.g. via analyses of patterns within academic papers themselves.
- This technique was used by Hannah Devinney, for example, to highlight that many papers in fair-ML fixate their efforts on prescriptive theories of gender.
- Results are often presented discursively, or with basic tabulations of counts, leaving the statistical analysis work for readers.

Progress



CO*IR: A Greedy and Individually Fair Re-ranker

Seán Healy

ADAPT Centre (Dublin City University), Ireland

Abstract
An open-source and *greedy* approach to individually fair re-ranking is presented, evaluated and tested. Previous literature on individually fair search suggested greedy-style heuristics, but no such designs or implementations were introduced before this writing. We release our re-ranker as a zero-dependency utility compatible with all major operating systems.
We publish our re-ranker under a permissive software license (GPLv3). By explicitly considering individual fairness a post-processing (re-ranking) task, we implement the notion of individual fairness as a microservice, de-coupled from specific IR systems. Our software package works on commodity hardware, and finds application in small-to-medium-sized businesses, federated social networks, and under the broader open web search initiative.

Keywords
fairness, ranking, individual fairness, open source

AIMMES'24: Workshop on AI bias: Measurements, Mitigation, Explanation Strategies, March 20, 2024, Amsterdam
sean.healy@adaptcentre.ie (S. Healy)
© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Experiment 1

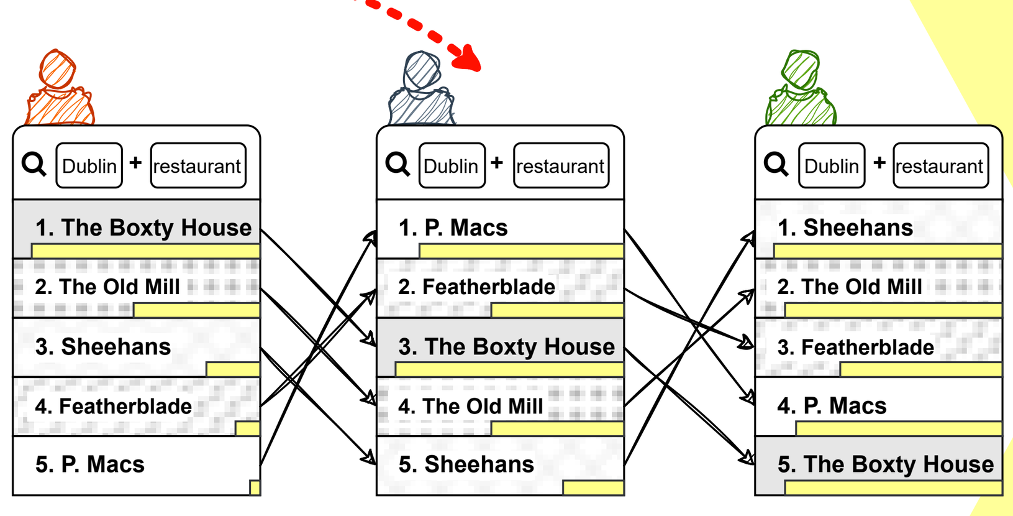
I used a TCS methodology to develop a new algorithm for something called "individual amortized fairness" with the aim of making this technique applicable in a vast array of computing environments.

The previously proposed methods were only suited for re-ranking a small number of results in ranking pages, e.g. for search and recommendation, and relied on the use of a proprietary tool for integer linear programming.

My method also allows us to 'optimise for fairness' with a more realistic quality threshold as a constraint (ERR instead of NDCG). These metrics are often used in IR evaluation as well.

I validated that this method works through statistical methods (ECS), but further comparisons to existing methods are pending.

The drawback of this new method is that it is sub-optimal. I will investigate more closely how sub-optimal it is in future work.



Research Methodology

For individual/group fairness techniques to work on an open web, each of the following research questions needs to be answered while assuming low computational resources:

1. Are the salient bias issues across varying search queries computable?
2. Is there a computationally performant theoretical model for individually fair re-ranking?
3. Can the decision of how much query merging to apply be partially automated?
   - Are there query merging techniques that will reliably never result in system degradation?

https://librecoir.com/

I maintain a C/C++ library for the fair re-ranking techniques I am developing as part of this PhD.

