From Bias to Balance: Fairness-Aware Paper Recommendation for Equitable Peer Review

Uttamasha Anjally Oyshi, Susan Gauch

Department of Electrical Engineering & Computer Science

University of Arkansas

Fayetteville, USA

e-mails: {uoyshi, sgauch}@uark.edu

Abstract—Despite frequent double-blind review, systemic biases related to author demographics still disadvantage underrepresented groups. We start from a simple hypothesis: if a post-review recommender is trained with an explicit fairness regularizer, it should increase inclusion *without* degrading quality. To test this, we introduce *Fair-PaperRec*, a Multi-Layer Perceptron (MLP) with a differentiable fairness loss over intersectional attributes (e.g., race, country) that re-ranks papers after double-blind review. We first probe the hypothesis on *synthetic* datasets spanning high, moderate, and near-fair biases. Across multiple randomized runs, these controlled studies map where increasing the fairness weight strengthens macro/micro diversity while keeping utility approximately stable, demonstrating robustness and adaptability under varying disparity levels. We then carry the hypothesis into the *original* setting—conference data from ACM Special Interest Group on Computer-Human Interaction (SIGCHI), Designing Interactive Systems (DIS), and Intelligent User Interfaces (IUI). In this real-world scenario, an appropriately tuned configuration of Fair-PaperRec achieves *up to* a 42.03% increase in underrepresented-group participation with at most a 3.16% change in overall utility relative to the historical selection. Taken together, the synthetic-to-original progression shows that fairness regularization can act as both an equity mechanism and a mild quality regularizer, especially in highly biased regimes. By first analyzing the behavior of the fairness parameters under controlled conditions and then validating them on real submissions, Fair-PaperRec offers a practical, equity-focused framework for post-review paper selection that preserves—and in some settings can even enhance—measured scholarly quality.

Keywords-Fairness-aware recommendation; Paper selection; Demographic bias mitigation

I. INTRODUCTION

This journal version extends our earlier conference paper [1], offering a deeper and more comprehensive analysis of fairness-aware learning for mitigating post-review bias in academic peer review. Double-blind review often does not eradicate systemic biases linked to authors' demographics, reputations, or institutional

affiliations, despite attempts to ensure impartiality [2], [3], [4], [5]. Recent data indicate that even the most stringent anonymization techniques can be undermined by analyzing writing style or cross-referencing previous articles [6], [7]. This tendency can sustain biases against particular groups, including women, racial minorities, and researchers from underrepresented areas [4], [8], [9], [10]. Simultaneously, there is a growing dependence on recommendation algorithms to optimize processes such as paper selection, grant distribution, and significant publication identification [11], [12], [13]. While these systems can accelerate decision-making, they also pose a danger of perpetuating biases present in the training data, particularly if they focus only on predictive accuracy [14], [15], [16]. Therefore, it is imperative to devise novel methodologies that explicitly include demographic justice, preventing the perpetuation of historical inequalities.

We start from a simple, testable hypothesis on synthetic data: increasing the strength of fairness regularization (λ) should improve inclusion with limited impact on utility, and there should exist a sweet-spot range that depends on the underlying bias level. We map this hypothesis across controlled regimes (fair, moderate, high bias) for multiple protected attributes (race, country), then carry the learned settings into *original* conference data (ACM SIGCHI, DIS, IUI) to validate external relevance. This two-stage progression, from controlled what-if to real-world does-it-hold, reveals when fairness regularization uncovers under-selected high-quality work and when larger λ values begin to over-correct a system that is already close to balanced.

In this paper, we introduce *Fair-PaperRec*, a fairness-aware recommendation framework specifically designed to mitigate *post-review* bias through a differentiable fairness loss integrated with a prediction objective. Unlike heuristic approaches that often handle single-attribute constraints or overlook intersectionality, our approach:

- Surpasses single-attribute approaches by incorporating multiple demographic attributes (e.g., race, country) and constructing profiles that capture underlying biases.
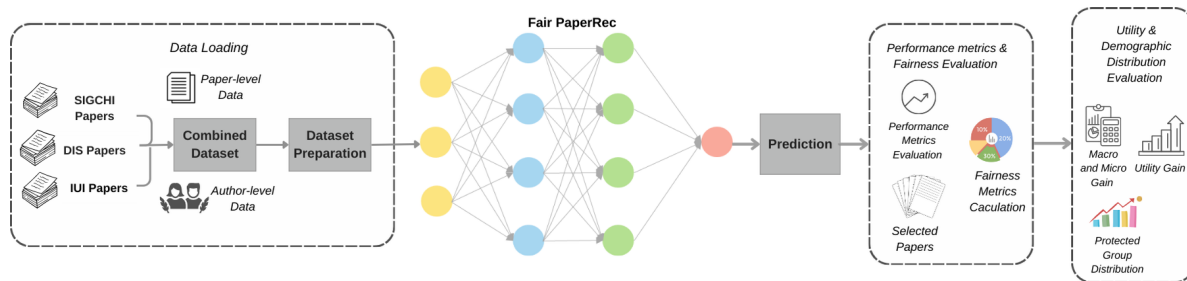
Figure 1. Overview of the Fair-PaperRec Architecture.

- After double-blind review, a specialized fairness penalty is implemented to address demographic disparities, correcting latent biases without replacing existing processes.
- The method explicitly targets demographic parity while monitoring utility, achieving equitable representation without compromising academic rigor.

Our results demonstrate improved representation in the participation of underrepresented groups, as well as an enhancement in overall paper quality. Notably, these findings reveal that enhanced inclusivity need not diminish academic rigor; a fairness-driven approach can yield greater demographic parity while simultaneously preserving, and at times even *enhancing*, the quality of accepted papers.

Contributions: This paper makes three contributions:

1) *A post-review, multi-attribute fairness framework* that integrates a differentiable fairness loss with prediction to address intersectional disparities while preserving quality.

2) *A two-stage evaluation methodology* that (i) charts fairness-utility trade-offs on synthetic bias regimes to identify λ sweet spots, then (ii) validates transfer to original conference data (SIGCHI, DIS, IUI).

3) *Actionable guidance* for choosing λ and attributing weights based on disparity levels, showing when fairness delivers win-win improvements and when smaller regularization suffices.

By mitigating biases in paper selection, our strategy promotes a richer academic discourse and amplifies the representation of marginalized communities, thereby paving the way toward more equitable, high-quality conferences. The paper is organized as follows: Section II reviews related work; Section III presents the methodology; Section IV details the experimental setup and metrics; Section V reports synthetic and original results with analysis; and Section VI concludes.

II. RELATED WORK

We review three strands of literature to motivate our two-stage story, starting from a simple hypothesis on synthetic data and ending in a real-world conclusion on original conference data. First, we examine where and why double-blind review can fail, emphasizing structural incentives beyond anonymity. Second, we connect these pressures to fairness in recommendation and ranking. Third, we cover post-review mitigation and neural approaches most relevant to our setting.

