

Assessing Algorithmic Bias in Two-Sided E-Commerce Marketplaces: Implications of Search, Ranking, and Recommendation Systems

Omar Rashdan^a, Yazan Khatib^b

Abstract: Two-sided e-commerce marketplaces increasingly rely on search, ranking, and recommendation systems to allocate attention, match buyers to sellers, and mediate competition among listings. These systems operate under strong feedback, since user interactions both reflect and reshape what is shown, purchased, and reviewed. Concerns about algorithmic bias in this context are not limited to interpersonal fairness; they also involve market power, quality discovery, seller entry incentives, and the distribution of consumer surplus. This paper develops a technical framework for assessing algorithmic bias in marketplace retrieval and recommendation pipelines, emphasizing measurable exposure and welfare outcomes under realistic behavioral and platform constraints. A unified model is presented for query-conditioned ranking, recommender-driven discovery, and blended result pages, allowing bias to be decomposed into data, model, objective, and interface components. The study formalizes group-conditional disparities in exposure, click propensity, conversion, and long-run seller viability, and shows how these disparities can arise even when relevance predictions are well calibrated on average. Statistical testing and uncertainty quantification are integrated with counterfactual evaluation methods that correct for position bias, missing-not-at-random logging, and interference across listings. To address dynamic effects, the paper introduces a numerical modeling layer that treats exposure as a propagating signal over the marketplace graph, enabling spectral diagnostics and PDE-inspired simulation with finite element discretizations. Finally, mitigation is treated as constrained optimization over ranking and recommendation policies, with attention to stability, robustness, and operational governance in production systems.

Copyright © Morphpublishing Ltd.

^a Department of Computer Science and Engineering, Al al-Bayt University, 12 King Abdullah II Street, Al-Mafraq 25113, Jordan ^b Department of Computer Science and Engineering, Tafila Technical University, 7 Prince Hasan Road, Al-Aqsa District, Tafila 66110, Jordan

This is an open-access article published by MorphPublishing Ltd. under a Creative Commons license. MorphPublishing Ltd. is not responsible for the views and opinions expressed in this publication, which are solely those of the authors.

1. Introduction

Two-sided e-commerce marketplaces coordinate interactions between heterogeneous buyers and heterogeneous sellers through a shared interface that surfaces an extremely small subset of available items at any moment [1]. Search, ranking, and recommendation components jointly determine which listings are discovered, which sellers accumulate reviews, and which products receive the behavioral signals that train subsequent models. The resulting allocation of attention is not merely an informational convenience; it functions as an economic mechanism that assigns scarce screen real estate, shapes competition, and converts platform design choices into distributional outcomes. In this setting, algorithmic bias is best treated as a family of measurable deviations between an intended allocation rule and the realized allocation induced by a learned policy interacting with users, sellers, and marketplace constraints. The relevant notion of bias depends on what is being allocated and who is being compared, and it can involve sellers grouped by protected attributes, sellers grouped by business type or geography, items grouped by price tier or novelty, or buyers grouped by demographic or preference clusters. Because the marketplace is two-sided, disparities can emerge on either side and propagate across sides through demand responses, seller investment, and platform revenue optimization.

Assessment is complicated by the fact that ranking and recommendation are typically trained on observational data generated by earlier versions of the same system. Logs encode position bias, selection bias, and strategic seller behavior; labels are sparse, delayed, and confounded by what was shown; and many relevant outcomes are counterfactual, such as what would have happened if an item had been placed higher, or if a seller had been recommended to different buyers. In addition, blended interfaces create entanglement across modules, for example when sponsored results, organic search results, and personalized recommendations share a page and compete for clicks. Even if each module is individually audited, the combined pipeline can produce emergent disparities. From a measurement perspective, bias can be defined on exposures, clicks, conversions, revenue, returns, complaint rates, or longer-run outcomes such as seller exit and buyer retention, each of which reflects different ethical and economic commitments [2]. From a modeling perspective, bias can appear in the score function, the loss function, the sampling scheme, the constraints imposed by retrieval, and the heuristics used for diversification and content moderation.

A technical approach to marketplace bias therefore needs several ingredients. It needs a formal system model that makes explicit the stochastic relationship between a policy and observed data, including the role of exploration, logging, and delayed feedback. It needs metrics that separate differences in opportunity from differences in realized performance, while accounting for heterogeneous relevance, price, and fulfillment quality. It needs statistical inference procedures that yield calibrated uncertainty under heavy-tailed outcomes, multiple testing, and nonstationarity. It needs counterfactual evaluation methods that can estimate disparities under alternative policies while correcting for logging bias and handling interference across items. Finally, it needs mitigation strategies that are deployable under latency and business constraints, and that remain stable under dynamic feedback. This paper develops these ingredients in a unified mathematical framework designed for two-sided marketplaces, with an emphasis on search, ranking, and recommendation systems.

The central object is an exposure allocation induced by a policy that maps a buyer context and a candidate set to an ordered slate. Exposure is the primary scarce resource in the interface, and it mediates downstream outcomes such as click probability, purchase probability, and the accumulation of reviews that influence future ranking. A seller's long-run welfare depends on exposure over time, not solely on instantaneous relevance [3]. Conversely, buyer welfare depends on discovering items that match preferences under constraints such as shipping time, trust, and price. Marketplaces often optimize a proxy objective combining predicted engagement, predicted conversion, predicted revenue, and risk or quality penalties. The proxy is only partially aligned with welfare, and the

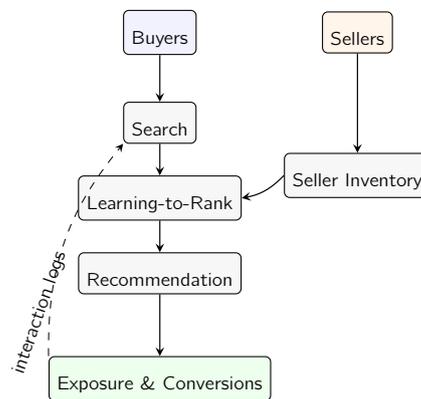


Figure 1. High-level view of a two-sided marketplace where buyers interact with search, learning-to-rank, and recommendation modules, and sellers supply inventory that competes for exposure and conversions. Logged interactions are recycled into the training data for subsequent model updates, creating feedback loops that can either intensify or mitigate existing exposure disparities.

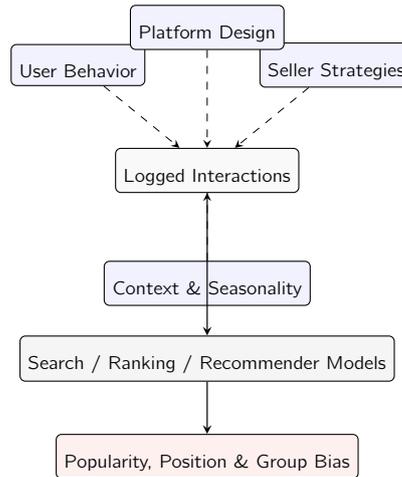


Figure 2. Sources of algorithmic bias in two-sided marketplaces. User behavior, seller strategies, platform design choices, and contextual dynamics jointly shape the logged data used to train search, ranking, and recommendation models. These models can then propagate and amplify skewed interaction patterns, leading to popularity bias, position bias, and group-level disparities in exposure and relevance.

optimization can amplify biases if the proxy is systematically mismeasured for some groups, or if the platform's constraints allocate exploration unevenly.

