Fairness in Search Systems

Other titles in Foundations and Trends® in Information Retrieval

User Simulation for Evaluating Information Access Systems
Krisztian Balog and ChengXiang Zhai
ISBN: 978-1-63828-378-2

Multi-hop Question Answering
Vaibhav Mavi, Anubhav Jangra and Adam Jatowt
ISBN: 978-1-63828-374-4

Conversational Information Seeking
Hamed Zamani, Johanne R. Trippas, Jeff Dalton and Filip Radlinski
ISBN: 978-1-63828-200-6

Perspectives of Neurodiverse Participants in Interactive Information Retrieval
Laurianne Sitbon, Gerd Berget and Margot Brereton
ISBN: 978-1-63828-202-0

Efficient and Effective Tree-based and Neural Learning to Rank
Sebastian Bruch, Claudio Lucchese and Franco Maria Nardini
ISBN: 978-1-63828-198-6

Quantum-Inspired Neural Language Representation, Matching and Understanding
Peng Zhang, Hui Gao, Jing Zhang and Dawei Song
ISBN: 978-1-63828-204-4

Fairness in Search Systems

Yi Fang
Santa Clara University
yfang@scu.edu

Ashudeep Singh
Microsoft
ashudeep.singh@microsoft.com

Zhiqiang Tao
Rochester Institute of Technology
zhiqiang.tao@rit.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Y. Fang *et al.*. *Fairness in Search Systems*. Foundations and Trends® in Information Retrieval, vol. 18, no. 3, pp. 262–416, 2024.

ISBN: 978-1-63828-499-4

Foundations and Trends® in Information Retrieval
Volume 18, Issue 3, 2024

Editorial Scope

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2024, Volume 18, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Fairness in Search Systems

Yi Fang[1], Ashudeep Singh[2] and Zhiqiang Tao[3]

[1] *Department of Computer Science and Engineering, Santa Clara University, USA; yfang@scu.edu*

[2] *Microsoft, USA; ashudeep.singh@microsoft.com*

[3] *School of Information, Rochester Institute of Technology, USA; zhiqiang.tao@rit.edu*

ABSTRACT

Search engines play a crucial role in organizing and delivering information to billions of users worldwide. However, these systems often reflect and amplify existing societal biases and stereotypes through their search results and rankings. This concern has prompted researchers to investigate methods for measuring and reducing algorithmic bias, with the goal of developing more equitable search systems. This monograph presents a comprehensive taxonomy of fairness in search systems and surveys the current research landscape. We systematically examine how bias manifests across key search components, including query interpretation and processing, document representation and indexing, result ranking algorithms, and system evaluation metrics. By critically analyzing the existing literature, we identify persistent challenges and promising research directions in the pursuit of fairer search systems. Our aim is to provide a foundation for future work in this rapidly evolving field while highlighting opportunities to create more inclusive and equitable information retrieval technologies.

Yi Fang, Ashudeep Singh and Zhiqiang Tao (2024), "Fairness in Search Systems", Foundations and Trends® in Information Retrieval: Vol. 18, No. 3, pp 262–416. DOI: 10.1561/1500000101.

1

Introduction

> Equals should be treated equally and unequals unequally.
>
> *Aristotle, 384–322 BC*

Search systems are ubiquitous across a wide array of platforms, from online information sources such as web search engines, e-commerce sites, and social media to sociotechnical systems encompassing admissions, housing, and employment platforms. They significantly influence the flow of information and transactions, dictating the content that gets consumed, the products purchased, employment decisions, and admissions processes. The impact of these systems extends to both sides of the spectrum: they serve not only consumers, such as web users, employers, purchasers, and admissions officials, who rely on them to make informed choices but also providers, such as content creators, sellers, job applicants, and media organizations, whose visibility and success are directly affected by how they are ranked and presented within these systems. This dual influence underscores the substantial role that search systems play in access to information, shaping economic opportunities, and social mobility. In recent years, there has been a growing focus within the Information Retrieval (IR) community on the *fairness* of search systems. This concern centers around whether the resources and benefits provided by these systems are equitably distributed among the various individuals or entities they impact. There is also a scrutiny of whether these systems perpetuate or introduce harms, especially those that are distributed in ways that are considered unfair or unjust.

Reflecting on the evolutionary trajectory of retrieval models over the past few decades reveals a significant shift towards data and machine learning driven methodologies. Initially, IR systems relied primarily on ranking algorithms that utilized various heuristics, such as TF-IDF weighting, to determine the relevance between a query and a document. The idea of aggregating multiple signals into the ranking process without resorting to heuristic methods led to the learning-to-rank techniques in the 2000s (Liu *et al.*, 2009), which involved defining hand-crafted features that capture different notions of what constitutes a relevant match, with machine learning models then tasked with learning the optimal combination of these features from training data. Recent neural IR models further eliminated the need for manual feature design (Mitra and Craswell, 2017). The rise of large language models (LLMs) is expected to dramatically transform the field of IR through their remarkable capabilities in language understanding, generation, generalization, and reasoning (Zhu *et al.*, 2023). These models bring a new level of sophistication to responding to complex queries. With the evolution of search engines into predominantly data-driven AI systems, they are increasingly susceptible to data and algorithmic biases. These biases can significantly impact the fairness of search results, potentially disadvantaging certain groups of consumers or providers, or reinforcing stereotypes.

In this monograph, we provide an introduction to fairness in search systems, with the aim of offering a starting point for understanding the problem space, reviewing the body of existing research, and laying the groundwork for further exploration and study in this critical area. Our focus is primarily on the fairness of a search system in delivering results that meet a user's information needs as encoded in their queries. We address fairness-related biases and harms, rather than the wider spectrum of issues that search systems might encounter, such as the propagation of misinformation.

1.1 History of Fairness in Search

The history of fairness research in search has evolved over several decades, reflecting a growing understanding of how these factors impact the user experience and the ethical implications of IR systems.

In the early years of IR, dating back to the 1960s and 1970s, the primary goal was to provide users with a list of documents that contained the queried keywords. Early IR systems did not incorporate sophisticated algorithms for ranking these documents, and as a result, search results often lacked the depth and relevance of modern search engines. However, interestingly, unfair rankings were discussed by Cooper and Robertson in the probability ranking principle work (Robertson, 1977), even though they did not use the term "fairness" as such (Hiemstra, 2023). It was revealed that unfair rankings may arise from blindly applying the principle without checking whether its preconditions are met.

The 1990s saw a significant expansion in search with the advent of the Internet. The focus started shifting towards improving search algorithms for better relevance and precision. Google's PageRank algorithm (Page *et al.*, 1998) revolutionized search technology, which considered not only keywords but also the quality and relevance of web pages. As the commercial interests grew, search advertising became prominent. Advertisers could pay to have their content displayed when specific keywords were searched. This practice had the potential to introduce bias in search results, as the presence and ranking of content became influenced by commercial interests rather than purely by relevance and quality.

During this era, the aspects of diversity and novelty in search results began to gain attention, particularly in the context of providing a broad range of search results to users (Clarke *et al.*, 2008). As search engines became integral to daily life, concerns regarding bias in search results also began to surface. Algorithmic bias became a topic of discussion, especially as it related to the ranking of websites. Critics have argued that search engine algorithms sometimes favor authoritative sources while marginalizing smaller or less mainstream voices in search results, in effect leading to concerns about information monopolies (Segev, 2010).

Discussions about net neutrality in the late 2000s and early 2010s also brought search engine neutrality into the spotlight, as part of the broader debate about equal access to online information (Crane, 2011). Search engine neutrality refers to the idea that search engines should have no inherent biases in their algorithms and should treat all web pages and content sources equally without favoritism. The central question was whether search engines should serve as neutral platforms that provided unfiltered and uncurated search results. The discussions about neutrality raised complex questions about the role of search engines as information gatekeepers and the potential consequences of curating content. Search engine providers faced increased scrutiny from regulatory bodies. They were challenged on practices such as favoring their own services in search results, penalizing competitor websites, and lack of transparency in their ranking algorithms. Legal battles and antitrust investigations became more common, as seen in the European Commission Guidelines on Ranking Transparency (Commission, 2020), as governments sought to ensure that search engines operated fairly and did not abuse their market dominance.

In the realm of IR research, numerous early studies have shed light on various forms of unfairness in search results. These encompass a range of biases, including racial, gender, and political viewpoint biases, which have raised concerns about the perpetuation of stereotypes through biased search outcomes. This area of inquiry is part of the broader research landscape focusing on fairness in sociotechnical and AI systems (Mitchell *et al.*, 2021), yet IR systems present their unique challenges and opportunities (Ekstrand *et al.*, 2022). Early work (Friedman and Nissenbaum, 1996; Introna and Nissenbaum, 2000) recognized the inherent capacity of search engines to incorporate social, political, and moral values into their ranking algorithms. To quantify the impact of such embedded values, Mowshowitz and Kawaguchi (2002) proposed a metric for measuring a search engine's deviation from an ideal exposure of content. Beyond the study of bias in algorithmic ranking, Vaughan and Thelwall (2004) and Vaughan and Zhang (2007) discovered that biases can arise from skewed crawling and indexing processes. Furthermore, the concept of document retrievability (Azzopardi and Vinay, 2008) investigated the distribution skew in document retrievability across various retrieval systems, contributing valuable insights into the mechanics of search engine fairness.

In the 2020s, calls for ensuring fairness in search engine algorithms have intensified. Many raised concerns about the biases of AI and machine learning algorithms used in search engines (Baeza-Yates, 2018; Gao and Shah, 2020). The need to make these algorithms more equitable gained prominence. Ethical considerations became essential to the development and deployment of search engine algorithms. The relationship between the relevance of search algorithm results (and consequently, the revenue of the search engine) and the fairness of those results is not inherently contradictory. It has been shown that there are instances where enhancing the quality of the results, quantified by metrics such as Reciprocal Rank (RR), Average Precision (AP), or Normalized Cumulative Discounted Gain (nDCG), can also simultaneously improve the fairness of the outcomes (Hiemstra, 2023).

Fairness in search engines remains a dynamic and evolving field. In recent years, there has been a generally increasing number of publications on fair search as shown in Figure 1.1. The scope of this survey covers more than 400 papers including the representative papers about fairness studies in AI and the papers about fairness in search published in the top IR related conferences and journals such as SIGIR, CIKM, WSDM, WWW, KDD, ICTIR, ECIR, RecSys, FAccT, FnTIR, TOIS, ACL, EMNLP, NAACL, AAAI, IJCAI, NeurIPS, ICML, as well as some of the outstanding arXiv papers.

