A Framework for Fairness Evaluation in Academic Recommender Systems: Evidence from Scholarly Data

Gabriela Suntaxi*, César Sarango*, Marco Aguirre* and Lorena Recalde*
* Department of Informatics and Computer Sciences,
Escuela Politécnica Nacional,
Ladrón de Guevara E11-253, Quito, Ecuador
Email: {gabriela.suntaxi, cesar.sarango01, marco.aguirre, lorena.recalde}@epn.edu.ec

***Abstract*—Academic recommender systems play a central role in shaping how researchers discover knowledge, making fairness a critical concern. However, existing approaches remain fragmented and often transfer methods from general recommender systems without accounting for the structural characteristics of scholarly data. This paper proposes a domain-aware framework for fairness evaluation in academic recommender systems. Our proposed framework consists of four dimensions: bias sources, auditing and tooling, mitigation strategies, and evaluation and visualization. The framework is empirically validated using a large-scale dataset of Ecuadorian scholarly production. The results reveal measurable fairness risks related to gender imbalance, institutional concentration, and long-tail productivity distributions. They also show that while existing fairness toolkits can detect disparities, their effectiveness depends on domain-aware interpretation and alignment with ranking-based processes. The proposed framework provides a practical foundation for systematic fairness assessment in scholarly recommendation systems.**

***Index Terms*—Academic Recommender Systems, Fairness, Bias Auditing, Scholarly Data, Responsible AI**

I. Introduction

Academic recommender systems increasingly mediate how researchers discover papers, identify collaborators, explore venues, and navigate emerging topics [1], [2]. In doing so, they influence not only information access, but also patterns of visibility, recognition, and knowledge circulation within science. This role becomes especially important in digital scholarly infrastructures, where recommendation outputs interact with search interfaces, citation practices, and institutional prestige signals [3], [4].

This growing influence makes fairness a central concern in academic recommendation. In contrast to many commercial settings, where unfairness mainly affects exposure and consumption, in scholarly environments, it may also affect discoverability, citation opportunity, collaboration prospects, and the accumulation of academic reputation over time. Previous work has documented the effects of popularity bias, exposure concentration, feedback loops, and unequal treatment of groups or providers in recommender systems [5]–[7]. In academic contexts, these mechanisms may further benefit highly cited authors, prestigious institutions, and already visible topics, while underrepresented scholars, regions, and research areas remain comparatively invisible.

At the same time, fairness in academic recommender systems remains methodologically fragmented. Some works address scholarly recommendation directly, including paper recommendation and related academic recommendation tasks [8]–[10]. However, many of the concepts, metrics, and mitigation strategies used to reason about fairness in this domain come from broader fairness-aware recommender-systems literature [11]–[13]. Although this transfer is useful, it is not seamless. Academic recommendation is often ranking-based, scholarly entities are embedded in citation and affiliation networks, and relevance signals involve prestige and cumulative advantage. Consequently, methods that are effective in generic recommendation settings may only partially capture fairness risks in scholarly ecosystems.

To address this gap, this paper proposes a domain-aware framework for fairness evaluation in academic recommender systems. The framework organizes fairness analysis through four complementary dimensions: bias sources, auditing and tooling, mitigation strategies, and evaluation and visualization. These dimensions provide a structured approach for analyzing fairness risks throughout the recommendation lifecycle.

To demonstrate its applicability, the framework is empirically validated using a large-scale dataset of Ecuadorian scholarly production, comprising 69,392 articles, 65,547 authors, and rich relational structures linking authors, institutions, and topics [3]. The analysis reveals measurable fairness risks associated with gender imbalance, institutional concentration, and long-tail productivity and citation distributions, while also illustrating how existing fairness toolkits can be systematically applied and interpreted within academic data environments.

The main contribution of this paper is the proposal of a domain-aware framework for fairness evaluation in academic recommender systems. Unlike prior survey-oriented studies, which mainly review fairness concepts, bias types, or mitigation strategies across recommender systems, this work integrates and operationalizes these elements specifically for scholarly recommendation environments. In this sense, the reviewed literature provides the conceptual foundation, whereas the four-dimensional framework proposed here constitutes the paper's original contribution. The empirical study on Ecuadorian scholarly data is then used to validate the applicability of the framework in a real academic data setting.