To connect these pieces, the paper models the marketplace as a bipartite graph between buyers and items, with sellers as item owners and with latent features influencing matching. Ranking and recommendation are treated as policies on this graph, producing flows of exposure. The analysis uses tools from statistical modeling, causal inference, numerical methods, and spectral analysis. Frequency-domain diagnostics are used to detect periodicities and instabilities in exposure allocation across groups, while logarithmic and decibel-style scales are used to represent multiplicative disparities in attention. A PDE-inspired exposure propagation model is introduced to capture dynamic

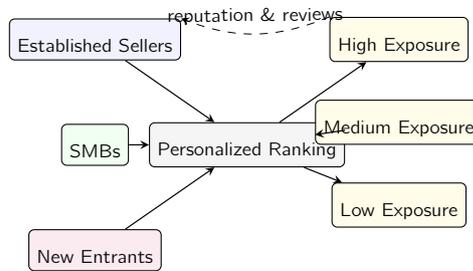


Figure 3. Group-level exposure in marketplace search and recommendation. Different seller segments feed into the ranking system with heterogeneous priors (driven by reputation, reviews, and historical performance), which can produce stratified exposure levels. Feedback from highly exposed segments further strengthens their standing, making it challenging to ensure equitable visibility for small or new sellers without explicit fairness-aware interventions.

Bias type	Definition	Marketplace example	Affected side
Position bias	Higher-ranked items receive disproportionately more attention	Users click mostly on items in the top 3 positions	Buyers and sellers
Popularity bias	Already-popular items are promoted frequently	Best-selling brands dominate search results	Primarily small sellers
Presentation bias	Visual or textual framing skews user choices	Sponsored tags blend with organic listings	Buyers and competing sellers
Price bias	Model overweights price signals in ranking	Low-price items consistently outrank higher-quality items	Quality-focused sellers
Geographic bias	Location features drive exposure differences	Urban sellers surface more than rural ones	Rural and remote sellers
Feedback-loop bias	Historic clicks reinforce existing patterns	Early winners keep getting exposure, others stagnate	New and niche sellers

Table 1. Typology of algorithmic biases in two-sided e-commerce marketplaces.

feedback and spillovers, and finite element approximations are used to solve the resulting equations on an irregular marketplace graph. Mitigation is treated as constrained optimization that trades off relevance and fairness under robust uncertainty, with operational considerations such as monitoring, governance, and auditability.

2. Two-Sided Marketplace and Algorithmic Pipeline Model

Consider a set of buyers indexed by $u \in \mathcal{U}$, a set of items indexed by $i \in \mathcal{I}$, and a set of sellers indexed by $s \in \mathcal{S}$, where each item belongs to a seller via a mapping $\sigma : \mathcal{I} \rightarrow \mathcal{S}$. Buyers arrive over time t with contexts $x_u(t)$ that include query text, session history, device, location, and inferred preferences. Items have features $x_i(t)$ capturing content, price, inventory, shipping promises, seller reputation, and policy compliance [4]. The system constructs a candidate set $\mathcal{C}(u, t) \subset \mathcal{I}$ using retrieval and filtering, then produces an ordered slate (or multiple blended slates) via a policy π_θ parameterized by θ . The displayed list induces exposures $e_{u,i}(t)$, which are random variables shaped by

Pipeline stage	Description	Potential bias source	Typical impact
Data collection	Logging user interactions and attributes	Skewed traffic, missing data for small sellers	Underrepresentation of minority items
Feature engineering	Transforming raw data into model inputs	Proxy variables for sensitive attributes	Indirect discrimination across groups
Model training	Learning relevance or ranking functions	Loss functions ignoring fairness constraints	Exposure imbalance by seller group
Online ranking	Real-time ordering of results	Business rules, ads, promotions	Systematic uplift for sponsored content
Recommendation	Personalized suggestions to users	Collaborative filtering on biased histories	Filter bubbles for buyers and sellers
Monitoring	Continuous performance tracking	Narrow metrics centered on revenue	Unnoticed long-term fairness degradation

Table 2. Key stages of the search, ranking, and recommendation pipeline and their bias risks.

Metric	Intuition	Target entity	Fairness objective
Exposure parity	Similar average visibility across groups	Seller groups	Align share of impressions with inventory share
Click-through parity	Comparable CTR across groups at similar relevance	Items or sellers	Avoid systematically lower engagement for a group
Conversion parity	Similar purchase likelihood conditional on relevance	Sellers	Ensure profitable demand for disadvantaged groups
Rank-weighted share	Fair share of top positions per group	Seller groups	Prevent dominance of a single group at the top
Calibration by group	Predicted scores match observed outcomes per group	Items or users	Maintain reliability of scores across demographics

Table 3. Representative fairness metrics for two-sided marketplace ranking and recommendation.

Evaluation setting	Advantages	Limitations	Typical metrics
Offline replay	Cheap, repeatable, no user risk	Reflects historic policy and logging bias	NDCG, MAP, exposure parity
A/B testing	Measures real user response	Costly, ethical risk for harmed groups	CTR, conversion rate, seller revenue
Interleaving	Fine-grained comparison of two rankers	Hard to scale across many variants	Win-rate, click preference
Simulation	Counterfactual and long-term analysis	Requires strong behavioral assumptions	Long-run exposure, entry and exit rates

Table 4. Comparison of evaluation strategies for assessing algorithmic bias in marketplaces.

viewport geometry, pagination, and interface design. A common abstraction is to define exposure as the probability

Seller group	Share of listings	Avg. rank position	Relative CTR
Large / professional	25%	8.2	1.20
Small businesses	40%	13.5	0.95
New entrants	15%	17.8	0.70
Minority-owned	10%	18.3	0.68
Niche / long-tail	10%	19.1	0.65

Table 5. Illustrative distribution of exposure and engagement across seller groups.

Mitigation strategy	Mechanism	Main trade-off	Implementation effort
Fairness-aware ranking	re-Adjust scores to enforce exposure constraints	Possible relevance or revenue loss	Medium
Diversity promotion	Encourage result variety within a page	Lower short-term clicks	Low–medium
Data reweighting	Upweight underrepresented groups in training	Increased variance, model instability	Medium
Regularization terms	Add group fairness to the loss function	Complex hyperparameter tuning	High
Transparency tools	Explain exposure and ranking outcomes	Possible gaming by strategic sellers	Medium

Table 6. Common mitigation strategies for reducing ranking and recommendation bias.