1.2 Fairness, Bias, and Diversity

While *fairness*, *bias*, and *diversity* are frequently discussed as interrelated concepts in the research community, their relationships remain complex and often misunderstood. According to the Cambridge Dictionary,[1] *bias* represents a disproportionate inclination for or against certain ideas or things, whereas *fairness* describes the equitable and reasonable

[1] https://dictionary.cambridge.org

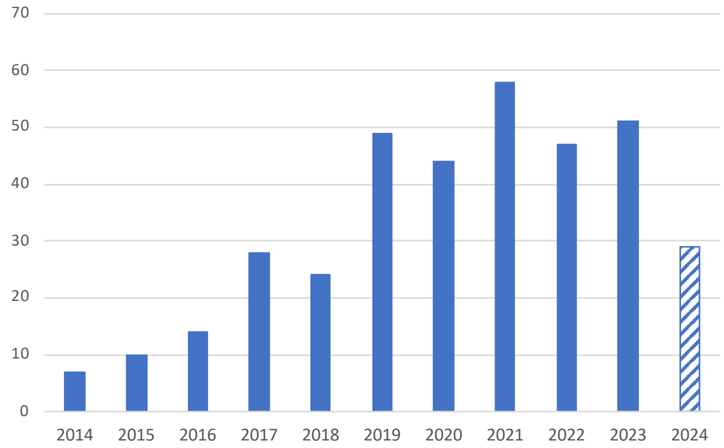

Figure 1.1: Publication trends in fairness in search (2014-2024). The data for 2024 is shaded to indicate that it represents an incomplete year at the time of this analysis.

treatment of individuals. This distinction is important: *bias* describes an observable characteristic of a system without making value judgments, while *fairness* addresses the ethical implications and societal impacts of system behavior (Ekstrand *et al.*, 2022).

Generally, different types of biases are key contributors to unfair outcomes in AI systems. The linkage between specific biases and resultant unfairness can be intricate (Li *et al.*, 2023). For instance, unfairness related to race and ethnicity might stem from biases in training data, model design, optimization algorithms, or evaluation benchmarks. Furthermore, a single type of bias, such as that in training data, can lead to various forms of unfairness such as individual and group unfairness.

On the other hand, the presence of bias does not inevitably lead to unfairness. For example, when a user searches for restaurants, a search engine shows results biased towards local establishments. This localization bias is based on the user's geographic location, which aligns with the user's likely intent. Beyond data and algorithmic biases, other factors can contribute to unfairness. It has been shown that certain fairness requirements are inherently conflicting, suggesting that upholding one type of fairness could inadvertently violate another (Kleinberg *et al.*, 2016).

Recent research in search systems has delved into various biases and debiasing methods (Zehlike *et al.*, 2022; Ekstrand *et al.*, 2022), but a clear distinction between research on bias and that on unfairness often remains elusive. Primarily, debiasing research tends to concentrate on enhancing retrieval performance, rather than explicitly promoting fairness. They usually conduct experiments based on improvements in relevance of results alone, using these gains to demonstrate the effectiveness of debiasing. In contrast, studies on fairness typically offer clear definitions and quantitative metrics for evaluating model unfairness, such as using performance disparities across groups to assess group-level unfairness. Fairness-focused research often assesses methods against both fairness metrics and traditional retrieval metrics.

While biases are recognized as key contributors to unfairness and debiasing methods can potentially improve fairness, many fairness studies do not rely on debiasing but instead directly incorporate fairness requirements into model design. This approach, like imposing fairness regularization during optimization, can sometimes compromise model accuracy. Hence, there is a discernible research gap between debiasing and fairness, despite their theoretical and practical interconnections (Li *et al.*, 2023). A more nuanced understanding of the relationship between bias, unfairness, and the interplay of debiasing and fairness enhancement methods could lead to more effective strategies that improve both fairness and accuracy in search systems.

Diversity in IR is about ensuring a wide range of information in search results. This means that the results should include a variety of sources, viewpoints, or content types, rather than being dominated by a few sources or perspectives. In many cases, efforts to improve fairness in IR systems also enhance diversity. For example, algorithms designed to reduce bias in search results often lead to a more diverse set of search results. On the other hand, there can be tensions between these two goals. For example, maximizing diversity in search results might sometimes lead to less fair outcomes for certain groups, or vice versa. In the literature, the notion of coverage-based diversity (Drosou *et al.*, 2017) is most closely related to fairness, which requires that members of multiple, possibly overlapping, groups, be sufficiently well-represented among the top-k, treated either as a set or as a ranked list. Both fairness and diversity should consider the user perspective. An IR system might be fair and diverse from a content perspective but still fail to meet the diverse needs and fairness expectations of different user groups.

Fairness is frequently encapsulated within the broader framework of FACTS-IR that stands for Fairness, Accountability, Confidentiality, Transparency, and Safety in Information Retrieval that also contains the other pivotal aspects of responsible IR. The report from the FACTS-IR Workshop (Olteanu *et al.*, 2021) delves into the interplay and significance of these concepts. In this survey, our primary focus is on fairness, although we will also touch upon the other aspects, particularly in contexts where they intersect with or influence fairness.

1.3 Biases in Search

The search process can be conceptualized as a feedback loop encompassing various stages, such as query formulation and understanding, document representation, retrieval (or candidate generation), ranking, user feedback, and evaluation. At each of these stages, biases may arise, and the cyclic nature of the feedback loop has the potential to sustain or even intensify these biases. While this survey primarily focuses on fairness, it is important to recognize that various types of biases are significant contributors to unfair outcomes in search systems. A thorough understanding of how these biases interplay is essential for delivering fair and accurate information to users. In this section, we outline the architecture of a typical search engine, highlighting potential biases at each stage as depicted in Figure 1.2. While this list of biases is not exhaustive, it aims to provide an initial understanding of how biases can manifest throughout the search process. More detailed discussions on biases and unfairness, their implications, and mitigation strategies are provided in the subsequent sections.

Given data sources, **crawling and indexing** are the foundational processes in search engines that determine what content becomes searchable. Crawling is the first step where crawlers, also known as spiders, systematically browse the web to collect data from accessible web pages. Due to the extensive nature of the web, *crawling bias* may occur when these crawlers favor certain pages over others based on factors such as

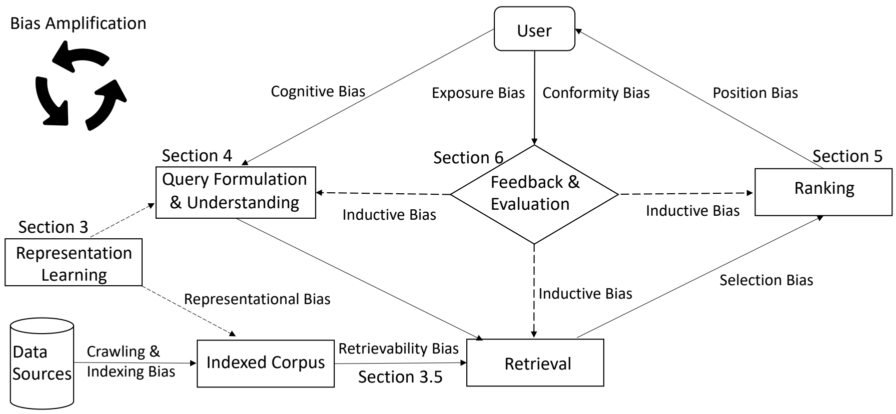

Figure 1.2: An overview of biases that can emerge at various stages in the life cycle of a search system. Section x in the figure refers to the specific section where the corresponding fairness issues are discussed.

page popularity or the quality and quantity of incoming links. This prioritization can result in the underrepresentation of less popular or newly established websites. Additionally, *indexing bias* can arise during the organization and storage of data, where a search engine might prioritize certain content, potentially distorting representation based on aspects like language, popularity, or perceived relevance. This can disproportionately represent cultural and linguistic content. Moreover, technical constraints and operational guidelines, such as the use of robots.txt files to guide crawler activities, can inadvertently introduce biases.

Query formulation and understanding begin with the user entering a query into the search engine. It involves a multi-faceted analysis of user queries to interpret their intent, context, and meaning. A *cognitive bias* is a systematic pattern of deviations in thinking which may lead to errors in judgments and decision-making (Azzopardi, 2021). Such biases may significantly influence how users formulate their queries. For instance, *confirmation bias* stems from people's tendency to prefer confirmatory information, where they discount information that does not conform to their existing beliefs. When querying, this may manifest as people employing positive test strategies where they try to find information that supports their hypotheses.

Representation learning involves transforming documents or queries into a format that can be efficiently processed by a search system. During this stage, each document/query is analyzed and converted into a structured form, often as a vector of features, which is then indexed and stored in the search system's database. This process also involves pre-processing steps such as tokenization, removal of stop words, and stemming or lemmatization. The goal is to distill the essence of each document into a representation that captures its main themes and content in a way that can be readily compared with user queries, facilitating effective and efficient retrieval in response to search requests.

Representational bias may emerge in representation learning. This bias can stem from a variety of factors related to the content, sources, and historical context of the documents. It manifests as skewed or unbalanced perspectives, representations, or information within the corpus itself, which can lead to a misrepresentation of certain demographics, viewpoints, or subject areas, affecting the fairness and accuracy of the search process. *Representational bias* is not introduced by the retrieval algorithms but rather originates from the intrinsic characteristics of the corpus. Bias inherent in training corpora can not only persist but also amplify (Papakyriakopoulos *et al.*, 2020; Wang *et al.*, 2024c) in learned latent representations through deep neural networks, such as pre-trained word embeddings (Brunet *et al.*, 2019), BERT (Kurita *et al.*, 2019), and more recently in LLMs (Gallegos *et al.*, 2023).

Retrieval is a process that retrieves all the candidates that match the user query from the index. In general, the retrieval system has to be fast and lightweight, as it considers the contents of the entire index. *Retrievability bias* measures how easily a document can be retrieved and exposed to the later ranking stage. A system with pronounced *retrievability bias* disproportionately favors certain documents over others (Azzopardi and Vinay, 2008), potentially resulting in unfair outcomes in the search results (Otterbacher *et al.*, 2017). *Popularity bias* can also be manifested in retrieval, which is the tendency to retrieve popular items more frequently than their intrinsic popularity justifies. This bias stems from several contributing factors. The sheer volume and visibility of content from popular sources can overshadow less popular but relevant content in the retrieval process. Many search engines use link analysis algorithms such as PageRank to infer its importance or relevance. Popular pages with many inbound links are more likely to be retrieved due to their perceived authority. Some retrieval algorithms may use historical user interaction data, like click-through rates as indicators of relevance. Popular items that have been clicked on or interacted with more frequently are likely to be considered more relevant, thus being retrieved more often.

Ranking involves reordering the top results obtained from the retrieval process. This can be based on chronological order, relevance criteria, or a combination of both. Learning-to-rank techniques are often employed at this stage to enhance the relevance of the results (Liu *et al.*, 2009). Beyond the *popularity bias* noted in the retrieval stage, the ranking stage is also subject to biases introduced during retrieval. Specifically, *selection bias* occurs when the initial set of documents retrieved dictates the subsequent ranking order (Wang *et al.*, 2023c). If this initial retrieval is biased or narrow in scope, the range of documents available for ranking becomes limited. As a result, the ranking stage is constrained to working with this pre-selected set, potentially overlooking more relevant or diverse documents that were not initially retrieved.