The remainder of the paper is organized as follows. Section II describes the review design. Section III presents the analytical framework. Section IV introduces the motivating case based on Ecuadorian scholarly data. Section V discusses implications for fair academic recommendation, and Section VI concludes the paper.

II. Review Design

This paper adopts a focused review design to support the construction of the proposed fairness-evaluation framework. This choice reflects the current state of the field: while fairness in recommender systems has been studied extensively, work addressing scholarly recommendation directly remains comparatively limited. At the same time, many relevant concepts, metrics, and mitigation strategies have been developed in adjacent areas such as fairness-aware recommender systems, ranking fairness, and trustworthy recommendations. Under these conditions, a focused review is better suited than an exhaustive systematic review for building a precise and domain-aware synthesis [11]–[14].

The review distinguishes between two complementary categories of evidence. The first includes studies centered on scholarly recommendation tasks, such as paper recommendation, collaborator recommendation, venue recommendation, and analyses of bias in scholarly platforms [8]–[10]. These studies provide the most direct basis for reasoning about fairness in academic recommender systems. The second includes broader methodological work on fairness-aware recommender systems and related ranking settings [11]–[13]. These contributions provide mature conceptual and technical foundations, but their applicability to scholarly recommendation must be assessed carefully due to the distinctive characteristics of academic data, citation networks, affiliation structures, and visibility dynamics.

The review is comparative and interpretive rather than frequency-based. It examines how previous studies contribute to the four analytical dimensions and where gaps remain for academic recommender systems.

This synthesis provides the conceptual foundation for the proposed framework. While the reviewed studies contribute existing fairness concepts, metrics, and mitigation approaches, the integration of these elements into the four-dimensional framework proposed in this paper constitutes the original contribution of this work.

III. Analytical Framework

To synthesize the literature in a way that is both domain-aware and methodologically coherent, this paper organizes fairness challenges in academic recommender systems through bias sources, auditing and tooling, mitigation strategies, and evaluation and visualization. These are not intended as isolated stages of a linear pipeline. Rather, they capture complementary dimensions through which fairness is defined, inspected, operationalized, and interpreted in recommendation settings. Taken together, they provide a compact framework for distinguishing what is already evidenced in scholarly recommendation contexts from what is transferred from broader fairness-aware recommender-system research [11]–[13].

Unlike prior survey-oriented studies, the proposed framework does not merely summarize fairness-aware recommendation literature but organizes and operationalizes it specifically for fairness evaluation in academic recommender systems.

Figure 1 illustrates the proposed framework and its four interrelated dimensions for fairness evaluation in academic recommender systems.

The framework represents the authors' original analytical contribution. Unlike previous studies that focus on bias characterization, auditing, mitigation, or evaluation separately, it integrates these perspectives into a unified structure tailored to academic recommender systems. While the framework supports systematic fairness assessment in scholarly environments, its validation remains exploratory and depends on partially observed protected attributes.

The first lens, *bias sources*, concerns the mechanisms through which unfairness may emerge before or during recommendation. In academic recommender systems, these mechanisms include not only well-known phenomena such as popularity bias and exposure concentration [15], but also domain-specific asymmetries linked to citation accumulation, institutional prestige, geographic concentration, topical imbalance, and incomplete metadata. In this context, bias is not limited to demographic unfairness alone; it also involves structural distortions in how scholarly visibility is distributed across authors, institutions, venues, and research areas. This lens identifies whether fairness issues originate from data, interaction dynamics, or ranking behavior [9], [10], [13], [16].

The second lens, *auditing and tooling*, focuses on the means through which fairness-related risks are inspected and characterized. The literature shows that much of the current auditing ecosystem in recommender systems relies on general-purpose fairness toolkits, interpretability libraries, and diagnostic procedures that were not originally designed for scholarly data [16]–[18]. These tools, including approaches such as AIF360, Fairlearn, and xQUAD commonly discussed in recent review articles, are valuable because they offer measurable criteria, reusable workflows, and practical entry points for bias inspection. However, their direct use in academic recommender systems is often limited by three factors: recommendation outcomes are typically ranking-based rather than purely classificatory, scholarly entities are relational and embedded in citation or affiliation networks, and many fairness-relevant attributes are only partially observed. Under these conditions, auditing becomes not only a matter of running metrics, but also of deciding which entities, groups, and exposure mechanisms should be treated as relevant units of analysis [13], [14], [19].