A. Double-Blind Review, Structural Incentives, and Bias in Academic Selection

Although double-blind review conceals identities [2], [3], [4], it often fails to eliminate disparities by gender, race, or geography [10], [17]. Re-identification via stylometry or linkage to prior work can undermine anonymity [6], while prestige effects still favor well-known institutions [10]. Consequently, underrepresented groups remain disadvantaged [18], and acceptance gaps persist [8], [9].

Beyond anonymity, *structural incentives* can shape outcomes. Conflicts of interest (COIs) and funding ties have become more visible as industry participation in ML/NLP expands. Recent large-scale audits report the prevalence and statistical correlates of industry involvement across venues and years [19], [20], with complementary bibliometric evidence that industry and academia contribute differently across problem types, data assets, and deployment pathways [21], [22], [23]. Co-authorship and collaboration networks also influence visibility and impact [24], which can indirectly translate into acceptance advantages. Field norms further matter: in computing, the centrality of *conferences* (deadlines, rapid iteration) versus *journals* can shape incentives and review practices [25], [26]. Together, these factors motivate *post-review* fairness controls that are sensitive to venue norms and network effects, not just anonymization fidelity.

Because peer selection often conflates contribution and credit, authorship order practices vary across subfields and team sizes [27], [28]. Likewise, h-index is field- and seniority-dependent, with wide normative ranges and known limitations as a sole quality proxy [29], [30]. These observations motivate our decision to (i) treat author-level "utility" as a *weighted h-index by*

career stage and (ii) evaluate fairness at *paper* and *author* levels (macro/micro).

B. Fairness in Recommendation and Ranking

When optimizing solely for accuracy, recommenders can amplify historical bias [12], [31]. Multi-objective [14], adversarial [15], and re-ranking approaches [16] offer alternatives but frequently target a *single* protected attribute or generic user–item contexts, leaving intersectional and post-review constraints underexplored. In scholarly selection, *provider fairness* aligns with *author fairness*; the focus is on equitable exposure and acceptance for authors, not only on predictive accuracy [32]. Contemporary venue trends (e.g., topic shifts and data/benchmark access highlighted at CVPR 2024) underscore how exposure and resources interact with selection signals [33], reinforcing the need to treat fairness and quality jointly.

In recommender-systems literature, *provider fairness* refers to ensuring equitable exposure or selection opportunities for content creators. In the academic peer-review context, the "providers" are the authors themselves; thus, provider fairness corresponds directly to *author fairness*, aiming to equalize acceptance probabilities across demographic groups. This distinction clarifies that our work focuses on fairness for authors (providers), rather than fairness for users of the recommendation system.

Our synthetic-to-original narrative builds on this literature by (i) explicitly modeling *multiple* protected attributes (race, country), (ii) mapping fairness-utility trade-offs across tunable bias regimes to locate λ "sweet spots," and (iii) transferring those settings to real conference data where industry/academia dynamics, network effects, and venue norms co-exist [21], [22], [23], [25], [26].

C. Post-Review Bias Mitigation and Neural Approaches

Heuristic post-review balancing can improve representation [32] but risks local optima and often omits multi-attribute fairness. Neural approaches such as *DeepFair* [12] and *Neural Fair Collaborative Filtering* [34] show that fairness and accuracy can align in commercial domains, yet peer review imposes hard quotas, reputational stakes, and limited items. Multi-stakeholder optimization [35], [36] argues for context-aware definitions (e.g., exposure vs. acceptance parity), while text-based relevance features (e.g., TF–IDF) can help ranking [37] but do not, on their own, guarantee equity for historically marginalized authors.

We focus on a *post-review* neural re-ranking stage with a differentiable fairness loss that handles *intersectional* attributes. The synthetic analysis provides a controlled test of the hypothesis that increasing fairness regularization improves inclusion with limited utility cost—and

reveals regime-dependent sweet spots. The original-data validation then asks whether these settings hold under real constraints shaped by industry–academia participation, collaboration networks, and conference practices [19], [20], [22], [24], [25]. This two-stage path converts a simple, testable idea into a practical recipe for equity-preserving paper selection.

III. METHODOLOGY

Our approach tackles demographic biases in conference data to enforce fairness post-review: we begin with a simple, controlled hypothesis tested on synthetic datasets, then validate its robustness on original conference data. The design therefore emphasizes both (i) tractability in controlled regimes and (ii) fidelity to the complex realities of peer review. The backbone of our approach is a lightweight Multilayer Perceptron (MLP) trained with an integrated fairness loss, as illustrated in Fig. 1. This choice reflects two guiding principles: (1) expose and alleviate demographic disparities instead of concealing them, and (2) use a simple, transparent neural architecture that harmonizes fairness and utility while remaining interpretable enough to trace effects across bias regimes.

A. Data Collection and Pre-processing

a) Synthetic stage (hypothesis testing): Synthetic datasets are constructed with tunable demographic distributions (fair, moderate, high bias). These provide a sandbox where the effect of fairness regularization λ can be precisely measured. By varying the demographic skew, we can test the hypothesis that fairness constraints are most effective when initial bias is high, and that "sweet spot" λ values differ by attribute. Table I summarizes the demographic distributions for each protected group across the fair, moderate, and high synthetic bias scenarios.

b) Original stage (validation): To evaluate transfer to real contexts, we utilize datasets from SIGCHI 2017, DIS 2017, and IUI 2017 [32]. These venues naturally exhibit systemic imbalances in demographics and prestige, reflecting the challenges of real peer review. Rather than discarding these biases, we preserve them to observe how fairness regularization reshapes outcomes in situ. Table II reports the participation rates of protected groups (gender, race, country) in the three conferences (SIGCHI, DIS, IUI).

c) Processing pipeline: At both stages, the dataset is curated at paper and author levels. Every paper record includes a title, author list, and conference label (1 = IUI, 2 = DIS, 3 = SIGCHI). Author records capture demographic variables (gender, race, nationality, career stage). Real-world datasets—particularly those drawn from academic conference submissions—often contain

TABLE I. Demographic Distribution in Synthetic Datasets Across Bias Scenarios For Protected Groups

Scenario	Gender (%)	Race (%)	Country (%)
Fair	48.80	51.50	52.30
Moderate	28.10	28.90	31.50
High	8.50	9.70	10.00
Average	28.47	30.03	31.27

latent biases that mirror systemic imbalances in the scholarly community (e.g., underrepresentation of certain demographics). We utilize datasets from conferences that naturally reflect systemic disparities (e.g., skewed demographics). Instead of eliminating such biases, our objective is to *recognize and rectify* them.

TABLE II. DEMOGRAPHIC PARTICIPATION FROM PROTECTED GROUPS IN THREE CONFERENCES.

Conference	Gender (%)	Race (%)	Country (%)
SIGCHI	41.88	6.84	21.94
DIS	65.79	35.09	24.56
IUI	43.75	51.56	39.06
Average	50.47	31.16	28.52

1) Data Description: The resulting dataset combines both paper-level and author-level perspectives. *Overall* refers to all submissions, while *Selected* refers to those recommended by Fair-PaperRec. SIGCHI 2017 papers act as a benchmark for high-impact work, offering a ground truth for quality. Author records include career stages (student, postdoc, faculty, industry), enabling *h-*index weighting that reflects differences in scholarly maturity.