When ranked results are presented to the users, *position bias* occurs when users engage more frequently with items at the top of a ranked list, often irrespective of the actual relevance of these results. Eye-tracking studies have shown that users typically focus on the initial items and are less likely to consider those positioned lower (Joachims *et al.*, 2007b). Other research indicates that users often place undue trust in the top-ranked results and may not evaluate subsequent items as thoroughly, leading to a lack of holistic assessment of all available results (O'Brien and Keane, 2006).

User feedback on ranked search results can be categorized into two types: explicit and implicit. Explicit feedback is provided directly by users in a clear and intentional manner such as ratings and surveys. It represents a deliberate effort to convey relevance satisfaction with the search results. Explicit feedback can also be done by third-party human annotators by providing relevance judgment on query-document pairs. Implicit feedback is gathered from user behavior and interactions that are not directly intended as feedback but can be interpreted as such. It is unobtrusively collected as users go about their normal activities. Examples include click-through rate (CTR), dwell time, scroll depth, mouse movements, query reformulations, bounce rate, and so on. **Evaluation** is required to continuously monitor the performance of a search engine, as well as for measuring the effect of new changes that are introduced to any of its components. Evaluation can be done either manually, using explicit feedback, or automatically by tracking the implicit feedback such as clicks and session metrics.

Conformity bias can skew user explicit feedback, as individuals often align their behaviors with group norms, sometimes overriding their personal judgment (Azzopardi, 2021). This can lead to feedback that does not accurately represent their true opinions. Similarly, *confirmation bias* occurs when users selectively favor or emphasize search results that align with their pre-existing beliefs. This bias can result in feedback that reflects personal preferences or beliefs rather than an impartial assessment of the search results' quality.

Unlike explicit feedback, implicit feedback only offers a limited indication of user preference, as it lacks accurate information on what users like or dislike. *Exposure bias* is a significant issue in this context, arising from the fact that users only interact with a subset of documents. Consequently, not all unobserved interactions imply a negative preference. This ambiguity stems from two potential reasons for an unobserved interaction: either the document was not relevant to the user, or the user was simply unaware of it. This makes it challenging to accurately differentiate between genuinely negative interactions where the user is exposed to but not interested in a document and potentially positive ones where the user is not exposed to the document. As a result, this inability to distinguish between different types of unobserved interactions can lead to substantial biases in the learning process (Chen *et al.*, 2023b).

User feedback and evaluations are pivotal to update the parameters of machine learning models in various components, including query understanding, retrieval, and ranking, thus creating a feedback loop. To enhance specific desirable properties, *inductive biases* can be intentionally incorporated into the model design. *Inductive biases* are the underlying assumptions that a model uses to better learn the target function and generalize beyond the training data. These biases are often not harmful but essential, as the core of machine learning is the ability to extrapolate predictions to new, unseen examples. Without making certain assumptions about the data or model, generalization is impossible, as the output for unseen examples could vary widely. The development of an effective search system requires the incorporation of specific assumptions about the nature of the target function to guide the learning process. Moreover, some unfairness mitigation strategies, such as the in-processing methods discussed in Section 5, leverage inductive bias to correct for certain biases.

As shown in Figure 1.2, the search process forms a feedback loop and biases emerge in different stages of the loop. These biases could be further amplified over time along the loop. Take *popularity bias* or *position bias* as an example. Initially, certain documents may be ranked higher due to their popularity or early user engagement. These documents then garner additional feedback, which influences future rankings, potentially fostering a rich-get-richer dynamic (Joachims *et al.*, 2017c). This phenomenon raises important fairness questions regarding how exposure should be distributed, ideally based on the merit of the documents or items, rather than their initial popularity or position (Biega *et al.*, 2018; Singh and Joachims, 2018). For instance, in a job applicant ranking system, such dynamics could exacerbate existing unfairness, such as gender disparities. Similarly, in an online marketplace, this bias could favor certain sellers (or groups), leading to monopolistic tendencies and potentially driving other sellers out of the market (Morik *et al.*, 2020). Both scenarios highlight the important need to address the biases and feedback loop to prevent the reinforcement of existing disparities in search systems.

1.4 Comparisons with Related Surveys

In recent years, a number of surveys discussing fairness and bias in general machine learning have been published (Caton and Haas, 2020; Castelnovo *et al.*, 2022). They usually focus on the fairness works in classification tasks. A few surveys provide an overview of fairness in recommendation tasks (Wang *et al.*, 2023b). Recommendation algorithms can usually be considered as a type of ranking algorithm, but they often represent different characteristics. Pitoura *et al.* (2021) addresses fairness in both ranking and recommendation, and Ekstrand *et al.* (2022) discusses fairness in information access systems such as information retrieval and recommendation. Chen *et al.* (2023b) provides a survey on bias and debias in recommender systems, which covers a part of the content about fairness in recommendation. Similarly, Li *et al.* (2023) offers a systematic survey of existing works on fairness in recommendation by focusing on the foundations for fairness in recommendation literature. Recently, Dai *et al.* (2024a) presents a survey on bias and unfairness in IR systems that incorporate large language models predominantly references studies from the recommendation systems domain. While covering a brief introduction about fairness in classification and ranking, our survey pays specific attention to organizing the concept of fairness in search through a comprehensive taxonomy of fairness notions proposed in search problems, the task-specific techniques for promoting ranking fairness, as well as the datasets specially suitable for fairness research in search.

Three surveys were focused on fairness in ranking and retrieval systems (Ekstrand *et al.*, 2022; Zehlike *et al.*, 2022; Patro *et al.*, 2022). One recent survey performed a systematic literature review of the field of fairness, accountability, transparency, and ethics in information retrieval (Bernard and Balog, 2023). Our survey distinguishes itself from existing literature by offering several key advantages: 1) it provides a holistic review of unfairness across the entire life cycle of a search process, in contrast to previous surveys that primarily concentrate on fairness in ranking; 2) it introduces a thorough taxonomy of fairness in search and retrieval, aiding readers in comprehending various fairness considerations within search systems and facilitating an organized framework for navigating the literature in this domain; and 3) it is designed to be accessible, enabling newcomers to the field to develop a systematic understanding of the subject.

It is also worth noting that there have been several tutorials and workshops related to investigating biases and fairness issues in IR including the following: Addressing Bias and Fairness in Search Systems at SIGIR 2021 (Gao and Shah, 2021), Fairness of Machine Learning in

Recommender Systems at CIKM 2021 (Li *et al.*, 2021b), Fair Graph Mining at CIKM 2021 (Kang and Tong, 2021), Gender Fairness in Information Retrieval Systems at SIGIR 2022 (Bigdeli *et al.*, 2022), Fairness of Machine Learning in Search Engines at CIKM 2022 (Fang *et al.*, 2022), Bias and Unfairness in Information Retrieval Systems: New Challenges in the LLM Era at KDD 2024 (Dai *et al.*, 2024a) and WSDM 2025, and the workshop series on Algorithmic Bias in Search and Recommendation (BIAS) at ECIR 2020-2023 (Boratto *et al.*, 2023) and SIGIR 2024 (Bellogin *et al.*, 2024).

1.5 Intended Audience and Scope

This survey is beneficial for a wide array of individuals in the information retrieval field, including: 1) newcomers seeking a comprehensive guide to quickly delve into fairness issues in search systems; 2) those grappling with various sources of bias and requiring a systematic study to grasp the nuances of unfairness in search; 3) researchers aiming to stay up-to-date with cutting-edge techniques for mitigating unfairness in search; and 4) practitioners confronting unfairness challenges in the development of search systems and searching for effective solutions.

Primarily written for the IR community, this monograph also caters to diverse backgrounds such as machine learning, natural language processing, and AI ethics. It serves as an accessible entry point to the concept of fair search, enriched with numerous practical insights. We envision this resource as valuable for students, researchers, and software practitioners alike. Offering a holistic perspective and a thorough exploration of key ideas, it is essential for understanding and constructing modern search systems. These systems are crucial in enabling billions of users to access a wealth of global knowledge and services while ensuring fairness and equity in access.

1.6 Structure of the Survey

The monograph is structured as follows.

- Section 1 describes the architecture of a modern search system with important components and highlights various biases that may arise in the search process. We also briefly review the history of fairness in search.

- Section 2 provides background information about the bias in algorithmic decision-making in general and in search in particular. We review the existing work on bias mitigation in machine learning and discuss the challenges in this space.
- Section 3 focuses on representation learning and content analysis, and on how to learn an unbiased data representation.
- Section 4 investigates fairness in query understanding, specifically in query formulation, query suggestion, and non-textual queries.
- Section 5 studies fair ranking and how to mitigate unfairness in rankings.
- Section 6 discusses bias in relevance judgment (both explicit and implicit) and how to learn and evaluate with biased feedback.
- Section 7 discusses emerging research directions, prompted by the rise of large language models (LLMs) and the growing imperative for responsible AI. This section also examines the open challenges that define this evolving landscape.

References

Abdollahpouri, H., G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. Pizzato. (2019). "Beyond personalization: Research directions in multistakeholder recommendation". *arXiv preprint arXiv:1905.01986*.

Abdollahpouri, H. and R. Burke. (2019). "Multi-stakeholder recommendation and its connection to multi-sided fairness". *arXiv preprint arXiv:1907.13158*.

Abdollahpouri, H., R. Burke, and B. Mobasher. (2017). "Controlling popularity bias in learning-to-rank recommendation". In: *RecSys*.

Abid, A., M. Farooqi, and J. Zou. (2021). "Large language models associate Muslims with violence". *Nature Machine Intelligence*. 3(June): 461–463. DOI: 10.1038/s42256-021-00359-2.

Abolghasemi, A., L. Azzopardi, A. Askari, M. de Rijke, and S. Verberne. (2024). "Measuring Bias in a Ranked List Using Term-Based Representations". In: *European Conference on Information Retrieval*. Springer. 3–19.

Agarwal, A., K. Takatsu, I. Zaitsev, and T. Joachims. (2019a). "A General Framework for Counterfactual Learning-to-Rank". In: *SIGIR*.

Agarwal, A., I. Zaitsev, X. Wang, C. Li, M. Najork, and T. Joachims. (2019b). "Estimating Position Bias Without Intrusive Interventions". In: *WSDM*.

Agarwal, A., S. Basu, T. Schnabel, and T. Joachims. (2017). "Effective Evaluation Using Logged Bandit Feedback from Multiple Loggers". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery. 687–696.

Agarwal, A., X. Wang, C. Li, M. Bendersky, and M. Najork. (2019c). "Addressing Trust Bias for Unbiased Learning-to-Rank". In: *The World Wide Web Conference*. 4–14.