The third lens, *mitigation strategies*, addresses how fairness interventions are introduced into recommendation processes [20]. In the broader literature, mitigation is commonly dis-

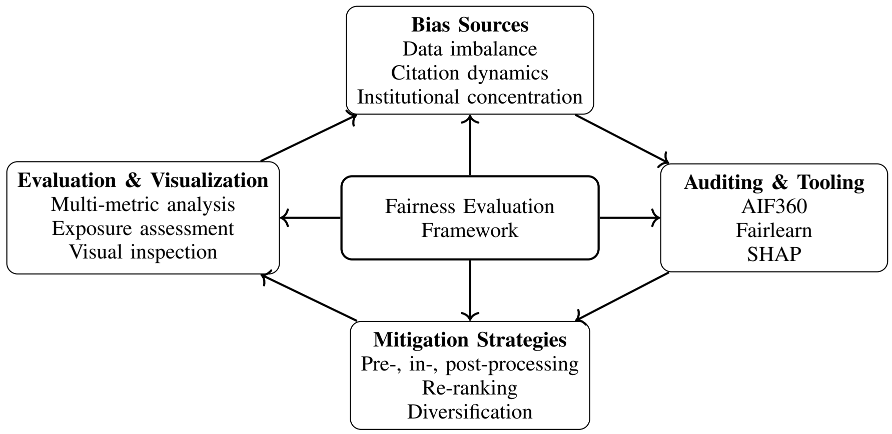

Fig. 1. Proposed framework for fairness evaluation in academic recommender systems.

cussed in terms of preprocessing, in-processing, and post-processing approaches [21]. This categorization remains useful in academic recommender systems, but its practical meaning changes in important ways. Data-level interventions may be constrained by the limited control that scholarly platforms have over external bibliographic sources. Model-level interventions may be difficult to deploy in real systems where recommendation pipelines are opaque, inherited, or optimized for relevance under institutional constraints. Output-level interventions, such as re-ranking or diversification, may therefore become more feasible in practice, even if they do not fully address the upstream origins of unfairness [14]. This lens is important because it highlights that the question is not only whether a mitigation method exists, but also whether it is compatible with the technical and organizational realities of academic recommendation platforms [11]–[13].

The fourth lens, *evaluation and visualization*, concerns how fairness outcomes are measured, compared, and interpreted. This dimension is critical, as fairness cannot be reduced to a single metric. Group-based measures capture disparities across categories but may overlook ranking position and exposure dynamics [5], [15]. Conversely, exposure-aware or diversity-oriented metrics may better reflect scholarly visibility dynamics, yet still fail to explain whether observed differences are normatively acceptable in a given academic context. For this reason, evaluation in academic recommender systems should be understood as inherently multi-dimensional. Visualization and human-centered analytical support also play an important role [22], since fairness assessment often requires the joint interpretation of ranking patterns, group distributions, and domain context rather than the isolated reading of a numerical score [12]–[14].

These four lenses are analytically distinct, but closely interdependent. Bias sources shape what auditing should inspect; auditing influences which mitigation strategies appear justified; mitigation outcomes depend on the evaluation criteria that are adopted; and visualization supports the interpretation of all three [23]. This interdependence is especially relevant in academic recommender systems, where fairness is tied to the governance of scholarly visibility rather than to a single notion of prediction quality. A method that improves one metric may leave exposure concentration unchanged; a mitigation strategy that increases diversity may alter perceived relevance; and an auditing workflow that is adequate for general recommendations may overlook the institutional and epistemic asymmetries embedded in scholarly data. The framework proposed here is intended precisely to make those tensions explicit.

For practical application, the framework relates its dimensions to methods commonly discussed in recent review articles. Fairness tools such as AIF360, Fairlearn, and Aequitas support group-disparity auditing; SHAP supports interpretability; and xQUAD supports re-ranking and exposure balancing. However, these methods require domain-aware adaptation to address ranking outcomes, citation dynamics, institutional concentration, and incomplete protected attributes in academic recommender systems.

IV. Empirical evaluation of the framework

To validate the proposed framework and demonstrate its practical applicability, we apply it to a real-world dataset of Ecuadorian scholarly production. Unlike the literature synthesis presented in previous sections, this empirical analysis constitutes the experimental component of the paper and operationalizes the four dimensions of the proposed framework in a scholarly data environment.

The data were obtained from Scopus through a query targeting Ecuadorian-affiliated production and processed into entity and relation tables [24]. The resulting collection includes 69,392 articles, 65,547 enriched author records, 8,143 affiliations, 122,124 topics, and relational links among authors, articles, affiliations, subject areas, and topics. This structure reflects the multi-entity environment in which scholarly recommendations typically operate.

The exploratory analysis reveals three fairness-relevant patterns. First, the author-level gender attribute is both imbalanced and incomplete. Among the 48,252 records that underwent gender labeling, 52.8% were labeled as male, 29.6% as female, and 17.6% as unknown. Second, the data show

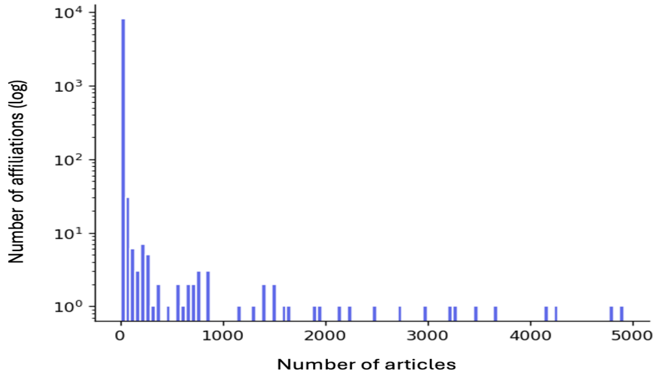

Fig. 2. Institutional concentration of scientific production across affiliations in the Ecuadorian bibliometric dataset

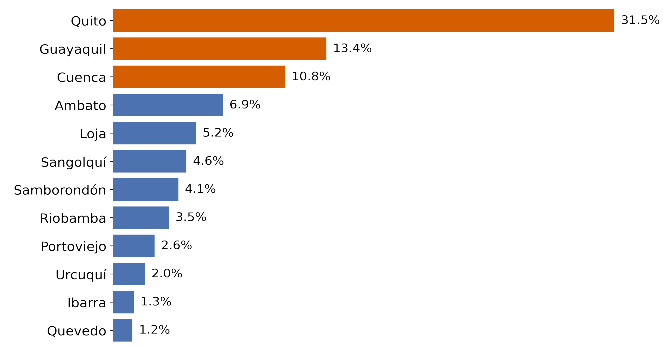

Fig. 3. Distribution of the published articles per city

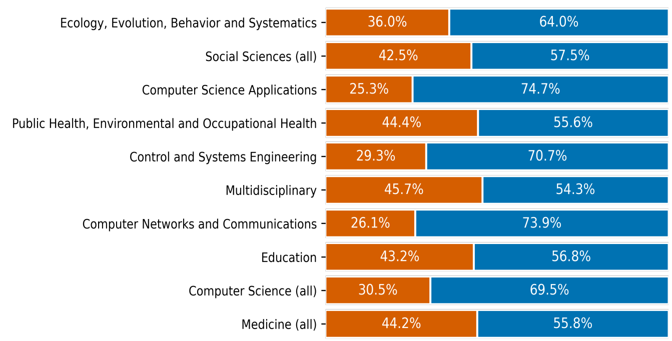

Fig. 4. Gender distribution across the ten most represented thematic areas (left: female, right: male).

strong institutional concentration given that only 61 of the 8,143 affiliations (0.7%) account for 78.1% of the scientific production in the dataset (Figure 2). Third, productivity and impact signals are highly skewed: the h-index distribution is strongly right-skewed, with a median of 1 and extreme values reaching 188, while citations follow a long-tail pattern with a mean of 14.2, a median of 3, and nearly one-third of articles receiving no citations. Together, these characteristics suggest that recommendation models relying on scholarly impact or productivity proxies may easily reinforce pre-existing visibility asymmetries.

These patterns remain relevant when viewed through protected-variable candidates derived from the data. The exploratory process identified four particularly important dimensions for fairness analysis: gender, geographic location, institutional size, and subject area. The top three cities account for 55.7% of the articles (Figure 3), while author–subject-area relations reveal 301 thematic areas with uneven concentration. Moreover, intersectional inspection suggests that gender disparities are not uniform across contexts. For instance, female representation is relatively higher in Medicine and Social Sciences, but substantially lower in areas such as Computer Science Applications and Computer Networks (Figure 4), while differences in h-index also vary by institutional size and city. This reinforces the idea that fairness assessment in academic recommender systems should not rely exclusively on single-attribute comparisons, since disparities may depend on the interaction between disciplinary, geographic, and organizational factors.

