A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-box Systems: A Case Study with COVID-related Searches

Binh Le, Damiano Spina, Falk Scholer, Hui Chia

binh.le@rmit.edu.au damiano.spina@rmit.edu.au falk.scholer@rmit.edu.au chia.h@unimelb.edu.au









Motivation



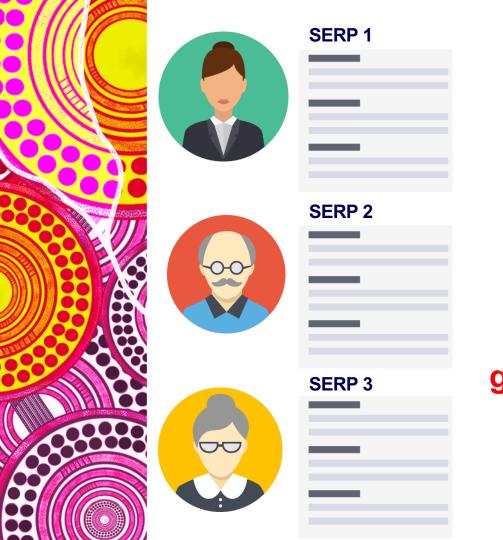
Should I get vaccinated from COVID-19?





Search Engine Results Page







Should I get vaccinated from COVID-19?



RQ1: Do different users get the same search results?





Should I get vaccinated for covid







Should I avoid getting vaccinated for covid





RQ2:

Do results vary between positive and negative query formulations?



Methodology

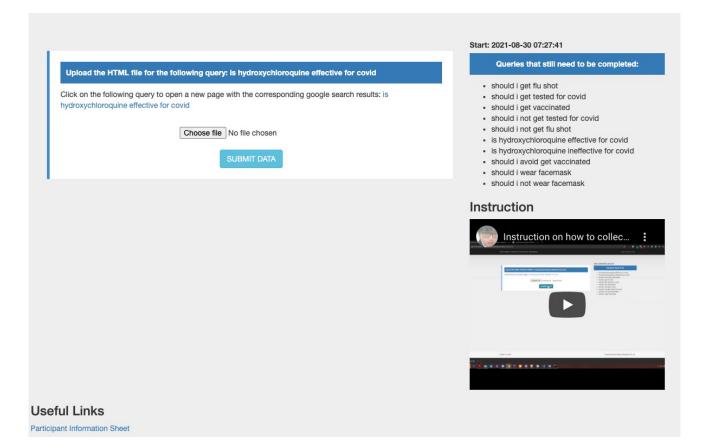
Crowdsourcing Platform to Audit a Web Search Engine

- Requesters: Publish micro-tasks or Human Intelligent Tasks (HITs)
 - Annotation tasks to obtain labels for machine learning development an evaluation)
- Crowd workers: Can choose to carry out micro-tasks at their convenience



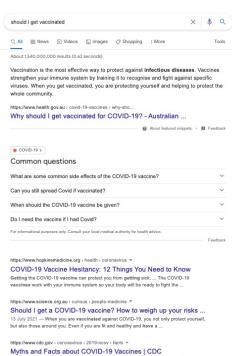


Crowdsourcing Task









Nyths and Pacts about COVID-19 vaccines | CDC Vaccines work by stimulating your immune system to produce antibodies, exactly like it would if you were exposed to the disease. After getting vaccinated, you ...

https://www.mayoclinic.org > in-depth > art-20484859 **

Get the facts about COVID-19 vaccines - Mayo Clinic

Find out about the COVID-19 vaccines, the benefits of a COVID-19 vaccination, ... Should I get the COVID-19 vaccine even if I've already had COVID-19?

https://www.mayoclinic.org > coronavirus-covid-19 **

Why get vaccinated? - Mayo Clinic

Prevent the COVID-19 virus from spreading and replicating, which allows it to mutate and possibly become more resistant to vaccines. More about why you should ...



Participant (Crowd Worker)

- Submit the query to Google using their browser
- 2. Download SERP as HTML file
- Upload HTML file into the web app

Our system will remove all the personal information i.e. avatar, email address



should i avoid get vaccinated X

About 115,000,000 results (0.37 seconds)

https://www.health.gov.au > ... > About immunisation =

Who can be immunised? - Australian Government Department ... 30 Nov 2017 — Should you get vaccinated if you have allergies? ... a minor illness and do not have a fever, you can be safely and effectively vaccinated.

https://www.hopkinsmedicine.org > health > coronavirus =

COVID-19 Vaccine Hesitancy: 12 Things You Need to Know
The vaccines do not contain live coronavirus, and you cannot and will not get COVID-19
from getting vaccinated. After the shots, you might experience a sore ...

https://www.healthline.com / health / vaccinations / im - v

Vaccines: Who Should Avoid Them and Why - Healthline
Vaccinations are safe — However, these vaccines may not be right for everyone. The CDC address that certain people not get specific vaccines, or to wait ...