Experimental Plan

Datasets:

- AOL (web search queries, results, and clicks, from 2006).
- Yandex (obfuscated web search queries, results, and clicks, from 2010).
- SogouQ (search queries, results, and clicks, from 2008).

Relevancy data predicted via the PSCM model. Researchers in Tsinghua previously demonstrated this model to be an effective replacement for real user relevance feedback in ML tasks like Learning to Rank.

Experiment 3: Validate and quantify the need for 'query merging' in fair web search settings. (amortized fairness techniques rely on a large number of duplicate queries).

Experiment 4: Explore techniques for 'query merging' which work in low-resource settings (open web). Work-in-progress: improvements to finite-state-based spelling correction techniques (TCS).

Experiment 5: When do we stop the computer looking for a solution? (graph-theoretic fair re-ranking model.) In this experiment, I will investigate how greedy we can permit our graph-theoretic fair re-ranking model to be, before differences in fair outcomes become noticeable.

Bibliography

De Jonge, T., & Hiemstra, D. (2023, June). Unfair: Search engine manipulation, undetectable by amortized inequity. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 830-839).

Devinney, H., Björklund, J., & Björklund, H. (2022, June). Theories of "gender" in NLP bias research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2083-2102).

Zehlike, M., Yang, K., & Stoyanovich, J. (2022). Fairness in ranking, part i: Score-based Ranking. *ACM Computing Surveys*, *55*(6), 1-36.

Biega, A. J., Gummadi, K. P., & Weikum, G. (2018, June). Equity of Attention: Amortizing Individual Fairness in Rankings. In *The 41st International ACM SIGIR Conferences on Research & Development in Information Retrieval* (pp. 405-414).

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

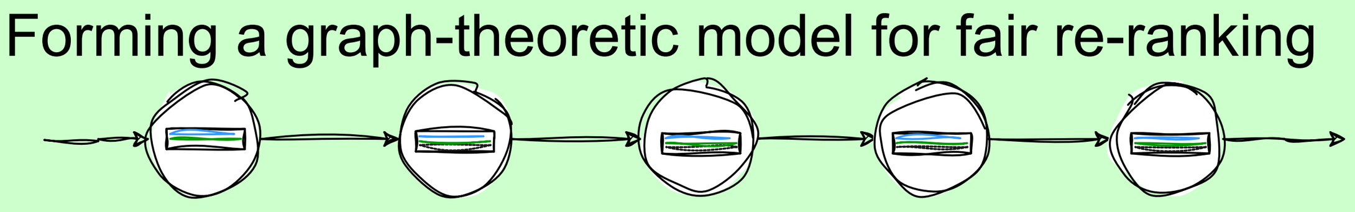
Wang, C., Liu, Y., Wang, M., Zhou, K., Nie, J. Y., & Ma, S. (2015, August). Incorporating Non-sequential Behavior Into Click Models. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 283-292).

Chapelle, O., Metlzer, D., Zhang, Y., & Grinspan, P. (2009, November). Expected Reciprocal Rank For Graded Relevance. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 621-630).

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415-444.

Next Steps

Forming a graph-theoretic model for fair re-ranking



- The CO*IR method I presented at AIMMES is efficient, but produces sub-optimal results.
- In theory you can get optimal results via an exhaustive graph search over possible rankings.
- Ideally some middle-ground would be best, and carefully deciding the right heuristic to guide the graph search towards this middle ground is necessary.

- Applications:
  - A lot of information access systems in the software industry now work through a technology dubbed 'NewSQL'. These systems can combine database queries with features from information retrieval: granular filters plus fulltext search. To get a fair re-ranking technique working in such an environment, you can't assume you'll be able to use a GPU. Current fairness techniques would therefore fail.
  - The current state-of-the-art within fair re-ranking lacks what we refer to in formal linguistics as 'expressive power'.
    - What if an industry practitioner wants their application to change dynamically between notions of 'individual fairness' and 'group fairness'?

