Fairness and Diversity in Information Access Systems

Lorenzo Porcaro[1], Carlos Castillo[2], Emilia Gómez[3] and João Vinagre[3,4]

[1]*Joint Research Centre, European Commission, Italy*
[2]*Web Science and Social Computing Group, UPF, & ICREA, Spain*
[3]*Joint Research Centre, European Commission, Spain*
[4]*University of Porto, Portugal*

1. Introduction

Among[1] the seven key requirements to achieve trustworthy AI proposed by the High-Level Expert Group on Artificial Intelligence (AI-HLEG) established by the European Commission (EC), the fifth requirement ("Diversity, non-discrimination and fairness") declares: "In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. [...] This requirement is closely linked with the principle of fairness"[Chapter 2, Section 1.5, AI-HLEG, 2019]. Hereafter, we try to shed light on how closely these two distinct concepts, diversity and fairness, may be treated by focusing on Information Access Systems literature. These concepts should not be used interchangeably because they do represent two different values, but what we argue is that they also cannot be considered totally unrelated or divergent. Having diversity does not imply fairness, but fostering diversity can effectively lead to fair outcomes, an intuition behind several methods proposed to mitigate the disparate impact of information access systems, i.e. recommender systems and search engines [2, 3, 4, 5].

2. Links between Fairness and Diversity

The first link can be found between the concepts of *group fairness* and *egalitarian diversity* [6, 7]. Indeed, the former, often referred to as demographic or statistical parity, is achieved when different groups, e.g., with regard to certain demographics, receive similar treatments. To maximise egalitarian diversity, hence having a population uniformly distributed among different groups [8], is identical to enforcing group fairness, wherein every group has equal representation i.e. similar treatment. This idea is behind the use of diversity constraints while intervening in the outcome of an automated decision-making system [9]. Moreover, group fairness relates to the concept of *coverage-based diversity*, an aggregated diversity metric often

[1]This work has been adapted from Lorenzo Porcaro PhD dissertation at the Universitat Pompeu Fabra, Spain.

used in Recommender Systems literature. Indeed, such metric is maximised when different groups of items are represented in the most heterogeneous way.

Second, both fairness and diversity relate to the treatment of, and consequently the impact on, protected/disadvantaged/minority groups (or classes). The definition of protected class is usually dependent upon laws and policies which may vary between countries, aiming at preventing any form of discrimination towards such classes. For instance, the *EU Charter of Fundamental Rights* states that: "Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited" [Article 21, EC, 2012].

As argued by Castillo [11], ensuring fairness can be seen as "emphasising not the presence of various groups but ensuring that those in protected groups are effectively included". Under this lens, it is evident that the construction of a group diverse in egalitarian terms may not result in a fair representation if disadvantaged classes are not effectively included. However, if we consider the exposure diversity with *adversarial perspective* as defined by Helberger [12], it explicitly aims at "promoting exposure to critical voices and disadvantaged views that otherwise might be silenced in the public debate". If defined as above, we notice that both fairness and diversity stress the importance of targeting a representation that is not only equal in terms of distribution but also that may give exposure to historically disadvantaged groups. We can further relate these concepts with the idea of *normative diversity* [8]. Indeed, if we imagine a scenario where the non-diverse norm coincides with the privileged group — for instance, the STEM community where the old-white-male represents the stereotype of the scientist — increasing the diversity in a normative sense would result in a wider inclusion of marginalised voices, which is what the exposure diversity under an adversarial perspective would target.

3. Differences and Limitations

So far we have discussed some intersections between diversity and fairness concepts, but in order to better clarify their nature it is useful to focus also on the differences between them. Early quantitative definitions of both values have been proposed several decades ago, but in their rationale we note a substantial difference. Indeed, whilst since the beginning fairness metrics have been proposed to tackle societal issues [13], most of the diversity indexes still widely used have been proposed in disparate fields, e.g., Simpson's Index in Ecology [14], and they have been originally formulated to measure diversity intended as heterogeneity, variety or entropy, e.g., Shannon's Index [15]. Even if this does not mine their use in measuring diversity, it is also true that their application needs to be contextualised for supporting the validity of the inferred results. Similarly, a lack of a value-oriented approach can be found in the design of the diversification techniques [16, 17]. Indeed, looking at the early proposals of the Information Retrieval and Recommender Systems communities, the main goal for diversifying is to tackle the problem of ambiguity of a query or the redundancy of the results, and also to deal with uncertainty. Great advancements have been made in this direction [18], but this utility-oriented definition of diversity has partly created ambiguity over the concept of diversity itself, at least in the communities where such approaches have been applied.

4. Conclusion

Whilst the aforementioned points are just a few among the several aspects that link diversity and fairness, we conclude by stressing their relevance in recent policies proposed in the European context. The Digital Service Act (DSA)[19] mandates that digital services powered by technologies such as recommender systems and search engines should be monitored to guarantee the avoidance of unfair or arbitrary outcomes. Under a different lens, the Artificial Intelligence Act (AI Act) proposal [20], also refers to the need for bias monitoring as part of the mandatory requirements for high-risk AI systems. Moreover, in terms of diversity the AI Act explicitly states that providers of AI systems should be encouraged to create code of conducts covering aspects such as accessibility, stakeholders participation and ensuring diversity of development teams. These two goals considered above, i.e. system-centric (ensuring bias and fairness in algorithmic systems) and a people-centric view (ensuring diversity of persons involved in the AI design process), are strongly related. Only fostering the diversity of development teams, and therefore embedding different perspectives, could lead to a future where Information Access Systems act in a trustworthy and fair way.

Acknowledgments

References

[1] AI-HLEG, High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, 2019. [Accessed: 2023-02-14].

[2] L. E. Celis, A. Deshpande, T. Kathuria, N. K. Vishnoi, How to be fair and diverse?, in: Fairness, Accountability and Transparency in Machine Learning (FAT/ML) 2016, 2016. URL: http://arxiv.org/abs/1610.07183. arXiv:1610.07183.

[3] P.-R. Lhérisson, F. Muhlenbach, P. Maret, Fair recommendations through diversity promotion, in: International Conference on Advanced Data Mining and Applications (ADMA 2017), volume 4093, 2017, pp. 89–103. URL: http://link.springer.com/10.1007/11811305. doi:10.1007/11811305.

[4] W. Liu, R. Burke, Personalizing fairness-aware re-ranking, in: FATREC 2018 Workshop: Responsible Recommendation, 2018. URL: http://arxiv.org/abs/1809.02921. arXiv:1809.02921.

[5] G. McDonald, C. Macdonald, I. Ounis, Search results diversification for effective fair ranking in academic search, Information Retrieval Journal 25 (2022) 1–26. URL: https://doi.org/10.1007/s10791-021-09399-z. doi:10.1007/s10791-021-09399-z.

[6] M. Drosou, H. Jagadish, E. Pitoura, J. Stoyanovich, Diversity in big data: A review, Big Data 5 (2017) 73–84. doi:10.1089/big.2016.0054.

[7] M. Mitchell, D. Baker, N. Moorosi, E. Denton, B. Hutchinson, A. Hanna, T. Gebru, J. Morgenstern, Diversity and inclusion metrics in subset selection, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, AIES '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 117–123. URL: https://doi.org/10.1145/3375627.3375832. doi:10.1145/3375627.3375832.

[8] D. Steel, S. Fazelpour, K. Gillette, B. Crewe, M. Burgess, Multiple diversity concepts and their ethical-epistemic implications, European Journal for Philosophy of Science 8 (2018) 761–780. doi:10.1007/s13194-018-0209-5.

[9] M. Zehlike, K. Yang, J. Stoyanovich, Fairness in ranking, part I: Score-based ranking, ACM Computing Surveys (2022). URL: https://doi.org/10.1145/3533379. doi:10.1145/3533379.

[10] EC, European Commission, The charter of fundamental rights, Official Journal of the European Communities (2012). URL: https://eur-lex.europa.eu/eli/treaty/char_2012/oj.

[11] C. Castillo, Fairness and transparency in ranking, ACM SIGIR Forum 52 (2018) 64–71. doi:10.1145/3308774.3308783.

[12] N. Helberger, K. Karppinen, L. D'Acunto, Exposure diversity as a design principle for recommender systems, Information, Communication and Society 21 (2018) 191–207. URL: https://doi.org/10.1080/1369118X.2016.1271900. doi:10.1080/1369118X.2016.1271900.

[13] B. Hutchinson, M. Mitchell, 50 years of test (un)fairness: Lessons for machine learning, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 49–58. URL: https://doi.org/10.1145/3287560.3287600. doi:10.1145/3287560.3287600.

[14] J. Simpson, Measurements of diversity, Nature 163 (1949) 688.

[15] C. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal Vol.27 (1948) 379–423.

[16] J. Carbonell, J. Goldstein, Use of MMR, diversity-based reranking for reordering documents and producing summaries, in: SIGIR Forum (ACM Special Interest Group on Information Retrieval), 1998, pp. 335–336. doi:10.1145/3130348.3130369.

[17] B. Smyth, P. McClave, Similarity vs. Diversity, in: Proceedings of the International Conference on Case-Based Reasoning (ICCBR) 2001, 2001, pp. 347–361. doi:10.1007/3-540-44593-5_25.

[18] P. Castells, N. J. Hurley, S. Vargas, Novelty and diversity in recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, third ed., Springer, 2022, pp. 603–647.

[19] EC, European Commission, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1666857835014, 2022. [Accessed: 2023-02-14].

[20] EC, European Commission, Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).COM(2021) 206, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, 2021. [Accessed: 2023-02-14].