2) Data Pre-processing: Several preprocessing steps were undertaken to prepare the dataset for training:

Categorical Encoding: Gender, Country, and Race are subjected to one-hot encoding. Gender is binary (0 = male, 1 = female), Country is categorized as *developed* or *underdeveloped*, and Race comprises {White, Asian, Hispanic, Black}, with Hispanic and Black designated as protected groups (Table II). *Normalization:* Numerical attributes (e.g., h-index) employ min-max scaling for consistent magnitude. *Training and Validation Division:* An 80%/20% stratified division guarantees equitable distribution of labels and protected attributes in both subsets.

B. Problem Definition

This study develops a *fairness-aware paper recommendation system* that ensures demographic parity with respect to authors' race and country while preserving high academic standards. We frame acceptance decisions as a *recommendation* task, where *conference organizers (users)* seek to select from *530 papers (items)* spanning

SIGCHI, DIS, and IUI. Each paper (item) includes an *h-index* for quality, demographic data (race, country), and a conference rating.

Our approach enforces fairness constraints on race and country independently, excluding *gender* due to its relatively balanced distribution in the dataset (Table II). Preliminary analysis further showed minimal disparity between accepted and rejected papers with respect to gender; because our fairness framework penalizes deviations from the observed baseline, incorporating gender would have produced negligible changes in the fairness–utility trade-offs examined here. To maintain clarity of exposition and focus on attributes with substantially higher disparity, we therefore restricted fairness constraints to race and country. Nevertheless, gender remains an important dimension of equity, and future work will extend Fair-PaperRec to gender-aware and multi-attribute fairness in larger or more imbalanced datasets.

Let D represent the dataset of submitted papers, where each paper $p \in D$ is associated with a set of features X_p (e.g., race, country, h-index) and a target variable y_p indicating acceptance (1) or rejection (0). The *race* attribute R_p and *country* attribute C_p are the protected attributes.

We aim to optimize a predictive model $f : X_p \rightarrow \hat{y}_p$ that minimizes the following objective function:

$$\min_f \left(\mathcal{L}(f(X_p), y_p) + \lambda \cdot \mathcal{L}_{\text{fairness}}(f, D) \right) \quad (1)$$

Here, $\mathcal{L}(f(X_p), y_p)$ is the *prediction loss* (e.g., Binary Cross-Entropy Loss), $\mathcal{L}_{\text{fairness}}(f, D)$ is the *fairness loss*, penalizing deviations from demographic parity across race and country, and λ is a hyperparameter that balances the trade-off between prediction accuracy and fairness.

Algorithm 1. FAIR-PAPERREC LOSS FUNCTION.

1: **Input**: Model M, Epochs E, Batch size B, Data D, Protected attributes A, Hyperparameter λ
2: **Output**: Trained Model M
3: **Initialize** Model M
4: **for** each $e \in E$ **do**
5: Shuffle Data D
6: **for** each batch $\{(X, Y)\} \in D$ with size B **do**
7:   Predict $\hat{Y} \leftarrow M(X)$
8:   Calculate Loss:
9:    $L_{\text{prediction}} \leftarrow \text{PredictionLoss}(Y, \hat{Y})$
10:    $L_{\text{fairness}} \leftarrow \text{FairnessLoss}(A, \hat{Y})$
11:   Calculate Total Loss:
12:    $L_{\text{total}} \leftarrow \lambda \cdot L_{\text{fairness}} + L_{\text{prediction}}$
13:   Compute gradients $\nabla L_{\text{total}} \leftarrow \frac{\partial L_{\text{total}}}{\partial M}$
14:   Update Model parameters: $M \leftarrow M - \alpha \nabla L_{\text{total}}$
15: **end for**
16: **end for**

C. Demographic Parity

We aim to ensure that the probability of a paper being accepted is independent of the protected attributes:

$$P(\hat{y}_p = 1 \mid R_p = r) = P(\hat{y}_p = 1), \quad \forall r \in \text{Race}$$

$$P(\hat{y}_p = 1 \mid C_p = c) = P(\hat{y}_p = 1), \quad \forall c \in \text{Country}$$

Utilizing these equations ensures that the papers authored by individuals from different races and countries have an equal probability of acceptance.

D. Fairness Loss

The fairness loss from the objective function in Equation (1) is constructed to minimize statistical parity differences between the protected and non-protected groups:

$$\mathcal{L}_{\text{fairness}} = \left(P(\hat{y}_p = 1 \mid G_{\text{p}}) - P(\hat{y}_p = 1 \mid G_{\text{np}}) \right)^2 \quad (2)$$

Here, $P(\hat{y}_p = 1 \mid G_{\text{p}})$ denotes the acceptance probability for the protected group and $P(\hat{y}_p = 1 \mid G_{\text{np}})$ is the acceptance probability for the non-protected group.

E. Combined Fairness Loss

Furthermore, we define a combined fairness loss to minimize statistical parity differences across race and country attributes between the protected and unprotected groups, as shown in Equation (3).

$$
\begin{aligned}
\mathcal{L}_{\text{fairness}} = W_r &\left(\frac{1}{N_r} \sum_{p \in G_r} \hat{y}_p - \frac{1}{N} \sum_{p=1}^{N} \hat{y}_p \right)^2 \\
+ W_c &\left(\frac{1}{N_c} \sum_{p \in G_c} \hat{y}_p - \frac{1}{N} \sum_{p=1}^{N} \hat{y}_p \right)^2
\end{aligned}
\quad (3)
$$

G_{r} and G_{c} denote the race and country groups, respectively. N_r and N_c are the number of papers in each group and weights W_r and W_c reflect group distributions.

F. Total Loss

The total loss is the combination of prediction and fairness losses:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{prediction}} + \lambda \cdot \mathcal{L}_{\text{fairness}}$$

G. Constraints and Considerations

We assess fairness by training our model separately on *race* and *country*, as well as jointly on both attributes to evaluate selection fairness across multiple dimensions.

a) Exclusion of Protected Attributes: Race R_p and country C_p are excluded from the input feature set X_p to mitigate direct bias amplification. To achieve joint fairness, both attributes are omitted during training, preventing the model from learning acceptance outcomes influenced by race or country.

b) Indirect Bias Mitigation: A fairness loss promotes demographic parity, addressing indirect biases associated with features related to race or country. The model maintains neutrality by penalizing selection disparities, even in the absence of protected attributes.

c) Scalability: The MLP architecture is deliberately simple, enabling reproducibility across venues and bias levels. Synthetic results provide controlled insight into when fairness is most beneficial; original results demonstrate external validity. Together, they build a methodological bridge from a simple hypothesis ("fairness regularization should help most under bias") to a powerful conclusion ("in real conferences, tuning λ uncovers under-selected, high-quality work without compromising rigor").

IV. MODEL OVERVIEW

To achieve demographic parity while preserving quality in paper selection, we present a MLP-based neural network (see Fig. 1), explicitly engineered to balance the trade-off between fairness and accuracy. It illustrates the correlations between input features, like author demographic attributes and paper quality, while alleviating biases during selection.