Agarwal, A., I. Zaitsev, X. Wang, C. Li, M. Najork, and T. Joachims. (2019d). "Estimating Position Bias without Intrusive Interventions". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 474–482.

Ahmad, W. U., K.-W. Chang, and H. Wang. (2018). "Multi-Task Learning for Document Ranking and Query Suggestion". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=SJ1nzBeA-.

Ai, Q., J. Mao, Y. Liu, and W. B. Croft. (2018). "Unbiased Learning to Rank: Theory and Practice". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management. CIKM '18*. Torino, Italy: Association for Computing Machinery. 2305–2306.

Alayrac, J.-B., J. Donahue, P. Luc, A. Miech, I. Barr, Y. Hasson, K. Lenc, A. Mensch, K. Millican, M. Reynolds, R. Ring, E. Rutherford, S. Cabi, T. Han, Z. Gong, S. Samangooei, M. Monteiro, J. Menick, S. Borgeaud, A. Brock, A. Nematzadeh, S. Sharifzadeh, M. Binkowski, R. Barreira, O. Vinyals, A. Zisserman, and K. Simonyan. (2022). "Flamingo: a Visual Language Model for Few-Shot Learning". In: *Advances in Neural Information Processing Systems*.

Ali, M., P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke. (2019). "Discrimination through optimization: How Facebook's Ad delivery can lead to biased outcomes". *CSCW*.

Angwin, J. and J. Larson. (2016a). "Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say". URL: https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say.

Angwin, J. and J. Larson. (2016b). "How We Analyzed the COMPAS Recidivism Algorithm". URL: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Anthis, J., K. Lum, M. Ekstrand, A. Feller, A. D'Amour, and C. Tan. (2024). "The Impossibility of Fair LLMs". In: *Proceedings of the 1st Human-Centered Evaluation and Auditing of Language Models (HEAL) workshop at CHI 2024*.

Aslam, J. A., V. Pavlu, and E. Yilmaz. (2006). "A Statistical Method for System Evaluation Using Incomplete Judgments". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '06*. Seattle, Washington, USA: Association for Computing Machinery. 541–548.

Aslam, J. A. and E. Yilmaz. (2007). "Inferring Document Relevance from Incomplete Information". In: *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management. CIKM '07*. Lisbon, Portugal: Association for Computing Machinery. 633–642.

Asudeh, A., H. Jagadish, J. Stoyanovichz, and G. Das. (2019). "Designing Fair Ranking Schemes". *ICDM*.

Azzopardi, L. (2021). "Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval". In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery. 27–37.

Azzopardi, L. and V. Vinay. (2008). "Retrievability: An evaluation measure for higher order information access tasks". In: *Proceedings of the 17th ACM conference on Information and knowledge management*. 561–570.

Badilla, P., F. Bravo-Marquez, and J. Pérez. (2020). "WEFE: The Word Embeddings Fairness Evaluation Framework". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. Ed. by C. Bessiere. International Joint Conferences on Artificial Intelligence Organization. 430–436. DOI: 10.24963/ijcai.2020/60.

Baeza-Yates, R. (2018). "Bias on the web". *Communications of the ACM*. 61(6): 54–61.

Bagdasaryan, E., O. Poursaeed, and V. Shmatikov. (2019). “Differential privacy has disparate impact on model accuracy”. *Advances in neural information processing systems*. 32.

Bakalar, C., R. Barreto, S. Bergman, M. Bogen, B. Chern, S. Corbett-Davies, M. Hall, I. Kloumann, M. Lam, J. Q. Candela, *et al.* (2021). “Fairness on the ground: Applying algorithmic fairness approaches to production systems”. *arXiv preprint arXiv:2103.06172*.

Balagopalan, A., A. Z. Jacobs, and A. J. Biega. (2023). “The Role of Relevance in Fair Ranking”. In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’23*. New York, NY, USA: Association for Computing Machinery. 2650–2660.

Balaneshin-kordan, S. and A. Kotov. (2017). “Embedding-Based Query Expansion for Weighted Sequential Dependence Retrieval Model”. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’17*. Shinjuku, Tokyo, Japan: Association for Computing Machinery. 1213–1216.

Banerjee, A., G. K. Patro, L. W. Dietz, and A. Chakraborty. (2020). “Analyzing ‘Near Me’Services: Potential for Exposure Bias in Location-based Retrieval”. In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. 3642–3651.

Barocas, S., M. Hardt, and A. Narayanan. (2019). *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org.

Barocas, S. and A. D. Selbst. (2016). “Big data’s disparate impact”. *California law review*: 671–732.

Bashir, S. and A. Rauber. (2010). “Improving retrievability of patents in prior-art search”. In: *European Conference on Information Retrieval*. Springer. 457–470.

Bellamy, R. K., K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, *et al.* (2019). “AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias”. *IBM Journal of Research and Development*. 63(4/5): 4–1.

Bellogin, A., L. Boratto, S. Kleanthous, E. Lex, F. M. Malloci, and M. Marras. (2024). "International Workshop on Algorithmic Bias in Search and Recommendation (BIAS)". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3033–3035.

Bengio, Y., A. Courville, and P. Vincent. (2013). "Representation Learning: A Review and New Perspectives". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 35(8): 1798–1828. DOI: 10.1109/TPAMI.2013.50.

Bernard, N. and K. Balog. (2023). "A Systematic Review of Fairness, Accountability, Transparency and Ethics in Information Retrieval". *ACM Computing Surveys*.

Betker, J., G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee, Y. Guo, *et al.* (2023). "Improving image generation with better captions". *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*.

Beutel, A., J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, and C. Goodrow. (2019a). "Fairness in Recommendation Ranking through Pairwise Comparisons". In: *KDD*.

Beutel, A., J. Chen, T. Doshi, H. Qian, A. Woodruff, C. Luu, P. Kreitmann, J. Bischof, and E. H. Chi. (2019b). "Putting fairness principles into practice: Challenges, metrics, and improvements". In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 453–459.

Beutel, A., J. Chen, Z. Zhao, and E. H. Chi. (2017). "Data Decisions and Theoretical Implications when Adversarially Learning Fair Representations". *ArXiv*. abs/1707.00075. URL: https://api.semanticscholar.org/CorpusID:24990444.

Beygelzimer, A. and J. Langford. (2009). "The Offset Tree for Learning with Partial Labels". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '09*. Paris, France: Association for Computing Machinery. 129–138.

Biega, A., F. Diaz, M. Ekstrand, and S. Kohlmeier. (2019). "TREC 2019 Fair Ranking Track".

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

Biega, A. J., P. Potash, H. Daume, F. Diaz, and M. Finck. (2020a). "Operationalizing the legal principle of data minimization for personalization". In: *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 399–408.

Biega, A. J., F. Diaz, M. D. Ekstrand, and S. Kohlmeier. (2020b). "Overview of the TREC 2019 Fair Ranking Track". *CoRR*. abs/2003.11650. URL: https://arxiv.org/abs/2003.11650.

Bigdeli, A., N. Arabzadeh, S. Seyedsalehi, B. Mitra, M. Zihayat, and E. Bagheri. (2023). "De-Biasing Relevance Judgements for Fair Ranking". In: *Advances in Information Retrieval: 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2–6, 2023, Proceedings, Part II*. Dublin, Ireland: Springer-Verlag. 350–358. DOI: 10.1007/978-3-031-28238-6_24.

Bigdeli, A., N. Arabzadeh, S. SeyedSalehi, M. Zihayat, and E. Bagheri. (2022). "Gender Fairness in Information Retrieval Systems". In: *SIGIR*.

Bigdeli, A., N. Arabzadeh, S. Seyedsalehi, M. Zihayat, and E. Bagheri. (2021a). "On the orthogonality of bias and utility in ad hoc retrieval". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1748–1752.

Bigdeli, A., N. Arabzadeh, M. Zihayat, and E. Bagheri. (2021b). "Exploring Gender Biases in Information Retrieval Relevance Judgement Datasets". In: *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 – April 1, 2021, Proceedings, Part II*. Springer-Verlag. 216–224.

Bird, S., M. Dudik, R. Edgar, B. Horn, R. Lutz, V. Milan, M. Sameki, H. Wallach, and K. Walker. (2020). "Fairlearn: A toolkit for assessing and improving fairness in AI". *Microsoft, Tech. Rep. MSR-TR-2020-32*.

Birhane, A., V. U. Prabhu, and E. Kahembwe. (2021). “Multimodal datasets: misogyny, pornography, and malignant stereotypes”. *arXiv preprint arXiv:2110.01963*.

Bithel, S. and S. Bedathur. (2023). “Evaluating Cross-Modal Generative Models Using Retrieval Task”. In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’23*. New York, NY, USA: Association for Computing Machinery. 1960–1965.

Bolukbasi, T., K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai. (2016a). “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings”. In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett. Vol. 29. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf.

Bolukbasi, T., K. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai. (2016b). “Quantifying and Reducing Stereotypes in Word Embeddings”. *CoRR*. abs/1606.06121. arXiv: 1606.06121. URL: http://arxiv.org/abs/1606.06121.

Bonart, M., A. Samokhina, G. Heisenberg, and P. Schaer. (2020). “An investigation of biases in web search engine query suggestions”. *Online Information Review*. 44(2): 365–381.

Boratto, L., S. Faralli, M. Marras, and G. Stilo. (2023). “Fourth International Workshop on Algorithmic Bias in Search and Recommendation (Bias 2023)”. In: *European Conference on Information Retrieval*. Springer. 373–376.

Borisov, A., I. Markov, M. de Rijke, and P. Serdyukov. (2016). “A Neural Click Model for Web Search”. In: *Proceedings of the 25th International Conference on World Wide Web*. 531–541.

Borisov, A., M. Wardenaar, I. Markov, and M. de Rijke. (2018). “A Click Sequence Model for Web Search”. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. SIGIR ’18*. Ann Arbor, MI, USA: Association for Computing Machinery. 45–54.

Bottou, L., J. Peters, J. Quiñonero-Candela, D. X. Charles, D. M. Chickering, E. Portugaly, D. Ray, P. Simard, and E. Snelson. (2013). "Counterfactual Reasoning and Learning Systems: The Example of Computational Advertising". *Journal of Machine Learning Research*. 14(101): 3207–3260. URL: http://jmlr.org/papers/v14/bottou13a.html.

Boyce, B. (1982). "Beyond topicality: A two stage view of relevance and the retrieval process". *Information Processing & Management*. 18(3): 105–109.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.* (2020). "Language models are few-shot learners". *Advances in neural information processing systems*. 33: 1877–1901.

Brunet, M.-E., C. Alkalay-Houlihan, A. Anderson, and R. Zemel. (2019). "Understanding the Origins of Bias in Word Embeddings". In: *Proceedings of the 36th International Conference on Machine Learning*. 803–811.

Buckley, C. and E. M. Voorhees. (2004). "Retrieval Evaluation with Incomplete Information". In: *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '04*. Sheffield, United Kingdom: Association for Computing Machinery. 25–32.