The data exploration also provides a practical view of how generic fairness tooling behaves in a scholarly-data setting. A comparative assessment selected AIF360 and Fairlearn as the most suitable core options, with Aequitas retained as a complementary reporting tool (Table I). In a proof-of-concept focused on gender and productivity, AIF360 and Fairlearn produced consistent results for disparate impact, with expected sign differences in statistical parity due to privilege-group conventions. In a second proof-of-concept based on a LightGBM model, SHAP analysis showed that *publication_count* dominated the model contribution (76.6%), while institutional-size analysis revealed substantial predictive disparity, with a disparate impact of 0.6224 and markedly lower rates of high predictions for small institutions. These results do not constitute a full fairness audit, but they do show that generic fairness and interpretability libraries can reveal meaningful signals in scholarly data while still requiring adaptation to academic recommendation contexts.

At the same time, this case highlights important limitations. The gender attribute was not obtained directly from Scopus and was only partially labeled; the analyzed files correspond to a processed layer rather than direct production-database extraction; and the use of Scopus as a single source likely underrepresents Spanish-language outputs as well as some social-science and humanities production. This shows that fairness analysis in academic recommender systems is shaped not only by modeling choices, but also by the quality, provenance, and coverage of scholarly data.

V. Discussion

The results of the empirical analysis provide validation for the proposed framework for fairness evaluation in academic recommender systems. By applying the four analytical dimensions: i) bias sources, ii) auditing and tooling, iii) mitigation strategies, and iv) evaluation and visualization to a real-world scholarly dataset, the study demonstrates that fairness in academic recommendation is inherently multi-dimensional

TABLE I
SUMMARY OF FAIRNESS-TOOLING RESULTS IN THE ECUADORIAN SCHOLARLY-DATA VIGNETTE.

Aspect	Evidence	Main insight
Toolkit selection	AIF360 and Fairlearn selected as core tools; Aequitas retained as complementary; Fairkit-learn discarded. Criteria: Python 3.11 compatibility, tabular-data support, metrics, mitigation, maintenance, scikit-learn integration, visualization.	Fairness analysis in scholarly data requires tools that are not only theoretically relevant, but also compatible with the predictive and analytical stack used in practice.
Gender/productivity PoC	AIF360 and Fairlearn yielded consistent Disparate Impact values (0.9863); SPD differed only in sign (-0.0102 vs. 0.0102) due to privilege-group conventions.	Generic fairness toolkits can yield consistent signals in scholarly data, although metric interpretation still depends on implementation details.
LightGBM + SHAP PoC	*publication_count* explained 76.6% of the model contribution; *distinct_authors*, *year*, and *affiliation_encoded* contributed 11.0%, 9.9%, and 2.5%, respectively.	The model is strongly driven by prior productivity, which may amplify popularity bias and cumulative advantage.
Institution-size disparity	SPD = -0.2992; DI = 0.6224; high-prediction rates: 79.2% (large), 82.5% (medium), 45.4% (small).	Institutional size is associated with substantial predictive disparity, suggesting that smaller institutions may be systematically disadvantaged.
Final fairness stack	AIF360, Fairlearn, SHAP, Aequitas, xQUAD, and a custom visualization module.	A workable fairness pipeline for academic recommendation benefits from combining auditing, explainability, re-ranking, and visualization.

and cannot be adequately captured through isolated metrics or generic methodologies.

First, the analysis of bias sources confirms that fairness challenges in academic recommender systems are deeply rooted in the structural properties of scholarly data. The observed gender imbalance, institutional concentration, and long-tail distributions in productivity and citation-related signals indicate that bias is not only an algorithmic artifact, but also a consequence of data representation and knowledge production dynamics.

Second, the auditing and tooling dimension shows that existing fairness libraries, such as AIF360 and Fairlearn, can be effectively used to detect disparities in scholarly data when properly contextualized. The empirical results demonstrate that these tools produce consistent signals, such as disparate impact and statistical parity differences, even in complex academic datasets. However, the analysis also highlights that their interpretation is not straightforward because results depend on assumptions about protected groups, data completeness, and the nature of prediction tasks.

Third, the evaluation of mitigation strategies reveals important practical constraints. While standard approaches such as preprocessing, in-processing, and post-processing remain theoretically applicable, their feasibility in academic environments is uneven. Data-level interventions are limited by the dependence on external bibliographic sources, and model-level interventions may be difficult to deploy in operational systems where recommendation pipelines are fixed or opaque. As a result, output-level strategies, such as re-ranking and diversification, emerge as more viable options.

Fourth, the evaluation and visualization dimension demonstrates that fairness assessment in academic recommender systems must be multi-dimensional. The empirical findings show that group fairness metrics alone are insufficient to capture disparities in visibility and exposure. For instance, institutional disparities and productivity-driven biases require complementary analysis that considers ranking behavior and distributional effects. Visualization and exploratory analysis play a critical role in this process by enabling the interpretation of fairness outcomes in relation to domain context, rather than relying solely on numerical indicators.

These results show that the four dimensions are interdependent components of a unified fairness evaluation process. Bias sources determine what should be audited; auditing informs which mitigation strategies are appropriate; mitigation outcomes depend on the evaluation criteria adopted; and visualization supports the interpretation of all stages. This interdependence is particularly relevant in academic recommender systems, where fairness is closely tied to the governance of scholarly visibility and the distribution of academic opportunities.

From a system-design perspective, the findings suggest that fairness should not be treated as a post-hoc correction problem but as a core design principle. The empirical validation shows that addressing fairness requires coordinated consideration of data quality, model behavior, evaluation metrics, and user-facing interpretability. In this sense, the proposed framework provides not only an analytical structure but also a practical guide for designing and assessing fairness-aware academic recommender systems.

Finally, the validation highlights several directions for future work. These include the development of ranking-aware fairness metrics tailored to scholarly recommendation, improved handling of incomplete or proxy-based protected attributes, and the integration of fairness-aware re-ranking mechanisms within operational recommendation pipelines. Additionally, further research is needed to evaluate the long-term impact of fairness interventions on scholarly visibility and knowledge dissemination.

VI. CONCLUSION

This paper proposed and validated a domain-aware framework for fairness evaluation in academic recommender systems. The framework structures fairness analysis into four dimensions: bias sources, auditing and tooling, mitigation strategies, and evaluation and visualization, enabling a systematic and reproducible assessment process.

Through an empirical study on Ecuadorian scholarly data, the results demonstrated that fairness risks arise from structural properties such as gender imbalance, institutional concentration, and long-tail productivity distributions. The findings also showed that while existing fairness toolkits can detect disparities, their effectiveness depends on domain-aware interpretation and alignment with ranking-based recommendation processes.

Overall, the study confirms that fairness in academic recommender systems requires a multi-dimensional and context-sensitive approach, rather than isolated metrics or post-hoc corrections. The proposed framework provides a practical and interpretable foundation for analyzing and designing fairness-aware scholarly recommendation systems while remaining compatible with existing fairness-auditing and explainability toolkits.

Future work includes the development of ranking-aware metrics, improved handling of incomplete protected attributes, and integration of fairness-aware mechanisms into operational systems.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the *Escuela Politécnica Nacional* for supporting the conference registration associated with this publication.

References

[1] M. S. Pera and Y.-K. Ng, "A personalized recommendation system on scholarly publications," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 2133–2136.

[2] H. Alhoori and R. Furuta, "Recommendation of scholarly venues based on dynamic user interests," *Journal of Informetrics*, vol. 11, no. 2, pp. 553–563, 2017.

[3] L. Recalde, G. Suntaxi, D. Martinez-Mosquera, R. Masabanda, and D. Cabrera, "Centinela: An Intelligent System Based on an Integrated Architecture for Supporting Scholars," in *Information and Communication Technologies. TICEC 2024*. Cham, Suiza: Springer, 2025, pp. 160–177.

[4] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher, "The unfairness of popularity bias in recommendation," *arXiv preprint arXiv:1907.13286*, 2019.

[5] A. J. Biega, K. P. Gummadi, and G. Weikum, "Equity of attention: Amortizing individual fairness in rankings," in *The 41st international acm sigir conference on research & development in information retrieval*, 2018, pp. 405–414.

[6] R. Burke, "Multisided fairness for recommendation," *arXiv preprint arXiv:1707.00093*, 2017.