A unique fairness loss function was employed to ensure equity, imposing penalties on the model for substantial differences in selection rates between protected and non-protected groups. This loss function is integrated with the conventional prediction loss to attain a balance between diversity and accuracy; the algorithm is shown in Algorithm 1.

The acceptance probabilities for submitted papers are generated by the MLP, which are subsequently ranked to guarantee that the final selection meets both quality and fairness objectives. By selecting top papers according to these probabilities, we ensure equal representation of authors from both protected and non-protected groups while upholding the requisite standard of academic excellence.

Algorithm 2. FAIRNESS-AWARE PAPER SELECTION MECHANISM.

1: **Input**: Dataset D, Model M, Number of Accepted Papers N_a, Total Papers N_t
2: **Output**: Selected Papers P_{selected}
3: **Initialize**: $P_{\text{selected}} \leftarrow \emptyset$
4: **Step 1**: Apply trained model M to the entire dataset D
5: **for** each paper $p \in D$ **do**
6:      Compute acceptance probability: $\hat{y}_p \leftarrow M(p)$
7: **end for**
8: **Step 2**: Rank all papers p by acceptance probability $\hat{y}_p$
9:      Sort D in descending order of $\hat{y}_p$
10: **Step 3**: Select top N_a papers:
11:      $P_{\text{selected}} \leftarrow \{p \mid \hat{y}_p \geq \hat{y}_{(N_a)}\}$
12: **Step 4**: Ensure Fairness Constraints
13: **Return** P_{selected}

TABLE III. RELATIVE GAIN (%) FOR MACRO, MICRO, AND UTILITY METRICS ACROSS $\lambda \in [1, 10]$ FOR COUNTRY AND RACE FEATURES UNDER DIFFERENT BIAS LEVELS.

Bias Level	λ	Country Feature			Race Feature		
		Macro Gain (%)	Micro Gain (%)	Utility Gain (%)	Macro Gain (%)	Micro Gain (%)	Utility Gain (%)
High	1	-18.03	-4.17	-2.98	**15.03**	20.18	8.51
	2.5	-3.93	-1.65	-1.28	12.44	18.46	2.55
	3	-1.18	3.22	5.96	12.11	**21.10**	**15.32**
	5	-0.30	-2.61	8.09	10.12	16.84	14.47
	10	**4.98**	**9.22**	**10.21**	5.06	14.20	5.53
Moderate	1	2.90	**10.44**	**15.32**	3.00	**10.62**	11.06
	2.5	-1.11	-2.83	4.68	1.82	4.66	8.09
	3	**5.04**	3.97	7.23	**3.96**	9.54	**13.62**
	5	-3.15	-6.15	0.43	4.39	10.55	11.06
	10	2.29	0.49	-0.85	0.83	3.05	7.66
Fair	1	**5.35**	**7.26**	**20.43**	-6.51	0.79	-4.26
	2.5	0.51	0.02	8.94	-0.28	-0.30	11.06
	3	3.81	2.01	-0.85	-8.96	-12.78	8.09
	5	2.31	1.43	3.83	-2.12	-4.76	**12.77**
	10	3.93	2.12	10.64	**1.37**	-1.78	4.68

A. Selection Mechanism

The model calculates acceptance probabilities for all submitted papers after training. After calculating acceptance odds, the algorithm ranks candidate papers. This rating phase ensures that underrepresented groups are represented in the final admission decisions. Representing this as a suggestion list preserves the peer-review process and corrects residual biases. Algorithm 2 selects the best papers based on probability, ensuring fairness and preserving the desired number of accepted papers.

- *Prediction Aggregation:* The trained MLP model is applied to the entire dataset to obtain predicted acceptance probabilities $\hat{y}_p$ for each paper.
- *Ranking:* Papers are ranked in descending order based on their predicted probabilities.
- *Selection:* The papers with the highest predicted probabilities are selected for acceptance, ensuring that the total number of selected papers matches the required acceptance quota.

Mathematically, the selection process is represented as:

$$\text{Selected Papers} = \left\{ p \in D \mid \hat{y}_p \geq \hat{y}_{(N_a)} \right\}$$

Here, $\hat{y}_{(N_a)}$ is the N_a-th highest predicted probability in the set $\{\hat{y}_p \mid p \in D\}$ while N_a is the total number of accepted papers and N_t is the total number of submitted papers, where $N_a \leq N_t$.

This approach ensures that the selection process is both informed by the model's predictions and constrained to uphold demographic parity, fostering an equitable and meritocratic paper selection environment.

V. EVALUATION AND EXPERIMENTS

This section presents the experimental evaluation of our proposed Fair-PaperRec model on the chosen datasets. First, we test our hypothesis in controlled synthetic settings, then we validate it with original conference data. This design allows us to ask not only whether fairness interventions work, but also *when* and *why* they are most effective.

To guide this exploration, we pose three research questions:

- *RQ1:* How do fairness constraints affect the overall quality (utility) of recommended papers, as measured by author h-index?
- *RQ2:* Does handling race and country as separate protected attributes differ from treating them jointly in terms of fairness outcomes and selection decisions?
- *RQ3:* How do varying weight assignments to multiple protected attributes (race and country) influence the trade-off between fairness and utility?

These questions move us along the synthetic-to-original trajectory, from controlled hypothesis testing to real-world validation. RQ1 begins in the synthetic environment, where we test the basic hypothesis that fairness regularization (λ) improves inclusion with limited cost to utility. RQ2 and RQ3 extend to the real-world stage, asking whether the same patterns hold under authentic demographic distributions and multi-attribute interactions.

A. Experimental Setting

We evaluate Fair-PaperRec across both synthetic datasets (fair, moderate, and high-bias regimes) and original conference datasets. This dual evaluation ensures that our findings generalize from controlled testbeds to complex real-world settings where structural factors (prestige, career stage, institutional affiliation) interact with demographics.