Buolamwini, J. and T. Gebru. (2018). "Gender shades: Intersectional accuracy disparities in commercial gender classification". In: *Conference on fairness, accountability and transparency*. PMLR. 77–91.

Burke, R., N. Sonboli, and A. Ordonez-Gauger. (2018). "Balanced Neighborhoods for Multi-sided Fairness in Recommendation". In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*.

Büttcher, S., C. L. A. Clarke, P. C. K. Yeung, and I. Soboroff. (2007). "Reliable Information Retrieval Evaluation with Incomplete and Biased Judgements". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '07*. Amsterdam, The Netherlands: Association for Computing Machinery. 63–70.

Cachel, K. and E. Rundensteiner. (2024). "Wise Fusion: Group Fairness Enhanced Rank Fusion". In: *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 163–174.

Cai, M., H. Liu, S. K. Mustikovela, G. P. Meyer, Y. Chai, D. Park, and Y. J. Lee. (2024). "Making Large Multimodal Models Understand Arbitrary Visual Prompts". In: *IEEE Conference on Computer Vision and Pattern Recognition*.

Calders, T., F. Kamiran, and M. Pechenizkiy. (2009). "Building classifiers with independency constraints". In: *Data mining workshops, ICDMW*. 13–18.

Caliskan, A., J. J. Bryson, and A. Narayanan. (2017). "Semantics derived automatically from language corpora contain human-like biases". *Science*. 356(6334): 183–186. DOI: 10.1126/science.aal4230. eprint: https://www.science.org/doi/pdf/10.1126/science.aal4230.

Calmon, F., D. Wei, B. Vinzamuri, K. Natesan Ramamurthy, and K. R. Varshney. (2017). "Optimized pre-processing for discrimination prevention". *Advances in neural information processing systems*. 30.

Campos, P. G., F. Diez, and I. Cantador. (2014). "Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols". *User Modeling and User-Adapted Interaction*. 24: 67–119.

Carbonell, J. and J. Goldstein. (1998). "The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries". In: *SIGIR*. Melbourne, Australia. 335–336. DOI: 10.1145/290941.291025.

Castelnovo, A., R. Crupi, G. Greco, D. Regoli, I. G. Penco, and A. C. Cosentini. (2022). "A clarification of the nuances in the fairness metrics landscape". *Scientific Reports*. 12(1): 4209.

Caton, S. and C. Haas. (2020). "Fairness in machine learning: A survey". *ACM Computing Surveys*.

Celis, L. E., D. Straszak, and N. K. Vishnoi. (2018). "Ranking with Fairness Constraints". *ICALP*.

Celma, Ò. and P. Cano. (2008). "From hits to niches?: or how popular artists can bias music recommendation and discovery". In: *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*. ACM. 5.

Chakraborty, A., S. Ghosh, N. Ganguly, and K. P. Gummadi. (2017). "Optimizing the recency-relevancy trade-off in online news recommendations". In: *Proceedings of the 26th International Conference on World Wide Web*. 837–846.

Chaney, A. J., B. M. Stewart, and B. E. Engelhardt. (2018). "How algorithmic confounding in recommendation systems increases homogeneity and decreases utility". In: *Proceedings of the 12th ACM conference on recommender systems*. 224–232.

Chapelle, O. and Y. Zhang. (2009). "A Dynamic Bayesian Network Click Model for Web Search Ranking". In: *Proceedings of the 18th International Conference on World Wide Web*. 1–10.

Chen, C., J. Fu, and L. Lyu. (2023a). "A pathway towards responsible ai generated content". *arXiv preprint arXiv:2303.01325*.

Chen, F. and H. Fang. (2023). "Learn to Be Fair without Labels: A Distribution-Based Learning Framework for Fair Ranking". In: *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval. ICTIR '23*. Taipei, Taiwan: Association for Computing Machinery. 23–32.

Chen, F., D. Yang, and H. Fang. (2024). "Toward Automatic Group Membership Annotation for Group Fairness Evaluation". In: *International Conference on Applications of Natural Language to Information Systems*. Springer. 285–300.

Chen, J., H. Dong, X. Wang, F. Feng, M. Wang, and X. He. (2023b). "Bias and debias in recommender system: A survey and future directions". *ACM Transactions on Information Systems*. 41(3): 1–39.

Chen, M., C. Liu, J. Sun, and S. C. Hoi. (2021). "Adapting Interactional Observation Embedding for Counterfactual Learning to Rank". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 285–294.

Cherumanal, S. P., F. Scholer, J. R. Trippas, and D. Spina. (2024). "Towards Investigating Biases in Spoken Conversational Search". *arXiv preprint arXiv:2409.00890*.

Cho, J., A. Zala, and M. Bansal. (2023). "DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generation Models". In: *ICCV*.

Cho, K., B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. (2014). "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1724–1734.

Cho, S., K. W. Crenshaw, and L. McCall. (2013). "Toward a field of intersectionality studies: Theory, applications, and praxis". *Signs: Journal of women in culture and society*. 38(4): 785–810.

Chouldechova, A. (2017). "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments". *Big data*. 5(2): 153–163.

Chuklin, A., I. Markov, and M. de Rijke. (2015). "Click Models for Web Search". In: *Click Models for Web Search*. URL: https://api.semanticscholar.org/CorpusID:38886570.

Clarke, C. L., M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. (2008). "Novelty and Diversity in Information Retrieval Evaluation". In: *SIGIR*. Singapore, Singapore. 659–666. DOI: 10.1145/1390334.1390446.

Commission, E. (2020). "Guidelines on ranking transparency pursuant to Regulation (EU) 2019/1150". URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020XC1208(01)&from=EN.

Crane, D. A. (2011). "Search neutrality as an antitrust principle". *Geo. Mason L. Rev.* 19: 1199.

Craswell, N., O. Zoeter, M. Taylor, and B. Ramsey. (2008a). "An Experimental Comparison of Click Position-Bias Models". In: *Proceedings of the 2008 International Conference on Web Search and Data Mining. WSDM '08*. Palo Alto, California, USA: Association for Computing Machinery. 87–94.

Craswell, N., O. Zoeter, M. Taylor, and B. Ramsey. (2008b). "An experimental comparison of click position-bias models". In: *WSDM*.

Crawford, K. (2017). "The trouble with bias (Invited Talk)". In: *NIPS*.

Dai, S., C. Xu, S. Xu, L. Pang, Z. Dong, and J. Xu. (2024a). "Bias and unfairness in information retrieval systems: New challenges in the llm era". In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 6437–6447.

Dai, S., Y. Zhou, L. Pang, W. Liu, X. Hu, Y. Liu, X. Zhang, G. Wang, and J. Xu. (2024b). "Neural retrievers are biased towards llm-generated content". In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 526–537.

Dai, X., J. Hou, Q. Liu, Y. Xi, R. Tang, W. Zhang, X. He, J. Wang, and Y. Yu. (2020). "U-Rank: Utility-Oriented Learning to Rank with Implicit Feedback". In: *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 2373–2380.

Dai, Z., V. Y. Zhao, J. Ma, Y. Luan, J. Ni, J. Lu, A. Bakalov, K. Guu, K. Hall, and M.-W. Chang. (2023). "Promptagator: Few-shot Dense Retrieval From 8 Examples". In: *The Eleventh International Conference on Learning Representations*. URL: https://openreview.net/forum?id=gmL46YMpu2J.

Dambanemuya, H. K. and N. Diakopoulos. (2021). "Auditing the Information Quality of News-Related Queries on the Alexa Voice Assistant". *Proceedings of the ACM on Human-Computer Interaction*. 5(CSCW1): 1–21.

Dash, A., A. Chakraborty, S. Ghosh, A. Mukherjee, and K. P. Gummadi. (2022). "Alexa, in you, I trust! Fairness and Interpretability Issues in E-commerce Search through Smart Speakers". In: *Proceedings of the ACM Web Conference 2022*. 3695–3705.

Deshpande, A., V. Murahari, T. Rajpurohit, A. Kalyan, and K. Narasimhan. (2023). "Toxicity in chatgpt: Analyzing persona-assigned language models". *arXiv preprint arXiv:2304.05335*.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186.

Diaz, F., B. Mitra, and N. Craswell. (2016). "Query Expansion with Locally-Trained Word Embeddings". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by K. Erk and N. A. Smith. Berlin, Germany: Association for Computational Linguistics. 367–377. DOI: 10.18653/v1/P16-1035.

Diaz, F., B. Mitra, M. D. Ekstrand, A. J. Biega, and B. Carterette. (2020). "Evaluating stochastic rankings with expected exposure". In: *CIKM*.

Drosou, M., H. V. Jagadish, E. Pitoura, and J. Stoyanovich. (2017). "Diversity in big data: A review". *Big data*. 5(2): 73–84.

Dudík, M., J. Langford, and L. Li. (2011). "Doubly Robust Policy Evaluation and Learning". In: *Proceedings of the 28th International Conference on International Conference on Machine Learning. ICML'11*. Bellevue, Washington, USA: Omnipress. 1097–1104.

Dupret, G. E. and B. Piwowarski. (2008). "A User Browsing Model to Predict Search Engine Click Data from Past Observations." In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '08*. Singapore, Singapore: Association for Computing Machinery. 331–338.

Dwork, C. (2006). "Differential privacy". In: *International colloquium on automata, languages, and programming*. Springer. 1–12.

Dwork, C., M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. (2012). "Fairness through awareness". In: *ITCS*. 214–226.

Dwork, C., M. P. Kim, O. Reingold, G. N. Rothblum, and G. Yona. (2019). "Learning from outcomes: Evidence-based rankings". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 106–125.

Eickhoff, C. (2018). “Cognitive Biases in Crowdsourcing”. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. WSDM ’18.* Marina Del Rey, CA, USA: Association for Computing Machinery. 162–170.

Ekstrand, M. D., A. Das, R. Burke, F. Diaz, *et al.* (2022). “Fairness in information access systems”. *Foundations and Trends® in Information Retrieval.* 16(1-2): 1–177.

Ekstrand, M. D., M. Tian, I. M. Azpiazu, J. D. Ekstrand, O. Anuyah, D. McNeill, and M. S. Pera. (2018). “All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness”. In: *Conference on fairness, accountability and transparency.* PMLR. 172–186.

Fang, H. and P. Xie. (2020). “CERT: Contrastive Self-supervised Learning for Language Understanding”. *CoRR.* abs/2005.12766.

Fang, Y., H. Liu, Z. Tao, and M. Yurochkin. (2022). “Fairness of Machine Learning in Search Engines”. In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management.* 5132–5135.

Fang, Z., A. Agarwal, and T. Joachims. (2019). “Intervention Harvesting for Context-Dependent Examination-Bias Estimation”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR’19.* Paris, France: Association for Computing Machinery. 825–834.

Feldman, M., S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian. (2015). “Certifying and removing disparate impact”. In: *KDD.*

Ferrara, E. (2023). “Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models”. *CoRR.* abs/2304.03738. DOI: 10.48550/arXiv.2304.03738. arXiv: 2304.03738.

Fleder, D. and K. Hosanagar. (2009). “Blockbuster culture’s next rise or fall: The impact of recommender systems on sales diversity”. *Management science.* 55(5): 697–712.

Foulds, J. R., R. Islam, K. N. Keya, and S. Pan. (2020). “An intersectional definition of fairness”. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE).* IEEE. 1918–1921.

Friedler, S. A., C. Scheidegger, and S. Venkatasubramanian. (2021). "The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making". *Communications of the ACM*. 64(4): 136–143.

Friedman, B. and H. Nissenbaum. (1996). "Bias in computer systems". *ACM Transactions on information systems (TOIS)*. 14(3): 330–347.

Gallegos, I. O., R. A. Rossi, J. Barrow, M. M. Tanjim, S. Kim, F. Dernoncourt, T. Yu, R. Zhang, and N. K. Ahmed. (2023). "Bias and Fairness in Large Language Models: A Survey". arXiv: 2309.00770 [cs.CL].

Ganguly, D., D. Roy, M. Mitra, and G. J. Jones. (2015). "Word Embedding Based Generalized Language Model for Information Retrieval". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '15*. Santiago, Chile: Association for Computing Machinery. 795–798.

Gao, R., Y. Ge, and C. Shah. (2022). "FAIR: Fairness-aware information retrieval evaluation". *Journal of the Association for Information Science and Technology*. 73(10): 1461–1473.

Gao, R. and C. Shah. (2019). "How fair can we go: Detecting the boundaries of fairness optimization in information retrieval". In: *Proceedings of the 2019 ACM SIGIR international conference on theory of information retrieval*. 229–236.

Gao, R. and C. Shah. (2020). "Toward creating a fairer ranking in search engine results". *Information Processing & Management*. 57(1): 102138.

Gao, R. and C. Shah. (2021). "Addressing bias and fairness in search systems". In: *SIGIR*.

Gao, T., A. Fisch, and D. Chen. (2021a). "Making Pre-trained Language Models Better Few-shot Learners". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 3816–3830.

Gao, T., A. Fisch, and D. Chen. (2021b). “Making Pre-trained Language Models Better Few-shot Learners”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Ed. by C. Zong, F. Xia, W. Li, and R. Navigli. Online: Association for Computational Linguistics. 3816–3830. DOI: 10.18653/v1/2021.acl-long.295.

Gao, Y., G. Huzhang, W. Shen, Y. Liu, W.-J. Zhou, Q. Da, and Y. Yu. (2021c). “Imitate TheWorld: A Search Engine Simulation Platform”. *arXiv preprint arXiv:2107.07693*.

Gerritse, E. J., F. Hasibi, and A. P. de Vries. (2020). “Bias in conversational search: The double-edged sword of the personalized knowledge graph”. In: *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 133–136.

Geyik, S. C., S. Ambler, and K. Kenthapadi. (2019). “Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search”. *KDD*.

Ghazimatin, A., M. Kleindessner, C. Russell, Z. Abedjan, and J. Golebiowski. (2022). “Measuring fairness of rankings under noisy sensitive information”. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2263–2279.

Gilotte, A., C. Calauzènes, T. Nedelec, A. Abraham, and S. Dollé. (2018). “Offline A/B Testing for Recommender Systems”. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. WSDM ’18*. Marina Del Rey, CA, USA: Association for Computing Machinery. 198–206.

Goldfarb-Tarrant, S., R. Marchant, R. Muñoz Sánchez, M. Pandya, and A. Lopez. (2021). “Intrinsic Bias Metrics Do Not Correlate with Application Bias”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics. 1926–1940. DOI: 10.18653/v1/2021.acl-long.150.

Gomroki, G., H. Behzadi, R. Fattahi, and J. S. Fadardi. (2023). “Identifying effective cognitive biases in information retrieval”. *Journal of Information Science*. 49(2): 348–358.

Gonen, H. and Y. Goldberg. (2019). "Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them". In: *Proceedings of the 2019 Workshop on Widening NLP*. Ed. by A. Axelrod, D. Yang, R. Cunha, S. Shaikh, and Z. Waseem. Florence, Italy: Association for Computational Linguistics. 60–63. URL: https://aclanthology.org/W19-3621.

Gorti, S. K., N. Vouitsis, J. Ma, K. Golestan, M. Volkovs, A. Garg, and G. Yu. (2022). "X-Pool: Cross-Modal Language-Video Attention for Text-Video Retrieval". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5006–5015.

Gu, Y., W. Bao, D. Ou, X. Li, B. Cui, B. Ma, H. Huang, Q. Liu, and X. Zeng. (2021). "Self-Supervised Learning on Users' Spontaneous Behaviors for Multi-Scenario Ranking in E-Commerce". In: *Proceedings of the 30th ACM International Conference on Information and Knowledge Management. CIKM '21*. 3828–3837.

Guo, F., C. Liu, A. Kannan, T. Minka, M. Taylor, Y.-M. Wang, and C. Faloutsos. (2009a). "Click Chain Model in Web Search". In: *Proceedings of the 18th International Conference on World Wide Web. WWW '09*. Madrid, Spain: Association for Computing Machinery. 11–20.

Guo, F., C. Liu, and Y. M. Wang. (2009b). "Efficient Multiple-Click Models in Web Search". In: *Proceedings of the Second ACM International Conference on Web Search and Data Mining. WSDM '09*. Barcelona, Spain: Association for Computing Machinery. 124–131.

Guo, H., J. Yu, Q. Liu, R. Tang, and Y. Zhang. (2019). "PAL: A Position-Bias Aware Learning Framework for CTR Prediction in Live Recommender Systems". In: *Proceedings of the 13th ACM Conference on Recommender Systems. RecSys '19*. Copenhagen, Denmark: Association for Computing Machinery. 452–456.

Guo, S., L. Zou, Y. Liu, W. Ye, S. Cheng, S. Wang, H. Chen, D. Yin, and Y. Chang. (2021). "Enhanced Doubly Robust Learning for Debiasing Post-Click Conversion Rate Estimation". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '21*. New York, NY, USA: Association for Computing Machinery. 275–284.

Gursoy, F., R. Kennedy, and I. Kakadiaris. (2022). "A critical assessment of the algorithmic accountability act of 2022". *Available at SSRN 4193199*.

Haak, F. (2023). "Investigation of Bias in Web Search Queries". In: *Advances in Information Retrieval: 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2–6, 2023, Proceedings, Part III*. Springer. 443–449.

Haak, F., B. Engelmann, C. K. Kreutz, and P. Schaer. (2024). "Investigating Bias in Political Search Query Suggestions by Relative Comparison with LLMs". In: *Companion Publication of the 16th ACM Web Science Conference*. 5–7.

Haak, F. and P. Schaer. (2021). "Perception-Aware Bias Detection for Query Suggestions". In: *Advances in Bias and Fairness in Information Retrieval: Second International Workshop on Algorithmic Bias in Search and Recommendation, BIAS 2021, Lucca, Italy, April 1, 2021, Proceedings*. Springer. 130–142.

Haak, F. and P. Schaer. (2022). "Auditing Search Query Suggestion Bias Through Recursive Algorithm Interrogation". In: *14th ACM Web Science Conference 2022*. 219–227.

Haak, F. and P. Schaer. (2023). "Qbias-A Dataset on Media Bias in Search Queries and Query Suggestions". In: *Proceedings of the 15th ACM Web Science Conference 2023*. 239–244.

Han, S., X. Wang, M. Bendersky, and M. Najork. (2020). "Learning-to-Rank with BERT in TF-Ranking". *CoRR*. abs/2004.08476. arXiv: 2004.08476. URL: https://arxiv.org/abs/2004.08476.

Hardt, M., X. Chen, X. Cheng, M. Donini, J. Gelman, S. Gollaprolu, J. He, P. Larroy, X. Liu, N. McCarthy, *et al.* (2021). "Amazon sagemaker clarify: Machine learning bias detection and explainability in the cloud". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2974–2983.

Hardt, M., E. Price, and N. Srebro. (2016). "Equality of opportunity in supervised learning". In: *NIPS*. 3315–3323.

Hashimoto, T., M. Srivastava, H. Namkoong, and P. Liang. (2018). "Fairness Without Demographics in Repeated Loss Minimization". In: *ICML*.

Hiemstra, D. (2023). "Was Fairness in IR Discussed by Cooper and Robertson in the 1970's?" In: *ACM SIGIR Forum*. Vol. 56. No. 2. ACM New York, NY, USA. 1–5.

Hsu, B., R. Mazumder, P. Nandy, and K. Basu. (2022). "Pushing the limits of fairness impossibility: Who's the fairest of them all?" *Advances in Neural Information Processing Systems*. 35: 32749–32761.

Hu, M., H. Wu, Z. Guan, R. Zhu, D. Guo, D. Qi, and S. Li. (2024). "No Free Lunch: Retrieval-Augmented Generation Undermines Fairness in LLMs, Even for Vigilant Users". *arXiv preprint arXiv:2410.07589*.

Hu, P., L. Zhen, D. Peng, and P. Liu. (2019). "Scalable Deep Multimodal Learning for Cross-Modal Retrieval". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'19*. Paris, France: Association for Computing Machinery. 635–644.

Hu, X., S. Yu, C. Xiong, Z. Liu, Z. Liu, and G. Yu. (2022). "P3 Ranker: Mitigating the Gaps between Pre-Training and Ranking Fine-Tuning with Prompt-Based Learning and Pre-Finetuning". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '22*. New York, NY, USA: Association for Computing Machinery. 1956–1962.

Huang, Z., H. Zeng, H. Zamani, and J. Allan. (2023). "Soft Prompt Decoding for Multilingual Dense Retrieval". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '23*. New York, NY, USA: Association for Computing Machinery. 1208–1218.

Hutchinson, B., V. Prabhakaran, E. Denton, K. Webster, Y. Zhong, and S. Denuyl. (2020). "Social Biases in NLP Models as Barriers for Persons with Disabilities". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Ed. by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault. Online: Association for Computational Linguistics. 5491–5501.

Introna, L. D. and H. Nissenbaum. (2000). "Shaping the Web: Why the politics of search engines matters". *The information society*. 16(3): 169–185.

Jacobs, A. Z. and H. Wallach. (2021). "Measurement and fairness". In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 375–385.

Jaenich, T., G. McDonald, and I. Ounis. (2023). "ColBERT-FairPRF: Towards Fair Pseudo-Relevance Feedback in Dense Retrieval". In: *European Conference on Information Retrieval*. Springer. 457–465.

Jain, A., M. Guo, K. Srinivasan, T. Chen, S. Kudugunta, C. Jia, Y. Yang, and J. Baldridge. (2021). "MURAL: Multimodal, Multitask Representations Across Languages". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*.

Ji, K., S. Pathiyan Cherumanal, J. R. Trippas, D. Hettiachchi, F. D. Salim, F. Scholer, and D. Spina. (2024). "Towards Detecting and Mitigating Cognitive Bias in Spoken Conversational Search". In: *Adjunct Proceedings of the 26th International Conference on Mobile Human-Computer Interaction*. 1–10.

Jiang, B., Z. Tan, A. Nirmal, and H. Liu. (2023). "Disinformation Detection: An Evolving Challenge in the Age of LLMs". *arXiv preprint arXiv:2309.15847*.

Jiang, N. and L. Li. (2016). "Doubly Robust Off-policy Value Evaluation for Reinforcement Learning". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by M. F. Balcan and K. Q. Weinberger. Vol. 48. *Proceedings of Machine Learning Research*. New York, New York, USA: PMLR. 652–661. URL: https://proceedings.mlr.press/v48/jiang16.html.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. (2007a). "Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search". *ACM TOIS*.

Joachims, T., A. Swaminathan, and T. Schnabel. (2017a). "Unbiased Learning-to-Rank with Biased Feedback". In: *WSDM*.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, and G. Gay. (2005). "Accurately Interpreting Clickthrough Data as Implicit Feedback". In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 154–161.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. (2007b). “Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search”. *ACM Transactions on Information Systems (TOIS)*. 25(2): 7.

Joachims, T., A. Swaminathan, and T. Schnabel. (2017b). “Unbiased Learning-to-Rank with Biased Feedback”. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 781–789.

Joachims, T., A. Swaminathan, and T. Schnabel. (2017c). “Unbiased learning-to-rank with biased feedback”. In: *Proceedings of the tenth ACM international conference on web search and data mining*. 781–789.

Kairouz, P., H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, *et al.* (2021). “Advances and open problems in federated learning”. *Foundations and Trends® in Machine Learning*. 14(1–2): 1–210.

Kallus, N. and A. Zhou. (2019). “The Fairness of Risk Scores Beyond Classification: Bipartite Ranking and the XAUC Metric”. *NeurIPS*.

Kamiran, F. and T. Calders. (2012). “Data preprocessing techniques for classification without discrimination”. *Knowledge and information systems*. 33(1): 1–33.

Kamishima, T., S. Akaho, H. Asoh, and J. Sakuma. (2018). “Recommendation independence”. In: *Conference on fairness, accountability and transparency*. PMLR. 187–201.

Kang, J. and H. Tong. (2021). “Fair graph mining”. In: *CIKM*.

Karako, C. and P. Manggala. (2018). “Using image fairness representations in diversity-based re-ranking for recommendations”. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 23–28.

Kay, M., C. Matuszek, and S. A. Munson. (2015). “Unequal representation and gender stereotypes in image search results for occupations”. In: *Proceedings of the 33rd annual acm conference on human factors in computing systems*. 3819–3828.

Kearns, M., S. Neel, A. Roth, and Z. S. Wu. (2019). “An empirical study of rich subgroup fairness for machine learning”. In: *Proceedings of the conference on fairness, accountability, and transparency*. 100–109.

Kearns, M., A. Roth, and Z. S. Wu. (2017). "Meritocratic fairness for cross-population selection". In: *International Conference on Machine Learning*. PMLR. 1828–1836.

Khramtsova, E., S. Zhuang, M. Baktashmotlagh, and G. Zuccon. (2024). "Leveraging LLMs for Unsupervised Dense Retriever Ranking". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '24*. Washington DC, USA: Association for Computing Machinery. 1307–1317.

Kiesel, J., D. Spina, H. Wachsmuth, and B. Stein. (2021). "The meant, the said, and the understood: Conversational argument search and cognitive biases". In: *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–5.

Kilbertus, N., M. R. Carulla, G. Parascandolo, M. Hardt, D. Janzing, and B. Schölkopf. (2017). "Avoiding discrimination through causal reasoning". In: *NIPS*. 656–666.

Kim, T. E. and F. Diaz. (2024). "Towards Fair RAG: On the Impact of Fair Ranking in Retrieval-Augmented Generation". *arXiv preprint arXiv:2409.11598*.

Kim, Y. (2014). "Convolutional Neural Networks for Sentence Classification". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1746–1751.

Kiyohara, H., Y. Saito, T. Matsuhiro, Y. Narita, N. Shimizu, and Y. Yamamoto. (2022). "Doubly Robust Off-Policy Evaluation for Ranking Policies under the Cascade Behavior Model". In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. WSDM '22*. Virtual Event, AZ, USA: Association for Computing Machinery. 487–497.

Kiyohara, H., M. Uehara, Y. Narita, N. Shimizu, Y. Yamamoto, and Y. Saito. (2023). "Off-Policy Evaluation of Ranking Policies under Diverse User Behavior". In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '23*. New York, NY, USA: Association for Computing Machinery. 1154–1163.

Kırnap, Ö., F. Diaz, A. Biega, M. Ekstrand, B. Carterette, and E. Yilmaz. (2021). "Estimation of Fair Ranking Metrics with Incomplete Judgments". In: *Proceedings of the Web Conference 2021*. 1065–1075.

Kleinberg, J. (2018). "Inherent trade-offs in algorithmic fairness". In: *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*. 40–40.

Kleinberg, J., S. Mullainathan, and M. Raghavan. (2016). "Inherent trade-offs in the fair determination of risk scores". *arXiv preprint arXiv:1609.05807*.

Kleinberg, J., S. Mullainathan, and M. Raghavan. (2017). "Inherent Trade-Offs in the Fair Determination of Risk Scores". In: *ITCS*.

Koenecke, A., A. Nam, E. Lake, J. Nudell, M. Quartey, Z. Mengesha, C. Toups, J. R. Rickford, D. Jurafsky, and S. Goel. (2020). "Racial disparities in automated speech recognition". *Proceedings of the National Academy of Sciences*. 117(14): 7684–7689.

Kopeinik, S., M. Mara, L. Ratz, K. Krieg, M. Schedl, and N. Rekabsaz. (2023). "Show me a" Male Nurse"! How Gender Bias is Reflected in the Query Formulation of Search Engine Users".

Krieg, K., E. Parada-Cabaleiro, G. Medicus, O. Lesota, M. Schedl, and N. Rekabsaz. (2022a). "Grep-BiasIR: a dataset for investigating gender representation-bias in information retrieval results". In: *Proceeding of the 2023 ACM SIGIR Conference On Human Information Interaction And Retrieval (CHIIR)*.

Krieg, K., E. Parada-Cabaleiro, G. Medicus, O. Lesota, M. Schedl, and N. Rekabsaz. (2023). "Grep-BiasIR: A Dataset for Investigating Gender Representation Bias in Information Retrieval Results". In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery. 444–448.

Krieg, K., E. Parada-Cabaleiro, M. Schedl, and N. Rekabsaz. (2022b). "Do Perceived Gender Biases in Retrieval Results Affect Relevance Judgements?" In: *Advances in Bias and Fairness in Information Retrieval*. Springer International Publishing. 104–116.

Kurita, K., N. Vyas, A. Pareek, A. W. Black, and Y. Tsvetkov. (2019). "Measuring Bias in Contextualized Word Representations". In: *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*. Ed. by M. R. Costa-jussà, C. Hardmeier, W. Radford, and K. Webster. Florence, Italy: Association for Computational Linguistics. 166–172. DOI: 10.18653/v1/W19-3823.

Kusner, M., Y. Sun, N. Kolkin, and K. Weinberger. (2015). "From Word Embeddings To Document Distances". In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by F. Bach and D. Blei. Vol. 37. *Proceedings of Machine Learning Research*. Lille, France: PMLR. 957–966. URL: https://proceedings.mlr.press/v37/kusnerb15.html.

Kusner, M. J., J. Loftus, C. Russell, and R. Silva. (2017). "Counterfactual fairness". *Advances in neural information processing systems*. 30.

Lahoti, P., A. Beutel, J. Chen, K. Lee, F. Prost, N. Thain, X. Wang, and E. Chi. (2020). "Fairness without demographics through adversarially reweighted learning". *Advances in neural information processing systems*. 33: 728–740.

Lahoti, P., K. P. Gummadi, and G. Weikum. (2019). "Operationalizing individual fairness with pairwise fair representations". *VLDB Endowment*.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. (2020). "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations". In: *ICLR*.

Lazovich, T., L. Belli, A. Gonzales, A. Bower, U. Tantipongpipat, K. Lum, F. Huszar, and R. Chowdhury. (2022). "Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics". *Patterns*. 3(8).

Le, Q. and T. Mikolov. (2014). "Distributed Representations of Sentences and Documents". In: *Proceedings of the 31st International Conference on Machine Learning*. 1188–1196.

Le Quy, T., A. Roy, V. Iosifidis, W. Zhang, and E. Ntoutsi. (2022). "A survey on datasets for fairness-aware machine learning". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 12(3): e1452.

Lee, N., Y. Bang, H. Lovenia, S. Cahyawijaya, W. Dai, and P. Fung. (2023). "Survey of Social Bias in Vision-Language Models". arXiv: 2309.14381 [cs.CL].

Lester, B., R. Al-Rfou, and N. Constant. (2021). "The Power of Scale for Parameter-Efficient Prompt Tuning". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 3045–3059.

Li, J., D. Li, C. Xiong, and S. Hoi. (2022). "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation". In: *ICML*.

Li, L., S. Chen, J. Kleban, and A. Gupta. (2015). "Counterfactual Estimation and Optimization of Click Metrics in Search Engines: A Case Study". In: *Proceedings of the 24th International Conference on World Wide Web. WWW '15 Companion*. Association for Computing Machinery. 929–934.

Li, L. H., M. Yatskar, D. Yin, C.-J. Hsieh, and K.-W. Chang. (2019). "VisualBERT: A Simple and Performant Baseline for Vision and Language". In: *Arxiv*.

Li, S., Y. Abbasi-Yadkori, B. Kveton, S. Muthukrishnan, V. Vinay, and Z. Wen. (2018). "Offline Evaluation of Ranking Policies with Click Models". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18*. London, United Kingdom: Association for Computing Machinery. 1685–1694.

Li, Y., H. Chen, S. Xu, Y. Ge, J. Tan, S. Liu, and Y. Zhang. (2023). "Fairness in Recommendation: Foundations, Methods, and Applications". *ACM Transactions on Intelligent Systems and Technology*. 14(5): 1–48.

Li, Y., H. Chen, S. Xu, Y. Ge, and Y. Zhang. (2021a). "Towards personalized fairness based on causal notion". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1054–1063.

Li, Y., Y. Ge, and Y. Zhang. (2021b). "Tutorial on fairness of machine learning in recommender systems". In: *SIGIR*.