Stakeholder	Primary objective	Responsibility	Governance tools
Platform operator	Growth, trust, compliance	Design and audit algorithms	Fairness dashboards, internal audits
Sellers	Market access, stable demand	Provide accurate product data	Feedback channels, appeal mechanisms
Buyers	Quality, price, choice	Responsible use of platform	Preference controls, transparency reports
Regulators	Competition, consumer protection	Set and enforce rules	Reporting requirements, audits, sanctions
Civil society	Equity, accountability	Independent oversight	External studies, standards, advocacy

Table 7. Roles and responsibilities in governing bias in two-sided marketplaces.

of being observed at least once in a session, or as expected view time, both of which are functions of rank and UI features. Let k denote a position index and let a_k denote a position attention weight with $a_1 \geq a_2 \geq \dots \geq 0$. If item i is placed at position k for buyer u at time t , a baseline exposure model is $e_{u,i}(t) = a_k$, while a more realistic model introduces stochastic scrolling and heterogeneous attention as $e_{u,i}(t) = a_k \cdot \eta_u(t)$ where $\eta_u(t)$ captures session-specific viewing depth and may depend on query intent and trust.

Observed feedback includes clicks $c_{u,i}(t) \in \{0, 1\}$, purchases $p_{u,i}(t) \in \{0, 1\}$, revenue $r_{u,i}(t) \geq 0$, and post-purchase outcomes such as returns or complaints. A standard cascade structure assumes $c_{u,i}(t)$ is conditionally

Hypothesis	Description	Outcome metric	Expected direction
H1	Popularity bias increases exposure inequality across sellers	Gini coefficient of exposure	Positive
H2	Fairness-aware re-ranking reduces inequality with limited CTR loss	Exposure Gini, CTR	Exposure ↓, CTR ≈
H3	New sellers are systematically under-exposed at launch	Avg. rank of new vs. existing sellers	New sellers ranked lower
H4	Sponsored content amplifies existing exposure disparities	Share of top-10 positions by group	Higher share for advantaged groups
H5	Diversity constraints improve long-term marketplace participation	Seller survival rate	Higher for disadvantaged groups

Table 8. Example hypotheses for empirical analysis of algorithmic bias in marketplaces.

Variable	Type	Description	Level
Rank position	Integer	Displayed position in the result list	Item-impression
Click indicator	Binary	1 if the item was clicked, 0 otherwise	Item-impression
Conversion indicator	Binary	1 if a purchase occurred	Session-item
Seller group	Categorical	Group membership (e.g., size, ownership)	Seller
Historical popularity	Integer	Past clicks or sales count	Item
User segment	Categorical	Coarse demographic or behavioral segment	User

Table 9. Illustrative variables for analyzing search, ranking, and recommendation outcomes.

independent given exposure and latent attractiveness, and $p_{u,i}(t)$ depends on click and latent match quality. A flexible formulation is to define latent utilities $v_{u,i}(t)$ and to write

$$\Pr(c_{u,i}(t) = 1 \mid e_{u,i}(t), x_u(t), x_i(t)) = \sigma(\alpha e_{u,i}(t) + g_\theta(x_u(t), x_i(t))),$$

$$\Pr(p_{u,i}(t) = 1 \mid c_{u,i}(t) = 1, x_u(t), x_i(t)) = \sigma(h_\theta(x_u(t), x_i(t))),$$

where $\sigma(z) = (1 + \exp(-z))^{-1}$, and g_θ and h_θ are learned predictors. This separates a mechanical exposure effect from a relevance effect, but in practice the separation is partial because $e_{u,i}(t)$ is itself induced by the policy and correlates with unobserved match quality through ranking. For recommendation surfaces without explicit queries, $x_u(t)$ includes a representation of user intent inferred from history; for search, it includes a query embedding. In blended interfaces, multiple policies produce multiple candidate pools and a compositor chooses a final ordering; this can be modeled by a meta-policy acting on module outputs with module-specific attention weights.

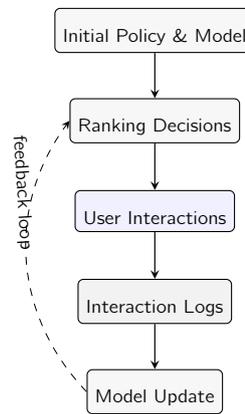


Figure 4. Feedback dynamics induced by search, ranking, and recommendation systems. Deployed models determine ranking decisions that shape user interactions and click behavior, which in turn populate the interaction logs used for subsequent model updates. Without careful monitoring and correction, this loop can entrench initial biases in exposure and conversion across users and sellers.

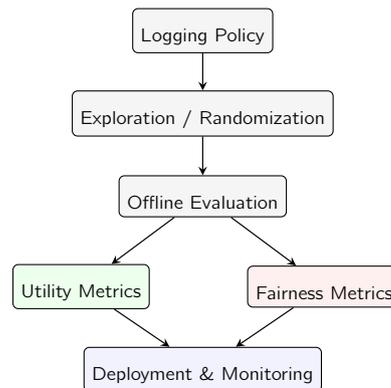


Figure 5. Evaluation and auditing pipeline for marketplace algorithms. A suitably randomized logging policy enables counterfactual or off-policy evaluation of new search, ranking, and recommendation policies. Offline experiments feed into both utility-focused and fairness-focused metrics, which jointly inform deployment decisions and continuous monitoring of marketplace outcomes.

Seller-side outcomes depend on aggregated exposure and demand. Let $E_s(T) = \sum_{t \leq T} \sum_u \sum_{i: \sigma(i)=s} e_{u,i}(t)$ denote cumulative exposure up to time T , and similarly define cumulative clicks $C_s(T)$, purchases $P_s(T)$, and revenue $R_s(T)$. Sellers also update their behavior in response to these signals by changing price, improving content, altering fulfillment options, or exiting the platform. A minimal dynamic model introduces seller state variables $z_s(t)$ that evolve according to

$$z_s(t+1) = z_s(t) + \Delta_s(E_s(t), R_s(t), \xi_s(t)),$$

where $\xi_s(t)$ captures exogenous shocks and constraints. This endogeneity implies that bias assessment cannot be purely myopic: disparities in exposure can alter seller states, which then affect future relevance and quality, forming feedback loops. On the buyer side, utility and retention evolve based on realized matches, which changes future traffic composition [5]. This two-sided feedback complicates causal identification because the system is not operating on a fixed population.

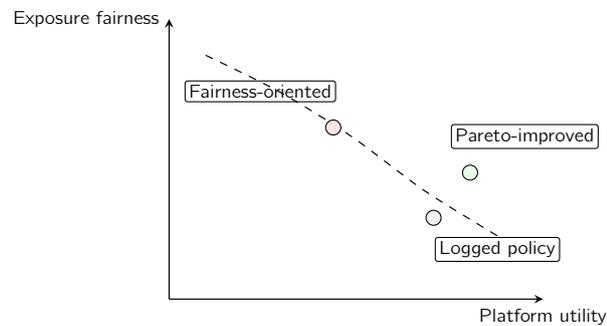


Figure 6. Illustrative fairness–utility frontier for ranking and recommendation policies in a two-sided marketplace. The logged policy may lie inside the frontier, allowing simultaneous gains in platform utility and exposure-based fairness. Moving along the frontier reveals trade-offs between utility and equity, guiding the selection of acceptable operating points for search, ranking, and recommendation systems.