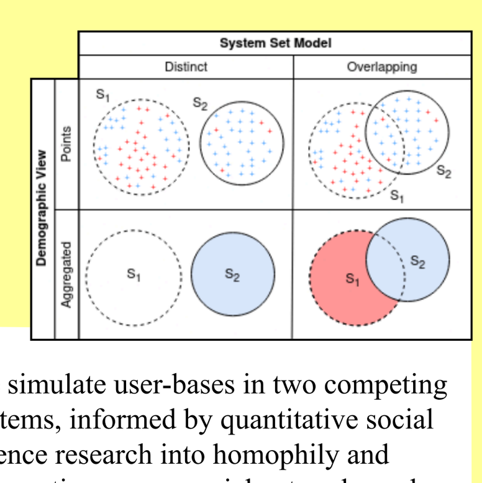
Experiment 2

Using the 'greedy' technique (named 'CO*IR'), you can still get optimally fair results by reducing your search quality threshold.

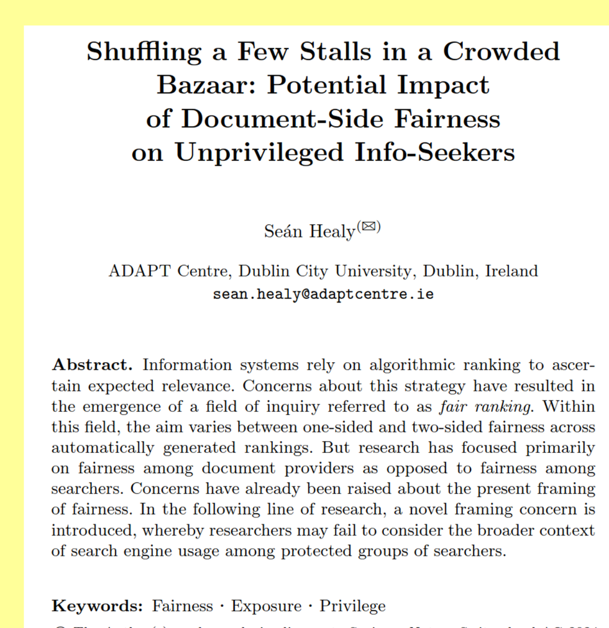
However, this can also lead to fairness issues.

I published an article at ECIR 2024 raising this 'framing' concern.

I also carried out a simulation study using datasets for two academic position opening mailing lists, demonstrating with a simple model for users' social capital (awareness of each system), that quality decrease can be framed a fairness issue with real losses of opportunity.



We simulate user-bases in two competing systems, informed by quantitative social science research into homophily and segregation across social networks and information systems.



Shuffling a Few Stalls in a Crowded Bazaar: Potential Impact of Document-Side Fairness on Unprivileged Info-Seekers

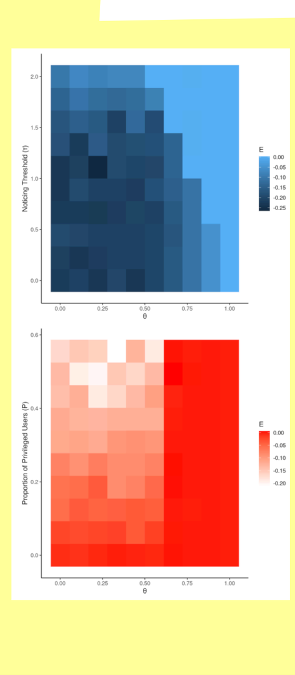
Seán Healy

ADAPT Centre, Dublin City University, Dublin, Ireland
sean.healy@adaptcentre.ie

Abstract. Information systems rely on algorithmic ranking to ascertain expected relevance. Concerns about this strategy have resulted in the emergence of a field of inquiry referred to as *fair ranking*. Within this field, the aim varies between one-sided and two-sided fairness across automatically generated rankings. But research has focused primarily on fairness among document providers as opposed to fairness among searchers. Concerns have already been raised about the present framing of fairness. In the following line of research, a novel framing concern is introduced, whereby researchers may fail to consider the broader context of search engine usage among protected groups of searchers.

Keywords: Fairness · Exposure · Privilege

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
N. Goharian et al. (Eds.): ECIR 2024, LNCS 14612, pp. 338–340, 2024.
https://doi.org/10.1007/978-3-031-56069-9_43



I repeated the simulation with varying values for the "noticing threshold" and privileged system segregation level. When users are less prone to noticing quality decreases, there is less privilege-based disparate impact. When segregation levels are low, disparate impact is also low (however, we know from the social sciences that identity-based segregation, i.e. homophily, is widespread, except in ubiquitous networks).

Conclusion: Varied fairness interventions within web search should be applied at appropriate moments and with appropriate degree.