Overview: Fig. Takeaway - Pos and cons

https://www.sydney.edu.au > news > 2021/02/22 > not-j... =

It's not just vaccine refusers who might not get the COVID-19 ... 22 Feb 2021 — The surprising reasons why people don't get vaccinated ... Leask doesn't do this thinking just for Australians. As chair of a World Health ...

https://www.bbc.com > future > article > 20210720-the-... *

Why some people don't want a Covid-19 vaccine - BBC Future 20 July 2021 — There should be no doubt about it: Covid-19 vaccines are saving lives. ... the fears of most vaccine hesitant people should not be confused ...

https://www.cdc.gov > vaccines > vpd > should-not-vacc *

Who Should not Get Vaccinated I CDC

Has a weakened immune system, or has a parent, brother, or sister with a history of hereditary or congenital immune system problems; Has ever had a condition ...

https://www.cdc.gov > parents > why-vaccinate > vacci... =

Making the Vaccine Decision: Common Concerns | CDC 5 Aug 2019 — The diseases vaccines prevent can be dangerous, or even deadly. Statistically, the chances of your child getting diseases such as measles, ...

https://www.science.org.au > curious > people-medicine =

Should I get a COVID-19 vaccine? How to weigh up your risks ...

13 July 2021 — and are more willing to accept an outcome if it's the result of doing nothing
("if I avoid the vaccine, I know I might catch COVID-19).

https://www.who.int > ... > COVID-19 vaccines *

COVID-19 advice for the public: Getting vaccinated - WHO ...

14 July 2021 — If you experience an immediate severe allergic reaction to a first dose of the COVID-19 vaccine, you should not receive additional doses of ...



Queries

10 queries – 5 pairs

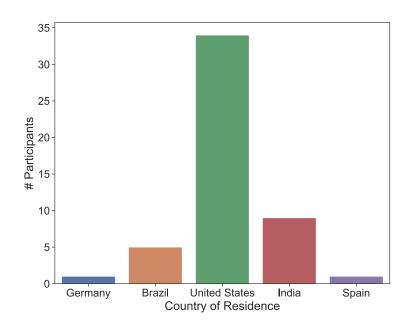
$\frac{1}{2}$	should i get tested for covid should i not get tested for covid	Pair 1
$\frac{3}{4}$	should i get flu shot should i not get flu shot	Pair 2
5 6	should i get vaccinated should i avoid get vaccinated	Pair 3
7 8	should i wear facemask should i not wear facemask	Pair 4
9 10	is hydroxychloroquine effective for covid is hydroxychloroquine ineffective for covid	Pair 5





Queries

- US\$ 1.20 per task
- 10 queries per task
- 1 task per participant
- N=50 participants
- Average time: 23' 30"
- All (50+1) crowdsourcing tasks completed in 2 days. Only 1 incorrect submission.
- Pre-questionnaire: Age, gender, country of origin, country of residence, level of education, ...





Comparing SERPs

Rank-Biased Overlap (RBO)

- Non-conjoint, top-weighted, and incomplete ranked lists
- Parameter "p" (patience): probability of the user to continue inspecting the ranking
 - It determines the strength of the weighting to top ranks
 - E.g., p=0.67: the user inspects only the top-3 search results.

Jaccard Similarity

- Overlap of items in two sets
- It does not consider ranking positions, i.e., it only measures conjointness

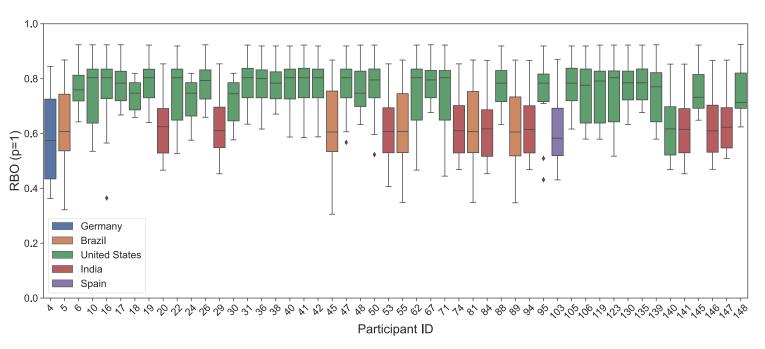
Preliminary Results





Do different users see different SERPs?