Grouping is central to bias definitions. Let G_s denote a seller group label and G_u a buyer group label. Group labels may be observed, partially observed, or inferred with error. For sellers, groupings can include protected attributes when legally permissible to use for auditing, or proxies such as region or business size. For items, groups can be derived from seller group, category, price tier, or novelty (for example, cold-start listings). For buyers, groupings can include geography or inferred preferences, and these can matter because personalization changes the distribution of exposures across buyers. A key subtlety in marketplaces is that relevance itself can differ across groups due to product mix, quality, and price, so equalized outcome constraints must be interpreted carefully. Metrics based on exposure attempt to characterize opportunity, while metrics based on conversion attempt to characterize realized benefit under demand. Both can be relevant, but they answer different questions.

The training pipeline typically uses logged data $\mathcal{D} = \{(x_u(t), x_i(t), e_{u,i}(t), c_{u,i}(t), p_{u,i}(t), r_{u,i}(t))\}$ to fit models for g_θ and h_θ and to learn a ranking score $s_\theta(u, i, t)$. The ranking policy may be a deterministic sort by score, or a stochastic policy introducing exploration [6]. Bias can be introduced by non-representative data, by label noise that differs by group, by objective functions that overweight short-term revenue, or by constraints such as minimum seller quality thresholds that are correlated with group membership due to historical disparities. Retrieval also matters: if the candidate set $\mathcal{C}(u, t)$ systematically omits certain items, then no ranking-stage fairness constraint can recover exposure for those items. Because retrieval is often based on approximate nearest neighbor search and learned embeddings, it can embed biases via representation collapse or domain mismatch.

Operationally, the platform may impose additional post-processing constraints such as diversity, content policy filtering, deduplication, and sponsored placements. These are best modeled as a feasible set Π of allowable slates, where $\pi_\theta(u, t) \in \Pi$ and Π depends on business rules. Fairness mitigation then becomes a constrained policy design problem: choose π within operational constraints to satisfy relevance and welfare objectives while bounding disparity metrics. The next sections formalize bias mechanisms, measurement, and counterfactual evaluation, and then develop numerical methods for dynamic analysis and optimization.

3. Bias Mechanisms in Search, Ranking, and Recommendation

Bias in marketplace ranking is often discussed as if it were a property of a model’s predictions, but in a two-sided setting it is more accurately a property of a closed-loop system. The same scoring function can produce

different disparities depending on the interface, the exploration strategy, the candidate generator, and the adaptive behavior of sellers and buyers. Several broad mechanisms are important. Data bias arises when training logs reflect historical exposure disparities, such that some sellers or items are underrepresented at high ranks and therefore contribute fewer positive labels. This is not merely a sampling issue; it changes the conditional distribution of labels given features, because labels are collected under an exposure process that is itself biased [7]. Label bias arises when clicks or purchases are imperfect proxies for relevance and differ systematically by group, for instance due to trust differentials, shipping constraints, or buyer priors about seller types. Objective bias arises when the platform optimizes a proxy such as predicted revenue that can trade off against buyer satisfaction or seller opportunity, and when the proxy's errors differ by group. Interface bias arises when the UI amplifies attention differences across ranks, such that small score differences produce large exposure differences.

Feedback dynamics can amplify small disparities into durable market structure effects. Consider a cold-start seller group with fewer historical reviews. If the ranking model uses review count as a feature, even implicitly through representations, then the group may be ranked lower, leading to fewer sales, which delays review accumulation, which sustains the lower rank. This is a positive feedback loop with a fixed point that can be inequitable even if initial differences were small. A simplified continuous-time representation treats a seller's reputation $q_s(t)$ as a state variable increasing with sales and decaying with time, such as $\dot{q}_s(t) = \beta P_s(t) - \gamma q_s(t)$. If ranking scores depend on $q_s(t)$, then exposure depends on $q_s(t)$, implying a coupled dynamical system. Nonlinearities arise because attention is concentrated in the top ranks. In such systems, stability depends on exploration and smoothing; without them, the dynamics can converge to a concentrated allocation where early random advantages lock in.

Strategic behavior by sellers introduces another mechanism [8]. Sellers may optimize for the ranking system by altering titles, images, pricing, or shipping promises in ways that correlate with resources and sophistication. If resource-constrained sellers are overrepresented in some groups, then strategic adaptation can produce group-correlated performance differences that appear as "merit" to the model but reflect differential capacity to game the system. Additionally, fraud and policy violations can be more prevalent in some observable proxies (for example, in certain regions or categories), leading to stronger filtering that may unintentionally suppress legitimate sellers sharing those proxies. If enforcement models have unequal error rates, then false positives can produce disparate delisting. Even when enforcement is accurate, the downstream effect on exposure can interact with ranking, because filtered items alter the competitive set and shift demand to substitutes.

Personalization introduces bias pathways distinct from generic ranking. For recommendations, the system aims to maximize predicted engagement by matching items to user profiles. If user profiles encode historical exposure bias, then recommendations can reinforce it. Suppose a user's embedding is updated via clicked items, and the system primarily shows items from a dominant group at high rank. Then the embedding update is driven by that group's items, which can reduce the similarity between the user and items from underexposed groups, further reducing their recommendation probability. This is a representational feedback loop [9]. Moreover, recommendation objectives can induce homophily: showing users more of what they previously consumed. If groups differ in initial exposure, homophily can produce persistent separation in consumption patterns, which can be undesirable if it reduces discovery or if it concentrates opportunity.

In search, query understanding and retrieval can embed linguistic and categorical biases. Query embeddings trained on click data can learn associations between query terms and seller or product attributes that reflect societal biases or platform history. For example, ambiguous queries can be resolved in ways that favor dominant categories because they received more clicks historically, even if minority categories are relevant. Retrieval bias can be expressed as a group-conditioned recall disparity. Let $\mathcal{R}(u, t)$ denote the set of truly relevant items under a latent relevance oracle, and let $\mathcal{C}(u, t)$ be the retrieved set. A retrieval stage is group-biased if $\Pr(i \in \mathcal{C}(u, t) \mid i \in \mathcal{R}(u, t), G_{\sigma(i)} = g)$ differs

substantially across g , after conditioning on category and other legitimate factors. Because downstream ranking only acts on $\mathcal{C}(u, t)$, retrieval bias creates hard exposure ceilings.

Blended ranking surfaces further complicate bias. Sponsored placements can displace organic results, and if the ability to bid for sponsored slots differs by seller group, the exposure distribution can shift. Even if sponsored slots are clearly labeled, users may click them due to their position, affecting training logs and organic ranking. The system can be modeled as allocating exposure among modules, where the exposure share to module m is $\omega_m(u, t)$ and depends on user context and monetization rules. Group disparities can arise from differences in module eligibility [10]. For example, a recommendation module might use fulfillment quality thresholds that exclude some sellers, while search results include them. The overall exposure disparity is then a mixture effect across modules.