Each experiment is conducted five times with independent random seeds. We report mean values with *standard deviations* to highlight stability across runs, ensuring
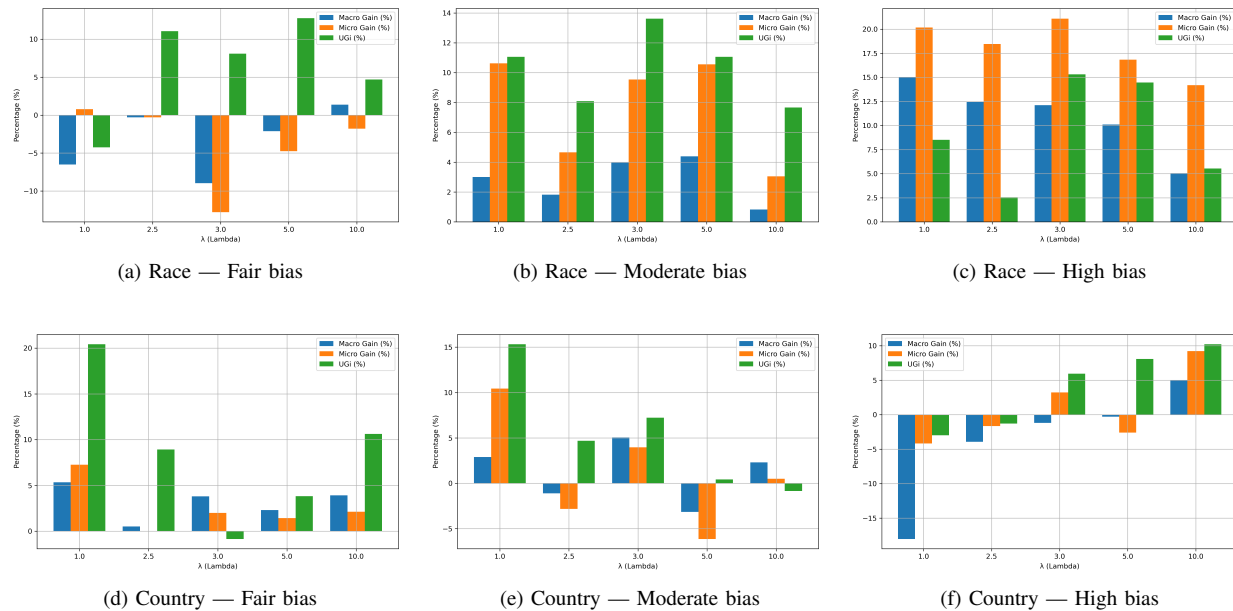
(a) Race — Fair bias (b) Race — Moderate bias (c) Race — High bias

(d) Country — Fair bias (e) Country — Moderate bias (f) Country — High bias

Figure 2. Relative gains (%) for Macro, Micro, and Utility across fairness regularization strengths ($\lambda \in [1, 10]$). Top row: Race; bottom row: Country.

that observed peaks or trade-offs are not artifacts of stochastic training.

1) Implementation Details: All experiments are implemented in `PyTorch` and executed on a high-performance server equipped with two NVIDIA Quadro RTX 4000 GPUs. The Fair-PaperRec model is a two-hidden-layer MLP using Rectified Linear Unit (ReLU) activations and Batch Normalization, culminating in a sigmoid output layer that produces acceptance probabilities.

TABLE IV. DISTRIBUTION OF RECOMMENDED PAPERS FROM EACH TIER FOR SYNTHETIC DATASET.

Label	Country	Race	Multi-Fair
Top-tier	87.12%	89.30%	91.41%
Mid-tier	6.45%	6.87%	5.64%
Low-tier	6.43%	3.83%	2.95%
# Papers	280	280	280

We train for 50 epochs using the Adam optimizer (learning rate = 0.001). Early stopping is applied if validation performance does not improve over 10 consecutive epochs. The fairness regularization parameter λ is swept across a range of values (1–10) to trace fairness–utility trade-offs. Datasets are split 80/20 (training/validation) using stratified sampling, preserving both label and demographic distributions.

This setup allows us to simulate the "synthetic hypothesis-testing stage" (controlled bias ratios) and then apply the same procedure in the "original validation stage" (real demographic skew).

TABLE V. DISTRIBUTION OF RECOMMENDED PAPERS FROM EACH CONFERENCE.

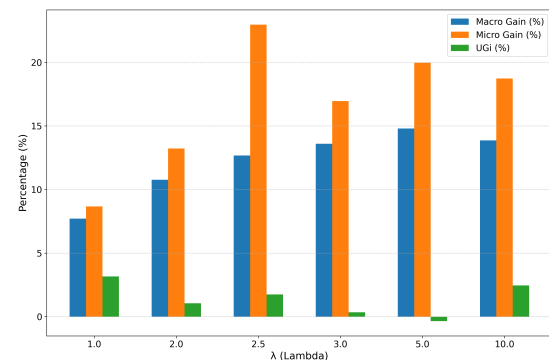| Label	Country	Race	Multi-Fair
SIGCHI	92.02%	92.00%	92.02%
DIS	4.84%	7.69%	7.40%
IUI	3.14%	0.31%	0.56%
# Papers	351	351	351

Figure 3. Comparison of Macro and Micro Gains for Country Across Different Fairness Configurations.

2) Baseline: We compare our model against a baseline demographic-blind model, which is a conventional (MLP) model that prioritizes quality and ignores fairness constraints. This model selects the original list of papers chosen by the SIGCHI 2017 program committee. By contrasting it with Fair-PaperRec, we identify λ values where fairness gains are achieved without utility loss (addressing RQ1). In original settings, the baseline rep-
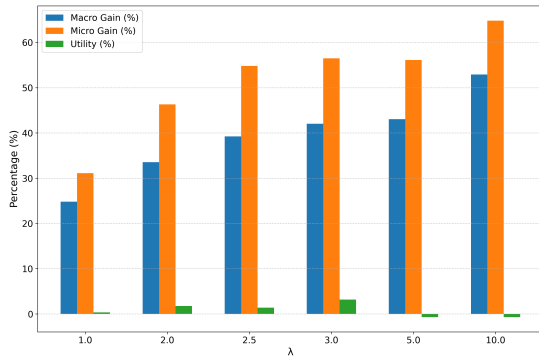
Figure 4. Comparison of Macro and Micro Gains for Race Across Different Fairness Configurations.

resents the actual peer-review outcomes; improvements over it demonstrate the practical value of fairness interventions in real conferences (addressing RQ2 and RQ3).

3) Parameters: A hyperparameter λ is used to control the trade-off between prediction accuracy and fairness. Higher values emphasize fairness more strongly.

The weights W_c and W_r, respectively, denote the weighting factors assigned to the country and race attributes in the fairness loss function, as shown in Equation (3).

B. Evaluation Metrics

Diversity is assessed at both the *paper level* and the *author level*. In particular:

- *Macro Gain* represents the percentage increase in the diversity of each feature within the selected papers compared with the baseline, assessing the overall representation of protected groups.
- *Micro Gain* is the percentage increase in the diversity of each feature among authors of the selected papers, providing a more detailed perspective on inclusivity.
- A *Diversity Gain* [32] further normalizes these macro-level changes in Equation (4), capping each feature at 100 to avoid any single attribute skewing the total.
- The *F - measure* [32] Equation (5) then combines this diversity improvement with the resulting utility, offering a harmonic balance between fairness gains and paper quality.

Table VII summarizes the fairness, diversity, and utility metrics used in our evaluation, including Macro Gain, Micro Gain, Diversity Gain, and the F-measure.