[7] M. D. Ekstrand, R. Burke, and F. Diaz, "Fairness and discrimination in retrieval and recommendation," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019, pp. 1403–1404.

[8] Z. Zhang, B. G. Patra, A. Yaseen, J. Zhu, R. Sabharwal, K. Roberts, T. Cao, and H. Wu, "Scholarly recommendation systems: A literature survey," *Knowledge and Information Systems*, vol. 65, no. 11, pp. 4433–4478, 2023.

[9] M. Färber, M. Coutinho, and S. Yuan, "Biases in scholarly recommender systems: Impact, prevalence, and mitigation," *Scientometrics*, vol. 128, no. 5, pp. 2703–2736, 2023.

[10] R. Alsaffar and S. Gauch, "Multidimensional demographic profiles for fair paper recommendation," in *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2021)-KDIR*, vol. 1. SCITEPRESS, 2021, pp. 199–208.

[11] Y. Wang, W. Ma, M. Zhang, Y. Liu, and S. Ma, "A survey on the fairness of recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 41, no. 3, pp. 1–35, 2023.

[12] Y. Zhao, Y. Wang, Y. Liu, X. Cheng, C. C. Aggarwal, and T. S. Derr, "Fairness and diversity in recommender systems: A survey," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 1, pp. 1–38, 2025.

[13] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, "Bias and debias in recommender system: A survey and future directions," *ACM Transactions on Information Systems (TOIS)*, vol. 41, no. 3, pp. 1–46, 2023.

[14] W. Fan, X. Zhao, X. Chen, J. Su, J. Gao, L. Wang, Q. Liu, Y. Wang, H. Xu, L. Chen, and Q. Li, "A comprehensive survey on trustworthy recommender systems," *arXiv preprint arXiv:2209.10117*, 2022. [Online]. Available: https://arxiv.org/abs/2209.10117

[15] A. Singh and T. Joachims, "Fairness of exposure in rankings," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 2219–2228.

[16] A. Balayn, C. Lofi, and G.-J. Houben, "Managing bias and unfairness in data for decision support: a survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems," *The VLDB Journal*, vol. 30, pp. 739–768, 2021.

[17] P. Reyero Lobo, E. Daga, H. Alani, and M. Fernandez, "Semantic web technologies and bias in artificial intelligence: A systematic literature review," *Semantic Web*, vol. 14, no. 6, pp. 1097–1122, 2023.

[18] A. Donald, A. Galanopoulos, E. Curry, E. Muñoz, I. Ullah, M. A. Waskow, M. Dabrowski, and M. Kalra, "Bias detection for customer interaction data: A survey on datasets, methods, and tools," *IEEE Access*, vol. 11, pp. 53 690–53 715, 2023.

[19] S. Essahraoui, I. Lamaakal, Y. Maleh, K. El Makkaoui, M. F. Bouami, I. Ouahbi, A. A. Abd El-Latif, M. Almousa, and J. J. P. C. Rodrigues, "Human behavior analysis: A comprehensive survey on techniques, applications, challenges, and future directions," *IEEE Access*, vol. 13, p. 128419, 2025.

[20] X. Meng, Y. Du, Y. Zhang, and X. Han, "A survey of context-aware recommender systems: From an evaluation perspective," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[21] T. B. Araújo, V. Efthymiou, and K. Stefanidis, "Fairness and explanations in entity resolution: An overview," *IEEE Access*, vol. 13, pp. 145 130–145 143, 2025.

[22] S. Ma, "Towards human-centered design of explainable artificial intelligence (xai): A survey of empirical studies," *arXiv preprint arXiv:2410.21183*, 2024. [Online]. Available: https://arxiv.org/abs/2410.21183

[23] S. Lacmanovic and M. Skare, "Artificial intelligence bias auditing–current approaches, challenges and lessons from practice," *Review of Accounting and Finance*, vol. 24, no. 3, pp. 375–400, 2025.

[24] A. Andrade, M. Aguirre, C. Ayala-Tipan, G. Suntaxi, D. Martinez-Mosquera, and L. Recalde, "A Graph-Based Analysis of Collaboration, Trends, and Ethical Dynamics: Perspectives on Ecuadorian Academic Research," in *2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG)*. IEEE, 2025, pp. 206–213.