Comparing items at domain level

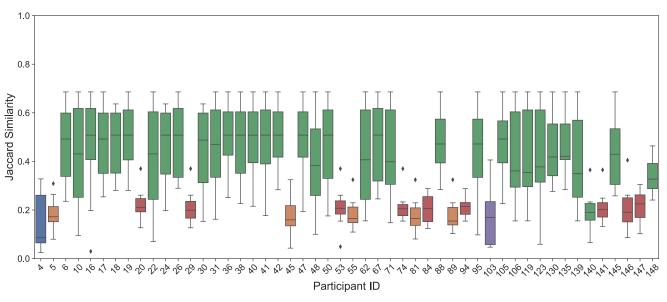






Do different users see different SERPs?

Comparing Items at domain level

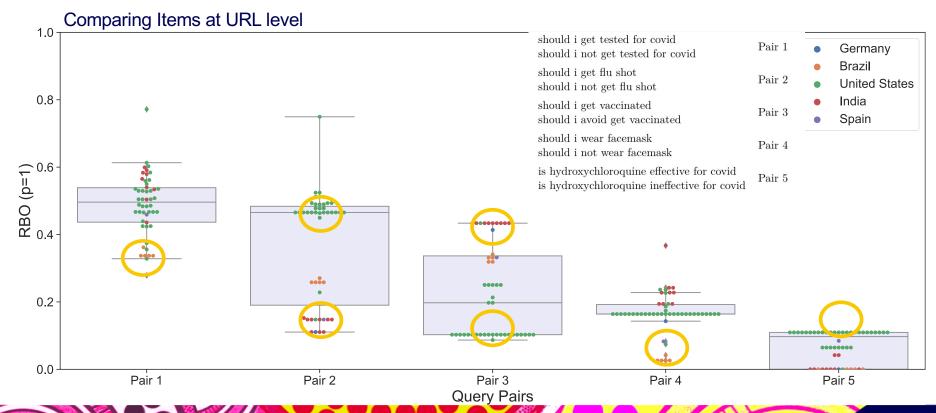




United States India

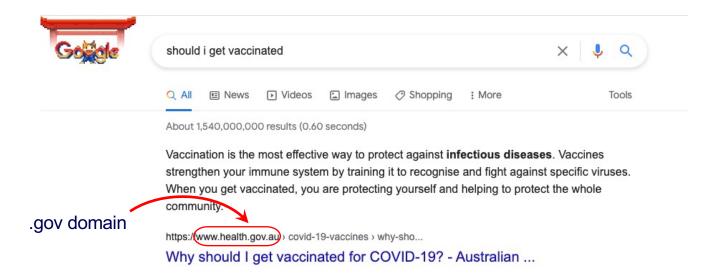
Spain

Does Query Formulation Change SERPs?



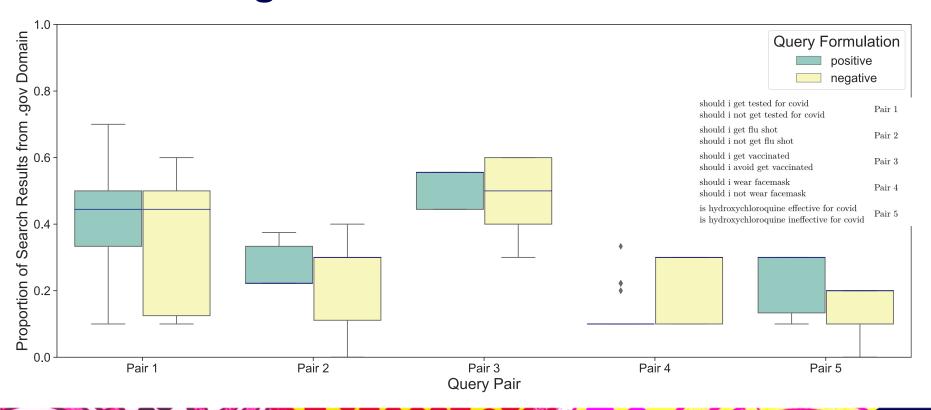


[EXTRA] Quality of Information Sources





Items from .gov Domain in SERPs





Conclusion

- We used a crowdsourcing platform to audit a web search engine (black-box system)
 - Preliminary study
 - We observed differences in the SERPs obtained by crowd workers / participants
- Our study validates the feasibility of this approach
 - Future work: Larger dataset / number of workers
 - A more scalable (but less controlled) methodology than *data donation*





A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-box Systems: A Case Study with COVID-related Searches

github.com/rmit-ir/crowdsourcing-algorithmic-bias



binh.le@rmit.edu.au
damiano.spina@rmit.edu.au
falk.scholer@rmit.edu.au
chia.h@unimelb.edu.au

Acknowledgments:









