Bias in Computer Science Abstracts and the Role of a Team's Diversity

marie-sophie.simon@ru.nl

Marie-Sophie Simon

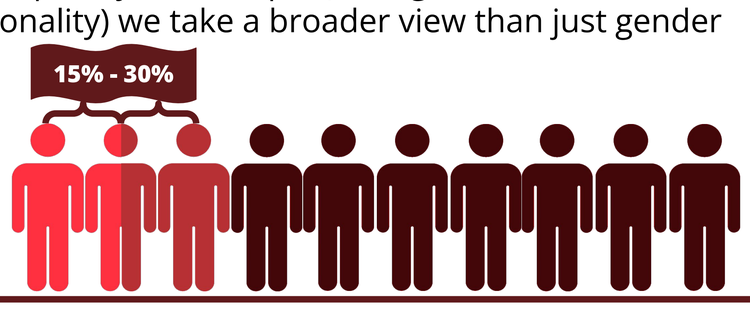
Research:

Does a less diverse research team, lead to more biased words in the produced research publications?

Our research investigates this hypothesis by correlating bias in terms of biased words in published abstracts of computer science with the diversity of a team of co-authors in terms of nationality and gender representation.

Background:

- Only about 15-30% of computer science researchers are female researchers [1]
- Gender imbalance is a self-exacerbating issue: factors reducing women's participation also amplify the effect of their reduced participation [2]
- Due to complexity of concepts (bias, gender, inclusion, diversity, intersectionality) we take a broader view than just gender



Methods:

Data:

- Our final dataset contained 2391 abstracts from 56 computer science conferences of 2017, covering 27 different subfields [3]
- Manually verified gender and nationality of each author per paper

Finding Bias – Dbias:

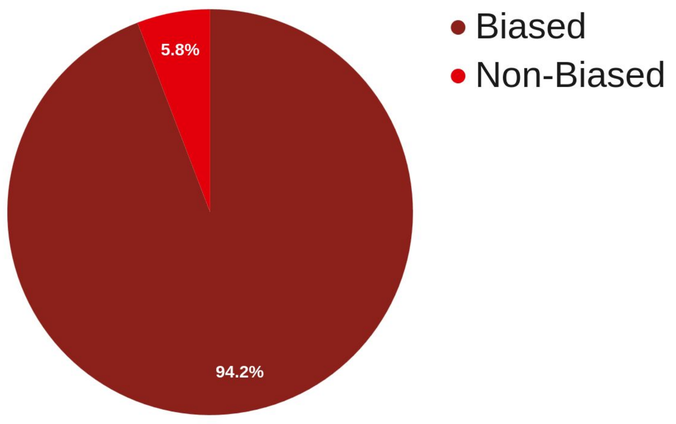
- Dbias detection module for binary classification into “biased” and “non-biased” [4]
- Dbias recognition module to find bias-bearing entities [4]

Correlating Bias and Diversity:

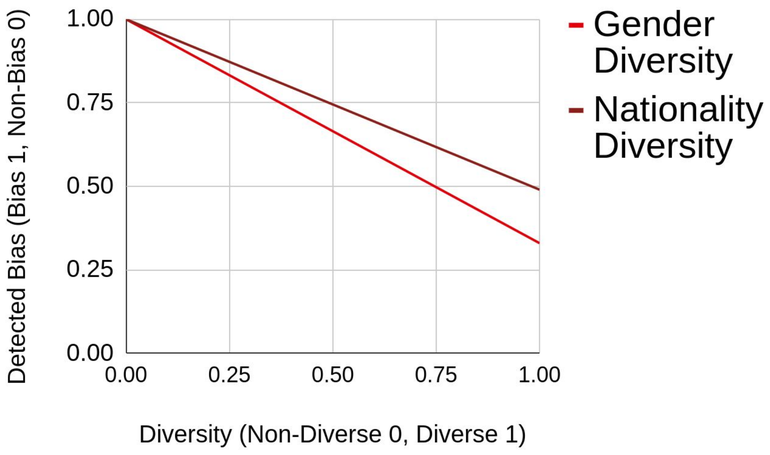
- Balance our dataset using random oversampling
- Logistic regression to analyse possible correlation

Results:

Classification of Bias: 2366 out of the original 2391 texts got classified

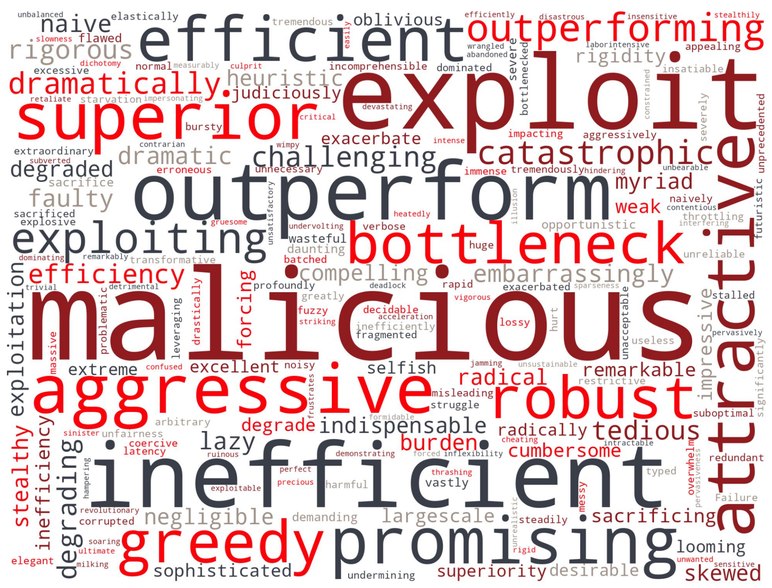


Correlation with Diversity: Abstracts are less likely to be classified as biased by Dbias if the research team had more different nationalities (-0.33, p-value 0.0005) or a more balanced gender distribution (-0.49, 0.000).



Computer Science terminology and bias:

Some of the most frequently used words that were recognized to contain bias, are typical computer science terminology



Discussion – Attribution Theory:

- “Fundamental attribution error” people tend to attribute other people's behaviour to internal and personal characteristics not taking into account external circumstances [5]
- Computer science terms such as “maliciousness” and “exploits” attribute qualities to another person (e.g. someone hacking into software may not be malicious but there may be external reason that drove them to do this)
- Example illustrates that dismissing the recognized biased words as necessary computer science terminology risks overlooking inherently biased words within our field
- Such biased words in an abstract convey a certain view of computer science, therefore attracting only certain people

Example of biased sentences recognized by Dbias:

News Media: “Democrats Planning to **Exploit** the Economic Crisis to Hurt Trump” [6]

Recognized bias-bearing entities

Computer Science article: “We show the existence of **exploitable** side channels” [7]

Limitations:

- Variable of gender is measured binary, instead of a fluid spectrum [8]
- Only two measures of diversity, no intersectionality and no comparison between different subfields
- Bias is conceptualised broadly in Dbias using words, not context
- Dbias was fine-tuned on news articles

References:

[1] Frachtenberg, E., & Kaner, R. D. (2022). Underrepresentation of women in computer systems research. PLOS ONE, 17(4), e0266439. https://doi.org/10.1371/journal.pone.0266439
[2] Richter, A., Yamamoto, J., & Frachtenberg, E. (2023). Why Are There So Few Women in Computer Systems Research? Computer, 56(2), 101–105. https://doi.org/10.1109/MC.2022.3219633
[3] Frachtenberg, E. (2021). Systems conferences analysis dataset. GitHub - Systems Conferences Analysis Dataset. https://github.com/eitanf/sysconf/tree/master
[4] Raza, S., Reji, D. J., & Ding, C. (2022). Dbias: Detecting biases and ensuring fairness in news articles. International Journal of Data Science and Analytics. https://doi.org/10.1007/s41060-022-00359-4
[5] Baumgartner, B. J., Bauer, L. M., & Bui, K. V. T. (2012). Reactions to homelessness: Social, cultural, and psychological sources of discrimination. Psi Chi Journal of Psychological Research, 17(1), 26–34.
[6] Kerik, B. (2020, March 23). Giddy Democrats Planning to Exploit the Economic Crisis to Hurt Trump › American Greatness. American Greatness. https://amgreatness.com/2020/03/22/giddy-democrats-planning-to-exploit-the-economic-crisis-to-hurt-trump/
[7] Wang, L., Grubbs, P., Lu, J., Bindschaedler, V., Cash, D., & Ristenpart, T. (2017). Side-Channel Attacks on Shared Search Indexes. 2017 IEEE Symposium on Security and Privacy (SP), 673–692. https://doi.org/10.1109/SP.2017.50
[8] Stanczak, K., & Augenstein, I. (2021). A Survey on Gender Bias in Natural Language Processing (arXiv:2112.14168). arXiv. http://arxiv.org/abs/2112.14168

Radboud University