To ensure that diversity enhancements do not compromise scholarly quality, we assess *Utility Gain* (UG_i) using a weighted h-index that reflects an author's career stage (professor, associate professor, lecturer, postdoctoral researcher, or graduate student). While the h-index offers a convenient proxy for scholarly influence, it also

encodes seniority and field-specific publication practices that may introduce their own biases. For this reason, we interpret utility gain cautiously, not as an absolute measure of "quality," but as a stability indicator demonstrating whether fairness interventions disproportionately select low impact papers. By comparing the resulting utility values against the baseline distribution, we evaluate whether fairness-aware re-ranking maintains, improves, or undermines the overall quality profile of accepted papers.

$$D_G = \frac{\sum_{i=1}^{n} \min(100, \text{Macro Gain}_{G_i})}{n} \quad (4)$$

$$F = 2 \times \frac{D_G \times (100 - UG_i)}{D_G + (100 - UG_i)} \quad (5)$$

C. Interpretation of the Results

We evaluate the fairness regularization parameter (λ) across values 1–10 to examine its impact on fairness (macro and micro gains) and quality (Utility Gain, UGi). The results span both synthetic datasets and original conference datasets. This two-stage design allows us to first test the hypothesis under controlled conditions and then validate it using real-world peer review data.

a) RQ1: Effect of fairness on utility: Fig. 2 (synthetic) and Figs. 4–3 (original) show that increasing λ generally improves macro and micro diversity, while the trajectory of utility differs across settings. Detailed gain calculations are reported in Table III.

- In synthetic *high-bias* settings, we observe a "sweet spot" ($\lambda \approx 3$) where both fairness and utility peak together. This supports our hypothesis that correcting bias can uncover under-selected, high-quality work.
- In synthetic *moderate/fair* settings, small λ values act as a useful regularizer, but large values λ over-correct, harming utility.
- In original data, utility is more stable for race (flat or mildly positive trend) but fluctuates for country, with error bars indicating greater uncertainty at higher λ. This confirms the trade-off: fairness gains are achievable, but utility must be monitored carefully.

TABLE VI. GAIN CALCULATIONS FOR COUNTRY AND RACE FEATURES WITH UTILITY GAIN (UG_i).

λ	Country Feature			Race Feature		
	Macro Gain (%)	Micro Gain (%)	UG_i (%)	Macro Gain (%)	Micro Gain (%)	UG_i (%)
1	7.71	8.67	3.16	24.81	31.11	0.35
2	10.77	13.23	1.05	33.54	46.30	1.75
2.5	**12.67**	**22.96**	**1.75**	39.25	54.81	1.40
3	13.60	16.96	0.35	**42.03**	**56.48**	**3.16**
5	14.80	19.97	-0.35	43.04	56.11	-0.70
10	13.86	18.73	2.46	52.91	64.81	-0.70

(a) Utility Gain.   (b) Macro/Micro for Race.   (c) Macro/Micro for Country.
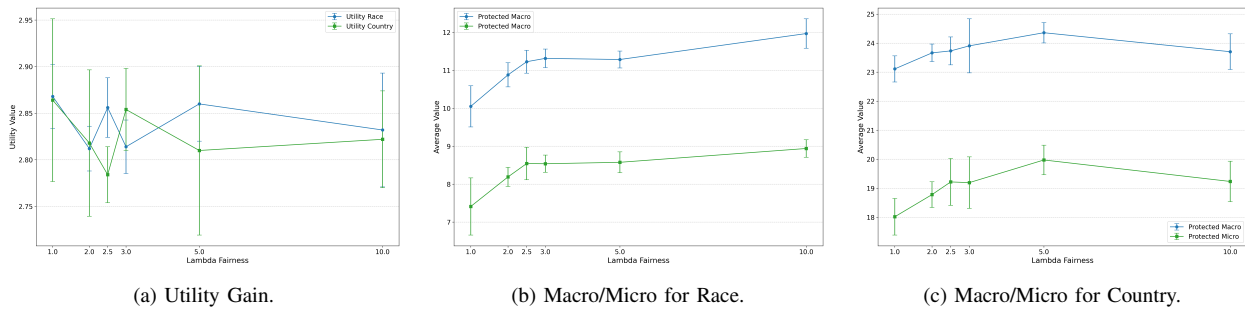
Figure 5. Comparison of gains across different fairness configurations.

b) RQ2: Attribute-specific effects (race vs. country): The optimal λ differs between attributes. Race requires larger λ values (around 3) to balance fairness and utility, reflecting the higher initial disparity ratios of racial minorities. Country achieves balance at lower λ values (around 2.5). Synthetic experiments revealed the same pattern: the more skewed the initial distribution, the stronger the regularization required. This validates our hypothesis that fairness policies must be tailored to group-specific disparities. A summary of these trade-offs for different λ values and weight assignments is provided in Table VII.

c) RQ3: Multi-attribute and weighting trade-offs: Experiments that combine race and country with varying weights show that emphasizing one attribute can reduce gains for the other. On synthetic data, joint optimization occasionally diluted improvements compared to single-attribute runs. On original data, this was visible in Micro diversity for country, which exhibited more volatility than race under joint fairness. As summarized in Fig. 5, these configurations reveal how different fairness weightings reshape the balance between utility and diversity, making the trade-offs between race-focused, country-focused, and multi-attribute fairness more explicit.

d) Macro vs. Micro Dynamics: Across both synthetic and original settings, macro diversity measures were consistently more responsive to fairness regularization than micro measures. Macro gains rose steadily with λ, especially for race, whereas micro gains were more variable (particularly for country). This suggests that fairness interventions are more effective at improving overall group-level representation than at equalizing outcomes for individual authors.

e) Conference-level distribution: The detailed gain calculations for each λ are shown in Table VI, which reports the breakdown of accepted papers across SIGCHI, DIS, and IUI under different fairness constraints. SIGCHI dominates acceptance at approximately 92% regardless of λ, reflecting its strong baseline prestige. However, fairness constraints modestly increase the share of DIS and IUI papers, broadening representation without destabilizing the overall distribution. This in-

dicates that fairness-aware re-ranking can diversify the conference portfolio while preserving dominant trends. For context, the original conference-level distribution of accepted papers is summarized in Table V. In the synthetic stage, the baseline distribution is configured to mimic this original selection profile under controlled bias levels; the corresponding breakdown is reported in Table IV.

D. Robustness & Sensitivity

For completeness, we report seed variability (mean±std over multiple random seeds) and, when space allows, confidence intervals around the λsweeps to demonstrate that peaks (e.g., Race–High at $\lambda \approx 3$) are stable rather than artifacts of stochasticity.

E. Ablation Study: Multi-Demographic Fairness

The objective of our ablation study was to evaluate the model's performance when optimizing fairness across multiple demographic attributes simultaneously, specifically with respect to both *country* and *race*. This ablation was conducted to address *RQ3*, which explores the impact of varying fairness weights for each attribute when multiple fairness attributes are considered together.

To ensure fairness, we removed these attributes from the input space, preventing the model from learning direct associations between them and the paper acceptance decisions. Instead, demographic parity loss was computed for each attribute during training, capturing deviations from fairness. The parity losses for both country and race were combined by assigning weights: W_c for country and W_r for race, with the initial weights set to $W_c = 0.68$ and $W_r = 0.32$, reflecting the distribution of protected groups.

To further explore the model's behavior and answer *RQ3*, we varied these weights, first increasing W_c while keeping W_r constant and then increasing W_r while keeping W_c fixed. Additionally, we experimented with different values of the fairness regularization parameter λ, which controls the trade-off between fairness and utility. These experiments allowed us to observe how different weight configurations and fairness constraints

TABLE VII. GAIN CALCULATIONS FOR COUNTRY AND RACE FEATURES WITH UTILITY GAIN.

λ	Weights	Country Feature		Race Feature		UG_i (%)	Avg. D_G (%)	Avg. F (%)
		Macro Gain (%)	Micro Gain (%)	Macro Gain (%)	Micro Gain (%)			
1	$(W_r = 0.32, W_c = 0.68)$	6.17	6.34	30.51	46.30	3.16	44.66	53.71
	$(W_r = 1, W_c = 2)$	6.73	9.15	-0.25	0.37	2.81	6.48	13.77
	$(W_r = 2, W_c = 1)$	7.43	11.43	12.91	16.11	3.16	25.63	40.36
2	$(W_r = 0.32, W_c = 0.68)$	**13.60**	**24.43**	30.51	42.22	**4.21**	55.38	68.47
	$(W_r = 1, W_c = 2)$	5.24	6.88	15.45	17.96	0.70	20.69	21.58
	$(W_r = 2, W_c = 1)$	8.36	12.86	39.49	54.26	1.75	26.31	21.58
2.5	$(W_r = 0.32, W_c = 0.68)$	8.63	17.33	36.58	50.37	2.46	56.46	66.31
	$(W_r = 1, W_c = 2)$	9.89	14.00	30.63	46.30	2.81	40.52	62.09
	$(W_r = 2, W_c = 1)$	9.60	17.11	42.53	56.48	1.40	59.25	69.98
3	$(W_r = 0.32, W_c = 0.68)$	7.15	11.42	39.49	53.89	1.40	55.98	63.45
	$(W_r = 1, W_c = 2)$	10.16	21.17	33.29	43.89	0.70	43.45	47.63
	$(W_r = 2, W_c = 1)$	9.60	18.35	42.53	55.37	2.81	61.90	47.63
5	$(W_r = 0.32, W_c = 0.68)$	10.80	19.38	45.82	58.52	0.70	**65.09**	**72.92**
	$(W_r = 1, W_c = 2)$	4.69	3.88	33.92	40.19	0.35	38.61	15.73
	$(W_r = 2, W_c = 1)$	7.43	11.90	39.49	52.96	5.26	52.26	15.73
10	$(W_r = 0.32, W_c = 0.68)$	9.60	18.34	42.53	55.37	1.40	62.92	70.89
	$(W_r = 1, W_c = 2)$	7.43	13.91	24.94	25.19	4.91	32.37	34.88
	$(W_r = 2, W_c = 1)$	7.43	11.72	35.44	47.41	-4.21	40.53	34.88

influenced the model's ability to achieve demographic fairness while maintaining utility and the quality of selected papers.

The results of the ablation study, shown in Table VII, reveal that at $\lambda = 1$, assigning equal weights to both race and country ($W_r = 0.32, W_c = 0.68$) produced significant gains for race, with a Macro Gain of 30.51% and a Micro Gain of 46.3%, while country showed relatively smaller improvements (6.17% and 6.34%, respectively). However, when the weight for country was increased ($W_c = 2 \times 0.68$), diversity gains for race dropped sharply, with a negative macro gain (-0.25%), while country experienced slight improvements. Conversely, increasing the weight for race ($W_r = 2 \times 0.32$) resulted in improved diversity for both race and country, indicating that assigning more weight to race enhances diversity for both attributes to some degree.

At $\lambda = 2.5$, the model achieved the best balance between diversity and utility. Equal weights for race and country yielded macro and micro gains of 36.58% and 50.37% for race, and 8.63% and 17.33% for country, with a low utility loss of 2.46%. This suggests that $\lambda = 2.5$ is optimal for balancing fairness and utility. As λ increases further, race diversity continues to improve (reaching 45.82% Macro Gain at $\lambda = 5$), but at the cost of decreasing utility. The different optimal λ values for race and country suggest that disparity ratios impact how fairness constraints should be weighted, with race requiring a higher λ due to its higher disparity ratio. This leads to greater race diversity gains at higher λ values, whereas country achieves optimal results at moderate λ values, such as 2.5.

These findings directly address *RQ3*, demonstrating that fairness weights must be carefully calibrated for each protected attribute. Assigning greater weight to race tends to improve diversity for both race and country, whereas increasing the weight for country may result in reduced fairness for race. The optimal balance between fairness and utility is achieved when fairness weights and λ values are adjusted based on the unique disparity ratios of each attribute.

VI. CONCLUSION AND FUTURE WORK

We began our experiment with a hypothesis tested on synthetic data: a lightweight fairness regularizer, controlled by λ, could improve demographic inclusion without compromising quality. Applying the same regularization to the original conference data confirmed and strengthened this insight: fairness regularization consistently improved representation, particularly for race, while leaving utility largely stable or mildly positive.

These results highlight three key lessons. First, fairness effects depend on the degree of underlying disparity: high-bias settings benefit from stronger regularization, while near-fair regimes require small adjustments to avoid over-correction. Second, attributes vary in sensitivity: race often demands more intervention than country, and multi-attribute optimization requires careful balancing. Third, in strongly biased systems, fairness and quality are not in conflict; debiasing uncovers underselected, high-quality work.

At the same time, our approach has limits. We rely on a simple MLP with post-hoc fairness loss, excluding protected attributes from the input but not explicitly modeling causal pathways or reviewer dynamics. Extending this framework with causal inference, graph-based architectures capturing author–institution–topic relations, or generative approaches such as VAEs could deepen bias mitigation while preserving interpretability. Broader definitions of fairness, adaptive tuning of λ and attribute weights, and human-in-the-loop evaluation are promising directions. Embedding fairness interventions into real-world workflows will require robustness to distribution shifts, transparency in reporting, and governance

mechanisms that account for privacy and accountability. Together, these steps can transform Fair-PaperRec from a research prototype into a practical, equitable tool for the future of scholarly peer review.

REFERENCES

[1] U. A. Oyshi and S. Gauch, "Fair learning for bias mitigation and quality optimization in paper recommendation," in *Proceedings of the 17th International Conference on Information, Process, and Knowledge Management (eKNOW 2025)*, ThinkMind Digital Library, Nice, France: IARIA Press, May 2025, pp. 43–50, ISBN: 978-1-68558-272-2.

[2] A. Tomkins, M. Zhang, and W. D. Heavlin, "Reviewer bias in single- versus double-blind peer review," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 48, pp. 12 708–12 713, 2017. DOI: 10.1073/pnas.1707323114.

[3] E. Shmidt and B. Jacobson, "Double-blind reviews: A step toward eliminating unconscious bias," *Clinical and Translational Gastroenterology*, vol. 13, no. 1, e00443, 2022. DOI: 10.14309/ctg.0000000000000443.

[4] V. P. Giannakakos, T. S. Karanfilian, A. D. Dimopoulos, and A. Barmettler, "Impact of author characteristics on outcomes of single- versus double-blind peer review: A systematic review of comparative studies in scientific abstracts and publications," *Scientometrics*, vol. 130, pp. 399–421, 2025. DOI: 10.1007/s11192-024-05213-x.