A useful way to conceptualize these mechanisms is to separate three levels: the scoring level, where a model produces predictions; the policy level, where predictions are mapped to slates under constraints; and the ecosystem level, where repeated interactions change the data distribution. Bias can exist at any level and can shift between levels during mitigation. For instance, enforcing exposure parity at the policy level can change the data distribution and eventually reduce scoring bias, but it can also create short-term relevance trade-offs that affect buyer trust. Conversely, improving calibration of predicted conversion can reduce scoring bias but may leave exposure concentration unchanged if the policy still concentrates attention in a winner-take-most regime. A robust assessment therefore needs both static and dynamic metrics, and methods that can estimate what would happen under alternative policies.

4. Measurement, Statistical Testing, and Spectrum-Based Diagnostics

Measurement begins with defining the outcome of interest and the unit of analysis. For marketplace bias, exposure is a natural starting point because it is directly controlled by ranking and recommendation policies. Let E_i denote the expected exposure of item i over a time window, defined as $E_i = \mathbb{E}[\sum_t e_{u_t, i}(t)]$ where u_t is the buyer arriving at time t and the expectation is over traffic and policy randomness. For a seller group g , define group exposure as $E_g = \sum_{i:G_{\sigma(i)}=g} E_i$. Since groups differ in inventory size and item relevance distributions, normalized metrics are often needed. One normalization is per-item exposure $\bar{E}_g = E_g/N_g$ where N_g is the number of active items in group g . Another is relevance-conditioned exposure, which compares exposure given predicted relevance [11]. If $\hat{y}_{u, i}(t)$ is a predicted relevance or conversion probability, then an opportunity metric can be defined as the ratio between realized exposure and expected exposure under a relevance-proportional allocation:

$$\text{Opp}_g = \frac{\sum_{i:G_{\sigma(i)}=g} E_i}{\sum_{i:G_{\sigma(i)}=g} \mathbb{E}[\sum_t \tilde{e}_{u_t, i}(t)]},$$

where \tilde{e} is computed under a counterfactual policy that allocates exposure proportionally to \hat{y} . This is not a fairness criterion by itself, but it helps isolate whether a group is systematically underexposed relative to the system's own predictions.

Disparity is inherently comparative. For two groups g and g' , a multiplicative exposure disparity can be expressed as $D_E(g, g') = \bar{E}_g/\bar{E}_{g'}$. Because multiplicative disparities are often easier to interpret on a logarithmic scale, it is useful to define a decibel-style attention disparity

$$\Delta_E^{\text{dB}}(g, g') = 10 \log_{10} \left(\frac{\bar{E}_g + \epsilon}{\bar{E}_{g'} + \epsilon} \right),$$

with a small ϵ to stabilize rare groups. A negative decibel value indicates lower exposure for g relative to g' , and additive differences in dB correspond to multiplicative ratios, which is helpful when exposure spans orders of

magnitude across sellers. Similar logarithmic transforms can be applied to revenue or conversion, with care for heavy tails.

Outcome-based metrics such as click-through rate, conversion rate, and revenue per exposure capture realized performance. For example, exposure-normalized conversion can be defined as $CVR_g = P_g/E_g$ where P_g is total purchases for group g . Differences in CVR_g can reflect true quality differences, but they can also reflect mismatched targeting, trust differences, or measurement bias. To avoid conflating opportunity with performance, a common approach is to measure both exposure disparity and conditional performance disparity, such as conversion conditional on comparable predicted relevance and price. A regression-based adjustment can model conversion as a function of covariates and group membership:

$$\log(\mathbb{E}[p_{u,i}(t) | x_u(t), x_i(t), G_{\sigma(i)}]) = \beta^\top \phi(x_u(t), x_i(t)) + \gamma_{G_{\sigma(i)}},$$

where ϕ are feature transformations and γ_g are group effects. Inference on γ_g provides an adjusted disparity estimate, but it depends on the adequacy of ϕ and on whether unobserved confounders remain [12].

Descriptive statistics should report not only means but also distributional shape, because attention and revenue are typically heavy-tailed. Quantiles, Lorenz curves, and Gini coefficients are relevant for concentration. If E_s denotes seller-level exposure, then the Gini coefficient can quantify exposure inequality across sellers, and group-conditional Gini values can reveal whether inequality is more severe in particular groups. Heavy tails also motivate robust estimation. For example, winsorized means or trimmed means of seller exposure can reduce sensitivity to a few dominant sellers. However, trimming can hide meaningful concentration, so it is best paired with explicit tail metrics such as the share of exposure captured by the top mass of sellers.

Statistical testing must account for non-independence and for multiple comparisons. Marketplace logs have clustering by user session and by seller. A naive test comparing group means of exposure may underestimate variance because exposures share traffic shocks and policy changes. A cluster-robust approach can treat days or sessions as clusters and compute sandwich estimators. If $Y_{d,g}$ is the daily exposure for group g on day d , then a difference-in-means test can be performed on the time series $\{Y_{d,g} - Y_{d,g'}\}$ with heteroskedasticity and autocorrelation corrections, because platform traffic and seasonal effects induce autocorrelation. When testing many groups and many metrics, false discovery control becomes important; practical audit systems often prefer effect size reporting with confidence intervals over binary significance decisions [13].

A major measurement challenge is position bias and selective logging. Clicks and purchases are observed only for exposed items, and exposure itself depends on the policy. In search logs, items below the fold may have uncertain exposure, and measuring $e_{u,i}(t)$ may require instrumenting scroll depth. If exposure is approximated by position weights a_k , then mis-specification of a_k can bias disparity estimates. A more empirical approach estimates a_k from randomized swaps or interleaving experiments. This can be framed as estimating the examination probability $q_k = \Pr(\text{examined} | \text{position} = k)$, then defining exposure as $e_{u,i}(t) = q_{k(i)}$. Group disparities can then be decomposed into disparities in slot assignments and disparities in examination due to UI interactions that differ by user group.

Frequency spectrum analysis provides an additional diagnostic for dynamic bias and instability. Exposure disparities often vary over time due to promotions, seasonal demand, model retraining, and seller campaigns. Let $d(t)$ be a time series of a disparity metric, such as $\Delta_E^{\text{dB}}(g, g')$ computed daily. The discrete Fourier transform

$$\hat{d}(\omega) = \sum_{t=0}^{T-1} d(t) \exp(-2\pi i \omega t / T)$$

reveals periodic components. Peaks at weekly frequencies can indicate that a group is disproportionately affected by weekend traffic, while peaks at retraining cadence frequencies can indicate model-induced oscillations [14]. Spectral density estimates can be computed using windowed periodograms to reduce leakage. While frequency-domain analysis does not identify causal mechanisms, it can flag systematic cycles that deserve investigation. In addition, cross-spectral analysis between disparity metrics and platform interventions, such as promotion schedules or policy enforcement surges, can reveal coupling. If $x(t)$ is an intervention indicator series and $d(t)$ is disparity, the coherence function between x and d quantifies frequency-specific correlation.

An analogy to sound propagation can be useful for interpreting exposure signals. In acoustics, energy spreads through a medium with attenuation and reflections; in marketplaces, exposure spreads through recommendation graphs with attenuation due to limited attention and reflections due to feedback and revisitation. Using decibel scales aligns with this analogy because both acoustic intensity and marketplace exposure can span orders of magnitude. However, the analogy is only a modeling device; it should not be mistaken for a claim of physical equivalence. The next section develops counterfactual methods that move beyond descriptive measurement toward estimates of what disparity would be under alternative policies.

5. Counterfactual Evaluation and Causal Identification

Counterfactual evaluation asks how outcomes would change if the marketplace used a different search, ranking, or recommendation policy. For bias assessment, the core counterfactual is whether a disparity metric would shrink or grow under an alternative policy, such as a fairness-constrained reranker or a different exploration scheme [15]. The difficulty is that logs are collected under a historical policy π_0 , and many outcomes are only observed for exposed items. Off-policy evaluation addresses this by reweighting observed data to approximate the distribution under a target policy π_1 . A standard approach uses inverse propensity scoring (IPS). If the logging system records the probability that an item was placed at a position under π_0 , then the expected value of an outcome under π_1 can be estimated by weighting observed outcomes by the ratio of target to logging propensities. In slate ranking, this ratio can be high-variance because it involves a combinatorial action. Practical implementations use marginal propensities for positions or use structured estimators based on swap randomization.

Let A_t denote the displayed slate at time t and let Y_t be an outcome such as total purchases in that session. Under assumptions of no unmeasured confounding beyond the logged propensities, an IPS estimator is

$$\widehat{\mathbb{E}}_{\pi_1}[Y] = \frac{1}{T} \sum_{t=1}^T \frac{\pi_1(A_t | X_t)}{\pi_0(A_t | X_t)} Y_t,$$

where X_t is the context. For group-specific outcomes, Y_t can be replaced with a contribution from group g , such as purchases of items owned by group g . The challenge is that $\pi_0(A_t | X_t)$ may be unknown because production ranking is often deterministic [16]. To enable off-policy evaluation, platforms introduce exploration, such as random swaps among near-tied candidates, which yields known propensities. This creates a trade-off: more exploration improves estimability but can reduce short-term relevance. From a bias perspective, exploration itself can be uneven if it is applied more in some categories or to some users, so the exploration design must be audited.

Doubly robust estimators reduce variance by combining IPS with an outcome model. If $\hat{m}(X_t, A_t)$ predicts Y_t , then a doubly robust estimator is

$$\widehat{\mathbb{E}}_{\pi_1}[Y] = \frac{1}{T} \sum_{t=1}^T \left[\hat{m}(X_t, A_t^{(1)}) + \frac{\pi_1(A_t | X_t)}{\pi_0(A_t | X_t)} (Y_t - \hat{m}(X_t, A_t)) \right],$$

where $A_t^{(1)}$ is a slate sampled or constructed under π_1 . This estimator is consistent if either the propensity model or the outcome model is correct. For group disparity estimation, the outcome model should include group-conditional features to avoid systematic residual errors that could bias disparity estimates.

Causal identification is further complicated by interference, because the outcome for a given seller depends on what other sellers are shown. If the policy boosts one group, it may reduce exposure for another group and change competition, which affects conversion rates through price and substitution effects. Standard potential outcomes assumptions of no interference are violated. A more appropriate approach models outcomes as functions of the entire slate, not just an individual item, and defines causal effects of policy changes at the slate or market level. This increases dimensionality [17]. One tractable approximation uses exposure as the intermediate variable and models conversion as a function of exposure shares. For seller s , define an exposure share vector across categories or competitor groups, and model P_s as depending on that vector. Causal effects are then defined in terms of changing exposure shares under a target policy. This is an approximation because exposure affects buyer choice among items, but it captures the competition-mediated nature of effects.

Randomized controlled experiments remain the most reliable approach for causal assessment, but they can be expensive and can risk adverse impacts. A fairness audit can use constrained experiments where small, targeted perturbations are applied to estimate local sensitivities. For example, randomized interleaving between two ranking variants can estimate differential exposure effects without fully switching the policy. If Z_t is a random assignment to variant and Y_t is a group disparity metric, then the average treatment effect can be estimated by comparing means across Z_t . Because disparity metrics are aggregated, power analysis should consider variance at the aggregation level. When experiments are not feasible, quasi-experimental approaches such as difference-in-differences can be used around policy changes, comparing affected and unaffected groups or categories, but these rely on parallel trends assumptions that can be fragile in marketplaces with seasonal demand.

Uncertainty quantification should reflect heavy tails and nonstationarity [18]. Bootstrapping is common, but naive bootstrap resampling of impressions can break dependence structures. A block bootstrap over days or sessions is often more appropriate. Bayesian hierarchical models can also be used to share strength across sellers while estimating group effects with partial pooling. For example, a seller-level conversion parameter can be modeled as drawn from a group-specific distribution, and posterior inference can yield uncertainty intervals for group disparities. Hierarchical modeling helps in small groups where raw estimates are noisy, but it introduces modeling assumptions that should be stress-tested.

Counterfactual assessment of long-run bias requires modeling dynamics. A short-run policy change can increase exposure for a group, but if it also increases negative feedback such as returns due to mismatch, it can harm long-run outcomes. Conversely, short-run reductions in conversion may be acceptable if they enable new sellers to accumulate reviews and improve future match quality. Dynamic causal inference can model state evolution with Markov decision processes or structural equations. Let S_t be the marketplace state summarizing seller reputations, inventory, and buyer preferences. A policy π induces transitions $S_{t+1} \sim P(\cdot | S_t, \pi(S_t))$. The long-run group welfare under π is then an expected discounted sum [19]. Estimating such quantities from logs is difficult due to compounding distribution shift. This motivates simulation-based approaches and numerical methods that approximate the dynamics of exposure and state variables, which are developed in the next section.

6. Numerical Methods, Finite Element Approximations, and Fairness-Constrained Optimization

To analyze dynamic exposure allocation and feedback, it is useful to introduce a numerical modeling layer that treats exposure as a flow on a graph. Let the marketplace be represented as a bipartite graph with adjacency weights capturing matching potential between buyers and items, or as a projected item-item graph capturing co-view and co-purchase relations. Define an exposure density $u(i, t)$ on items, representing expected exposure rate for item i at time t . Under a given policy, $u(i, t)$ is driven by traffic and by the policy's allocation rule, and it feeds back into model training signals and seller states. A simplified continuous approximation models u as evolving under a diffusion-reaction equation on a latent space in which similar items are near each other. Let Ω be a continuous embedding domain and let $x \in \Omega$ represent item coordinates. Then $u(x, t)$ can be modeled by

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(x, t)\nabla u) - \lambda(x, t)u + f(x, t),$$

where D is a diffusion tensor capturing how exposure spills over among similar items through recommendations and substitution, λ is an attenuation term capturing limited attention and churn, and f is a source term capturing incoming traffic allocated by search and recommendation. Group membership can be represented by a field $g(x)$, and group disparities correspond to integrals of u over regions associated with groups [20]. This PDE is not a literal physical law, but it provides a structured way to approximate how localized boosts or suppressions in exposure propagate through similarity structure and decay over time.