[5] C. Mebane, "Double-blind peer review is detrimental to scientific integrity," *Environmental Toxicology and Chemistry*, vol. 44, pp. 318–323, 2025. DOI: 10.1093/etojnl/vgae046.

[6] L. Bauersfeld, A. Romero, M. Muglikar, and D. Scaramuzza, "Cracking double-blind review: Authorship attribution with deep learning," *PLoS ONE*, vol. 18, no. 6, e0287611, Jun. 2023. DOI: 10.1371/journal.pone.0287611.

[7] N. B. Shah, "The role of author identities in peer review," *PLOS ONE*, vol. 18, no. 6, e0286206, Jun. 2023. DOI: 10.1371/journal.pone.0286206.

[8] J. Huber et al., "Nobel and novice: Author prominence affects peer review," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 119, no. 41, e2205779119, 2022. DOI: 10.1073/pnas.2205779119.

[9] E. Frachtenberg and K. McConville, "Metrics and methods in the evaluation of prestige bias in peer review: A case study in computer systems conferences," *PLoS ONE*, vol. 17, e0264131, 2022. DOI: 10.1371/journal.pone.0264131.

[10] C. J. Lee, C. R. Sugimoto, G. Zhang, and B. Cronin, "Bias in peer review," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 1, pp. 2–17, 2013. DOI: 10.1002/asi.22784.

[11] C. L. Goues et al., "Effectiveness of anonymization in double-blind review," *Communications of the ACM*, vol. 61, pp. 30–33, 2017. DOI: 10.1145/3208157.

[12] J. Bobadilla, R. Lara-Cabrera, A. González-Prieto, and F. Ortega, "DeepFair: Deep learning for improving fairness in recommender systems," *Information Processing & Management*, vol. 58, no. 3, p. 102 547, May 2021. DOI: 10.1016/j.ipm.2021.102547.

[13] Y. Peng, X. Qian, and W. Song, "A re-ranking approach for two-sided fairness on recommendation systems," in *Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things*, May 2023, pp. 312–316. DOI: 10.1145/3603781.3603836.

[14] M. Morik, A. Singh, J. Hong, and T. Joachims, "Controlling fairness and bias in dynamic learning-to-rank," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM)*, 2020, pp. 267–276. DOI: 10.1145/3340531.3412875.

[15] A. Beutel, J. Chen, Z. Zhao, and E. H. Chi, *Data decisions and theoretical implications when adversarially learning fair representations*, FAT/ML 2017 Workshop, Accessed: 2025-12-01, 2017.

[16] S. Yao and B. Huang, "Beyond parity: Fairness objectives for collaborative filtering," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[17] S. Abimbola, M. Pai, P. Nguyen, T. Greenhalgh, and E. H. Shadid, "How to address the geographical bias in academic publishing," *BMJ Global Health*, vol. 8, no. Suppl 2, e013111, Dec. 2023. DOI: 10.1136/bmjgh-2023-013111.

[18] W. M. Williams and S. J. Ceci, "National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 17, pp. 5360–5365, Apr. 2015. DOI: 10.1073/pnas.1418878112.

[19] T. Hagendorff and K. Meding, "Ethical considerations and statistical analysis of industry involvement in machine learning research," *AI & SOCIETY*, vol. 38, pp. 35–45, 2023. DOI: 10.1007/s00146-021-01284-z.

[20] M. Bosten and B. Kleinberg, *Conflicts of interest in published nlp research 2000–2024*, ArXiv, vol. abs/2502.16218, Accessed: 2025-12-01, 2025. DOI: 10.48550/arXiv.2502.16218.

[21] I. Kotseruba, M. Papagelis, and J. K. Tsotsos, *Industry and academic research in computer vision*, ArXiv, vol. abs/2107.04902, Accessed: 2025-12-01, 2021. DOI: 10.48550/arXiv.2107.04902.

[22] L. Liang, H. Zhuang, J. Zou, and D. E. Acuna, *The complementary contributions of academia and industry to AI research*, ArXiv, vol. abs/2401.10268, Accessed: 2025-12-01, 2024. DOI: 10.48550/arXiv.2401.10268.

[23] M. Färber and L. Tampakis, "Analyzing the impact of companies on AI research based on publications," *Scientometrics*, vol. 129, pp. 31–63, 2024. DOI: 10.1007/s11192-023-04867-3.

[24] J. L. Ortega, "Influence of co-authorship networks in the research impact: Ego network analyses from Microsoft Academic Search," *Journal of Informetrics*, vol. 8, no. 3, pp. 728–737, 2014. DOI: 10.1016/j.joi.2014.07.001.

[25] M. Y. Vardi, "Conferences vs. journals in computing research," *Communications of the ACM*, vol. 52, no. 5, p. 5, 2009. DOI: 10.1145/1506409.1506410.

[26] B. Meyer, *Reversing the fossilization of computer science conferences*, Communications of the ACM, BLOG@CACM, Accessed: 2025-12-01, Apr. 2025.

[27] J. M. Fernandes, A. Costa, and P. Cortez, "Author placement in computer science: A study based on the careers of ACM fellows," *Scientometrics*, vol. 127, no. 1, pp. 351–368, Jan. 2022. DOI: 10.1007/s11192-021-04035-5.

[28] Academia Stack Exchange, *What does authorship ordering really mean in field x?* Stack Exchange Q&A, Accessed: 2025-12-01, 2012.

[29] Samwell.ai, *What is a good h-index? practical guide for researchers*, Blog post, Accessed: 2025-12-01, Apr. 2025.

[30] Paperpile, *What is a good h-index? [with examples]*, Guide, Accessed: 2025-12-01, 2025.

[31] R. Burke, "Multisided fairness for recommendation," in *Proceedings of the ACM RecSys '17 Workshop on Responsible Recommendation*, Como, Italy: ACM, Aug. 2017, pp. 1–4. DOI: 10.1145/3109859.3109962.

[32] 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)*, Online: SCITEPRESS, Oct. 2021, pp. 199–208. DOI: 10.5220/0010655800003064.

[33] K. Kafle, *Four trends at CVPR 2024*, Adobe Research News, Accessed: 2025-12-01, Jul. 2024.

[34] R. Islam, K. N. Keya, Z. Zeng, S. Pan, and J. Foulds, "Neural Fair Collaborative Filtering," in *Proceedings of the 15th ACM Conference on Recommender Systems (RecSys '21)*, Amsterdam, Netherlands: ACM, Sep. 2021, pp. 148–159. DOI: 10.1145/3460231.3474603.

[35] H. Wu, C. Ma, B. Mitra, F. Diaz, and X. Liu, "A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation," *ACM Transactions on Information Systems (TOIS)*, vol. 41, no. 2, 47:1–47:29, 2022. DOI: 10.1145/3564285.

[36] 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, 52:1–52:43, Feb. 2023. DOI: 10.1145/3547333.

[37] B. Bulut, B. Kaya, R. Alhajj, and M. Kaya, "A Paper Recommendation System Based on User's Research Interests," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug. 2018, pp. 911–915. DOI: 10.1109/ASONAM.2018.8508313.