To incorporate oscillatory dynamics observed in practice, such as periodic traffic and retraining cycles, a wave-like component can be introduced. An analogy to sound propagation is a damped wave equation for a signal $w(x, t)$ representing rapid fluctuations in exposure due to promotions or viral trends:

$$\frac{\partial^2 w}{\partial t^2} + 2\zeta\omega_0 \frac{\partial w}{\partial t} - c(x)^2 \Delta w = s(x, t),$$

where $c(x)$ is a propagation speed in the embedding space, ζ is a damping ratio, ω_0 is a characteristic frequency, and s is a forcing term. While exposure itself is nonnegative and the wave equation yields signed oscillations, w can be interpreted as deviations from a baseline exposure field. Spectral analysis of w connects directly to the frequency diagnostics of disparity time series. If a platform exhibits retraining-induced oscillations in which certain groups' exposure rises and falls in a repeating pattern, the wave model provides a convenient linear approximation for studying damping and stability under mitigation.

Solving these equations on real marketplaces requires discretization on irregular domains. Finite element analysis is appropriate when the domain is complex or when coefficients vary spatially. Construct a mesh over the embedding space or, more directly, use a graph finite element approach where basis functions are defined on nodes corresponding to items or clusters. For the diffusion-reaction equation, a standard Galerkin finite element discretization yields a system [21]

$$M\dot{u} + Ku + Lu = b(t),$$

where u is now a vector of nodal exposure values, M is a mass matrix, K is a stiffness matrix derived from diffusion, L is a diagonal or sparse matrix from attenuation, and $b(t)$ is the discretized source term. Time integration can be performed using implicit schemes such as backward Euler for stability, especially when diffusion is stiff. This yields at each step a linear system $(M + \Delta t(K + L))u_{t+\Delta t} = Mu_t + \Delta t b(t + \Delta t)$. Large-scale marketplaces produce large sparse systems, motivating iterative solvers such as conjugate gradient for symmetric positive definite matrices, with preconditioning based on incomplete factorization or multigrid methods. The numerical solution provides a

smoothed estimate of how exposure evolves under an assumed policy and traffic pattern, enabling scenario analysis without fully deploying a policy.

The numerical layer becomes operationally relevant when coupled with fairness-constrained optimization. Let π denote a ranking or recommendation policy, and define an objective $J(\pi)$ representing expected utility, such as a weighted combination of predicted buyer satisfaction and revenue, possibly including risk penalties. Define disparity constraints based on exposure or welfare, such as requiring that group exposure per item satisfies $\bar{E}_g \geq \rho \bar{E}_{g'}$ for all pairs, for some ratio $\rho \in (0, 1]$, or requiring that a decibel disparity $\Delta_E^{\text{dB}}(g, g')$ remains above a threshold. Because strict parity can be infeasible when groups differ in inventory relevance distributions, constraints often condition on relevance bands. Let \mathcal{B}_ℓ be a set of items within a relevance score band, and impose constraints within each band to separate opportunity from quality. The policy design then becomes

$$\max_{\pi \in \Pi} J(\pi) \quad \text{subject to} \quad \Phi(\pi) \leq 0,$$

where $\Phi(\pi)$ encodes disparity constraints and Π encodes operational feasibility, such as latency, content policies, and diversity rules.

In practice, π is implemented as a reranker that modifies a base score. Let s_i be a base score for item i in a given context, and let δ_i be an adjustment [22]. A differentiable surrogate can treat the ranking as a softmax distribution over positions. If $P(i | u)$ is the probability of placing item i at a top position under a soft policy, one can write $P(i | u) \propto \exp(s_i + \delta_i)$ and approximate exposure as $\tilde{e}_i = \sum_k a_k P_k(i | u)$, where P_k is the probability of item i occupying position k under a Plackett–Luce style model. Disparity constraints then become differentiable functions of δ , enabling gradient-based optimization. Lagrangian methods introduce multipliers λ and solve $\max_\delta \min_\lambda \mathcal{L}(\delta, \lambda) = J(\delta) + \lambda^\top \Phi(\delta)$. Because the system is stochastic and nonconvex, stochastic gradient methods are used, with minibatches over contexts. Robustness can be improved by adding regularization on δ to limit changes and preserve relevance.

Parametric analysis examines how disparity and utility change with fairness constraint strength. Let τ parameterize a constraint, such as a maximum allowed decibel gap. Solving the constrained optimization for multiple τ values yields a Pareto frontier between utility and disparity. Sensitivity can be quantified by derivatives $dJ/d\tau$ and $d\Phi/d\tau$, estimated via implicit differentiation when the optimization converges. This analysis is important because fairness constraints can interact with market dynamics: tightening constraints can change seller behavior and future relevance distributions. A numerical simulation loop can approximate this by alternating between policy optimization and state evolution using the PDE or wave-like models. Even if the PDE model is only approximate, it can reveal qualitative behaviors such as whether exposure boosts dissipate quickly or persist due to feedback.

A key numerical challenge is that marketplace outcomes are heavy-tailed and noisy, making objective gradients unstable [23]. Variance reduction techniques from Monte Carlo estimation are relevant. When estimating group exposure under a stochastic policy, control variates can be used by subtracting a baseline exposure predictor. If \hat{E}_g is an estimator and Z is a correlated baseline with known expectation, then $\hat{E}_g^{\text{cv}} = \hat{E}_g - \alpha(Z - \mathbb{E}[Z])$ can reduce variance for a suitable α . Importance sampling weights can also be clipped to reduce variance at the cost of bias; this trade-off should be quantified with sensitivity analysis. For dynamic simulations, numerical stability matters; explicit time stepping can require small steps, while implicit methods increase computational cost but improve stability. Finite element discretizations can introduce numerical diffusion that smooths disparities, so calibration against observed time series is important.

Finally, the numerical layer can support stress testing under distribution shift. For example, a sudden category trend can change traffic sources $f(x, t)$, and a policy might respond in ways that alter group disparities. By injecting

shocks into f and solving the discretized dynamics, one can approximate how quickly disparities react and whether mitigation constraints remain satisfied. Spectral analysis of simulated disparity signals can indicate whether the system is likely to oscillate under retraining and feedback. If mitigation reduces disparity on average but introduces high-amplitude oscillations that periodically disadvantage a group, the average metric may look acceptable while practical harm occurs in bursts. Frequency-domain constraints can be introduced by penalizing high power in disparity signals at certain frequencies, for example by adding a term proportional to $\sum_{\omega} |\hat{d}(\omega)|^2 w(\omega)$ with weights emphasizing low-frequency drift or retraining cadence peaks. This connects fairness monitoring to control-theoretic stability considerations [24].

7. Conclusion

Assessing algorithmic bias in two-sided e-commerce marketplaces requires treating search, ranking, and recommendation not as isolated predictors but as a coupled allocation system that mediates attention, learning signals, and market dynamics. A technically useful framing centers on exposure as the scarce interface resource, with downstream outcomes such as clicks, purchases, and seller viability emerging through behavioral and strategic responses. Bias can arise from data and label processes shaped by historical exposure, from objective functions that encode proxy trade-offs, from retrieval and representation limitations that cap opportunity, and from feedback loops that amplify small differences into persistent concentration. Measurement therefore benefits from jointly reporting exposure and outcome disparities with descriptive distributional statistics, robust uncertainty quantification, and diagnostics that account for clustering and nonstationarity. Frequency spectrum analysis and logarithmic disparity scales offer additional visibility into instability and multiplicative gaps that can be hard to interpret on linear scales.

Counterfactual evaluation is necessary to move from description to actionable assessment, but it depends on exploration, logging, and careful handling of interference. Off-policy and doubly robust estimators can support policy comparison when propensities are available, while experiments and quasi-experiments provide stronger identification at higher operational cost. For long-run effects, dynamic modeling becomes important; PDE-inspired exposure propagation models, augmented with damped wave analogies for oscillatory behavior, provide a structured way to approximate how disparities evolve under repeated interaction. Finite element discretizations and stable time integration schemes offer scalable numerical tools for scenario analysis on irregular marketplace structures. Mitigation can be expressed as fairness-constrained optimization over ranking and recommendation policies, with parametric analysis clarifying trade-offs and stress testing highlighting robustness under demand shocks and retraining cycles. In operational use, these components support an audit loop in which disparity metrics, causal estimates, and stability diagnostics jointly inform interventions that are compatible with marketplace constraints and that remain interpretable under dynamic feedback [25].

References

- [1] K. Sasahara, W. Chen, H. Peng, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Social influence and unfollowing accelerate the emergence of echo chambers," *Journal of Computational Social Science*, vol. 4, no. 1, pp. 381–402, 9 2020.
- [2] M. Madianou, "The biometric assemblage: Surveillance, experimentation, profit, and the measuring of refugee bodies:," *Television & New Media*, vol. 20, no. 6, pp. 581–599, 7 2019.
- [3] D. Y. Bhomia, D. S. K. Mishra, and R. Pal, "The role of convolutional neural networks in automated diagnosis of neurological disorders: A critical analysis," *The Pharma Innovation*, vol. 8, no. 1S, pp. 1–4, 1 2019.

- [4] N. V. Patel, "Selection bias in two-sided e-commerce marketplaces: A framework for propensity score matching implementation," *Journal of Business Intelligence Systems and Computational Social Science Applications*, vol. 11, no. 5, pp. 1–20, 2021.
- [5] I. N. Cofone, "Ai and judicial decision-making," 11 2020.
- [6] M. D. Ekstrand, P. Kannan, J. A. Stemper, J. T. Butler, J. A. Konstan, and J. Riedl, "Recsys - automatically building research reading lists," in *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 9 2010, pp. 159–166.
- [7] T. Baer, *The Stability Bias of Algorithms*. Apress, 6 2019, pp. 79–86.
- [8] P. Sharad, "Evaluation of block chain and artificial intelligence using spss statistics analysis," *REST Journal on Emerging trends in Modelling and Manufacturing*, vol. 5, no. 4, pp. 139–146, 12 2019.
- [9] F. T. Chan, R. K. Tibrewal, A. Prakash, and M. K. Tiwari, *Inventory Based Multi-Item Lot-Sizing Problem in Uncertain Environment: BRKGA Approach*. United States: Springer International Publishing, 6 2013, pp. 1197–1206.
- [10] D. Strbenac, N. J. Armstrong, and J. Y. H. Yang, "Detection and classification of peaks in 5' cap rna sequencing data," *BMC genomics*, vol. 14, no. 5, pp. 1–11, 10 2013.
- [11] A. O. College and R. Libraries, "Acrl annual report 2019-2020," *College & Research Libraries News*, vol. 81, no. 11, pp. 559–559, 12 2020.
- [12] P. G. Fennell, Z. Zuo, and K. Lerman, "Predicting and explaining behavioral data with structured feature space decomposition," *EPJ Data Science*, vol. 8, no. 1, pp. 1–27, 6 2019.
- [13] C. Shulby, M. D. Ferreira, R. F. de Mello, and S. M. Aluísio, "Theoretical learning guarantees applied to acoustic modeling," *Journal of the Brazilian Computer Society*, vol. 25, no. 1, pp. 1–12, 1 2019.
- [14] E. S. Udoh, "Icegov - is the data fair?: An assessment of the data quality of algorithmic policing systems," in *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*. ACM, 9 2020, pp. 1–7.
- [15] Y. Chan, M. Tung, A. S. Garruss, S. G. A. W. Zaranek, Y. K. Chan, J. E. Lunshof, A. W. Zaranek, M. Ball, M. F. Chou, E. T. Lim, and G. M. Church, "An unbiased index to quantify participant's phenotypic contribution to an open-access cohort," *Scientific reports*, vol. 7, no. 1, pp. 46 148–46 148, 4 2017.
- [16] S. Vijayabaskar, "Harnessing generative ai for risk management and fraud detection in fintech: A new era of human-machine collaboration," *International Journal of Scientific Research and Management (IJSRM)*, vol. 8, no. 4, pp. 369–379, 4 2020.
- [17] B. Goankar, K. Cook, and L. Macyszyn, "Ethical issues arising due to bias in training a.i. algorithms in healthcare and data sharing as a potential solution," *AI Ethics Journal*, vol. 1, no. 2, 9 2020.
- [18] S. Sommer, S. Kozerke, E. Seifritz, and P. Staempfli, "Fiber up-sampling and quality assessment of tractograms - towards quantitative brain connectivity." *Brain and behavior*, vol. 7, no. 1, pp. e00 588–, 10 2016.
- [19] A. E. Hughes, D. Griffiths, J. Troscianko, and L. A. Kelley, "No evidence for motion dazzle in an evolutionary citizen science game," 10 2019.

- [20] J. Gellaerts, E. Bogdanov, F. Dadashi, and B. Mariani, "In-field validation of an inertial sensor-based system for movement analysis and classification in ski mountaineering." *Sensors (Basel, Switzerland)*, vol. 18, no. 3, pp. 885–, 3 2018.
- [21] A. L. H. Nelson and S. Zanti, "Framework for centering racial equity throughout the administrative data life cycle," *International Journal of Population Data Science*, vol. 5, no. 3, 9 2020.
- [22] C. Burke and C. S. Bloss, "Social media surveillance in schools: Rethinking public health interventions in the digital age." *Journal of medical Internet research*, vol. 22, no. 11, pp. e22612–, 11 2020.
- [23] L. J. Paxton, D. Knuth, S. Gribben, and E. Tunstel, "Cricket: Cryogenic reservoir inventory by cost-effective kinetically enhanced technology," 1 2019.
- [24] M. Ji, X. Chen, G. Z. Genchev, T. Xu, M. Wei, and G. Yu, "Status of ai-enabled clinical decision support systems clinical implementations in china," 7 2020.
- [25] S. G. Han, S. L. Guan, Y. Wu, Z. X. Bo, Y. Ding, and L. A. Chen, "Research of partial discharge monitoring for switchgear cabinet based on lssvm," *MATEC Web of Conferences*, vol. 63, pp. 04016–, 7 2016.