Lima, L., V. Furtado, E. Furtado, and V. Almeida. (2019). “Empirical analysis of bias in voice-based personal assistants”. In: *Companion Proceedings of the 2019 World Wide Web Conference*. 533–538.

Liu, H. and X. Zhao. (2021). “KDD 2021 IRS Workshop”. In:

Liu, H., C. Li, Q. Wu, and Y. J. Lee. (2023). “Visual Instruction Tuning”. In: *Advances in Neural Information Processing Systems*. Vol. 36. 34892–34916.

Liu, T.-Y. *et al.* (2009). “Learning to rank for information retrieval”. *Foundations and Trends® in Information Retrieval*. 3(3): 225–331.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. (2020). “Ro{BERT}a: A Robustly Optimized {BERT} Pretraining Approach”. URL: https://openreview.net/forum?id=SyxS0T4tvS.

Lu, J., D. Batra, D. Parikh, and S. Lee. (2019). “ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc.

Lu, J., G. Hernandez Abrego, J. Ma, J. Ni, and Y. Yang. (2021). “Multi-stage Training with Improved Negative Contrast for Neural Passage Retrieval”. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Ed. by M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics. 6091–6103. DOI: 10.18653/v1/2021.emnlp-main.492.

Luo, D., L. Zou, Q. Ai, Z. Chen, D. Yin, and B. D. Davison. (2023). “Model-Based Unbiased Learning to Rank”. In: *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 895–903.

Ma, H., S. Guan, C. Toomey, and Y. Wu. (2022a). “Diversified subgraph query generation with group fairness”. In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 686–694.

Ma, H., S. Guan, M. Wang, Y.-s. Chang, and Y. Wu. (2022b). "Subgraph query generation with fairness and diversity constraints". In: *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE. 3106–3118.

Ma, H., H. Zhao, Z. Lin, A. Kale, Z. Wang, T. Yu, J. Gu, S. Choudhary, and X. Xie. (2022c). "EI-CLIP: Entity-Aware Interventional Contrastive Learning for E-Commerce Cross-Modal Retrieval". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 18051–18061.

Ma, X., X. Zhang, R. Pradeep, and J. Lin. (2023). "Zero-Shot Listwise Document Reranking with a Large Language Model". arXiv: 2305.02156 [cs.IR].

Madaio, M., L. Egede, H. Subramonyam, J. Wortman Vaughan, and H. Wallach. (2022). "Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support". *Proceedings of the ACM on Human-Computer Interaction*. 6(CSCW1): 1–26.

Makhlouf, K., S. Zhioua, and C. Palamidessi. (2020). "Survey on causal-based machine learning fairness notions". *arXiv preprint arXiv:2010.09553*.

Makhortykh, M., A. Urman, and R. Ulloa. (2021). "Detecting race and gender bias in visual representation of AI on web search engines". In: *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer. 36–50.

Mandal, A., S. Leavy, and S. Little. (2021). "Dataset diversity: measuring and mitigating geographical bias in image search and retrieval".

May, C., A. Wang, S. Bordia, S. R. Bowman, and R. Rudinger. (2019). "On Measuring Social Biases in Sentence Encoders". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Ed. by J. Burstein, C. Doran, and T. Solorio. Minneapolis, Minnesota: Association for Computational Linguistics. 622–628. DOI: 10.18653/v1/N19-1063.

McInerney, J., B. Brost, P. Chandar, R. Mehrotra, and B. Carterette. (2020). "Counterfactual Evaluation of Slate Recommendations with Sequential Reward Interactions". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '20*. Virtual Event, CA, USA: Association for Computing Machinery. 1779–1788.

Meade, N., E. Poole-Dayan, and S. Reddy. (2022). "An Empirical Survey of the Effectiveness of Debiasing Techniques for Pre-trained Language Models". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Mehrotra, A. and N. Vishnoi. (2022). "Fair ranking with noisy protected attributes". *Advances in Neural Information Processing Systems*. 35: 31711–31725.

Mehrotra, R., A. Anderson, F. Diaz, A. Sharma, H. Wallach, and E. Yilmaz. (2017). "Auditing search engines for differential satisfaction across demographics". In: *Proceedings of the 26th international conference on World Wide Web companion*. 626–633.

Mehrotra, R., J. McInerney, H. Bouchard, M. Lalmas, and F. Diaz. (2018). "Towards a Fair Marketplace: Counterfactual Evaluation of the Trade-off Between Relevance, Fairness & Satisfaction in Recommendation Systems". In: *CIKM*. ACM.

Minka, T. and S. Robertson. (2008). "Selection Bias in the LETOR Datasets". In: URL: https://www.microsoft.com/en-us/research/publication/selection-bias-letor-datasets/.

Mitchell, S., E. Potash, S. Barocas, A. D'Amour, and K. Lum. (2021). "Algorithmic fairness: Choices, assumptions, and definitions". *Annual Review of Statistics and Its Application*. 8: 141–163.

Mitra, B. (2024). "Search and Society: Reimagining Information Access for Radical Futures". *arXiv preprint arXiv:2403.17901*.

Mitra, B. and N. Craswell. (2017). "Neural models for information retrieval". *arXiv preprint arXiv:1705.01509*.

Mitra, B., S. Hofstätter, H. Zamani, and N. Craswell. (2021). "Improving Transformer-Kernel Ranking Model Using Conformer and Query Term Independence". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '21*. New York, NY, USA: Association for Computing Machinery. 1697–1702.

Morik, M., A. Singh, J. Hong, and T. Joachims. (2020). "Controlling fairness and bias in dynamic learning-to-rank". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 429–438.

Mowshowitz, A. and A. Kawaguchi. (2002). "Assessing bias in search engines". *Information Processing & Management*. 38(1): 141–156.

Nabi, R. and I. Shpitser. (2018). "Fair inference on outcomes". *AAAI*.

Nadeem, M., A. Bethke, and S. Reddy. (2021). "StereoSet: Measuring stereotypical bias in pretrained language models". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Ed. by C. Zong, F. Xia, W. Li, and R. Navigli. Online: Association for Computational Linguistics. 5356–5371. DOI: 10.18653/v1/2021.acl-long.416.

Nag, P. and Ö. N. Yalçın. (2020). "Gender stereotypes in virtual agents". In: *Proceedings of the 20th ACM International conference on intelligent virtual agents*. 1–8.

Nalisnick, E., B. Mitra, N. Craswell, and R. Caruana. (2016). "Improving Document Ranking with Dual Word Embeddings". In: *Proceedings of the 25th International Conference Companion on World Wide Web. WWW '16 Companion*. Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee. 83–84.

Nangia, N., C. Vania, R. Bhalerao, and S. R. Bowman. (2020). "CrowS-Pairs: A Challenge Dataset for Measuring Social Biases in Masked Language Models". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Ed. by B. Webber, T. Cohn, Y. He, and Y. Liu. Online: Association for Computational Linguistics. 1953–1967. DOI: 10.18653/v1/2020.emnlp-main.154.

Narasimhan, H., A. Cotter, M. Gupta, and S. Wang. (2020). "Pairwise fairness for ranking and regression". In: *AAAI*.

Nogueira, R. and K. Cho. (2019). "Passage Re-ranking with BERT". *ArXiv*. abs/1901.04085. URL: https://api.semanticscholar.org/CorpusID:58004692.

Nogueira, R., Z. Jiang, R. Pradeep, and J. Lin. (2020). "Document Ranking with a Pretrained Sequence-to-Sequence Model". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics. 708–718. DOI: 10.18653/v1/2020.findings-emnlp.63.

Nogueira, R. F., W. Yang, K. Cho, and J. Lin. (2019). "Multi-Stage Document Ranking with BERT". *CoRR*. abs/1910.14424. arXiv: 1910.14424. URL: http://arxiv.org/abs/1910.14424.

Nozza, D., F. Bianchi, A. Lauscher, and D. Hovy. (2022). "Measuring Harmful Sentence Completion in Language Models for LGBTQIA+ Individuals". In: *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Dublin, Ireland: Association for Computational Linguistics. 26–34. DOI: 10.18653/v1/2022.ltedi-1.4.

O'Brien, M. and M. T. Keane. (2006). "Modeling result-list searching in the World Wide Web: The role of relevance topologies and trust bias". In: *Proceedings of the 28th annual conference of the cognitive science society*. Vol. 28. 1881–1886.

Olteanu, A., F. Diaz, and G. Kazai. (2020). "When are search completion suggestions problematic?" *Proceedings of the ACM on human-computer interaction*. 4(CSCW2): 1–25.

Olteanu, A., J. Garcia-Gathright, M. de Rijke, M. D. Ekstrand, A. Roegiest, A. Lipani, A. Beutel, A. Olteanu, A. Lucic, A.-A. Stoica, *et al.* (2021). "FACTS-IR: fairness, accountability, confidentiality, transparency, and safety in information retrieval". In: *ACM SIGIR Forum*. Vol. 53. No. 2. ACM New York, NY, USA. 20–43.

Oord, A. van den, Y. Li, and O. Vinyals. (2019). "Representation Learning with Contrastive Predictive Coding". arXiv: 1807.03748 [cs.LG].

Oosterhuis, H. (2022). “Reaching the End of Unbiasedness: Uncovering Implicit Limitations of Click-Based Learning to Rank”. In: *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval. ICTIR ’22*. Madrid, Spain: Association for Computing Machinery. 264–274.

Oosterhuis, H. (2023). “Doubly Robust Estimation for Correcting Position Bias in Click Feedback for Unbiased Learning to Rank”. *ACM Trans. Inf. Syst.* 41(3).

Oosterhuis, H. and M. de Rijke. (2020). “Policy-Aware Unbiased Learning to Rank for Top-k Rankings”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.

OpenAI. (2023). “GPT-4 Technical Report”. arXiv: 2303.08774 [cs.CL].

Otterbacher, J., J. Bates, and P. Clough. (2017). “Competent men and warm women: Gender stereotypes and backlash in image search results”. In: *Proceedings of the 2017 chi conference on human factors in computing systems*. 6620–6631.

Ovaisi, Z., R. Ahsan, Y. Zhang, K. Vasilaky, and E. Zheleva. (2020). “Correcting for Selection Bias in Learning-to-Rank Systems”. In: *Proceedings of The Web Conference 2020*. 1863–1873.

Page, L., S. Brin, R. Motwani, and T. Winograd. (1998). “The pagerank citation ranking: Bring order to the web”. *Tech. rep.* Technical report, stanford University.

Papakyriakopoulos, O., S. Hegelich, J. C. M. Serrano, and F. Marco. (2020). “Bias in Word Embeddings”. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 446–457.

Patro, G. K., L. Porcaro, L. Mitchell, Q. Zhang, M. Zehlike, and N. Garg. (2022). “Fair ranking: a critical review, challenges, and future directions”. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1929–1942.

Penha, G., E. Palumbo, M. Aziz, A. Wang, and H. Bouchard. (2023). “Improving content retrievability in search with controllable query generation”. In: *Proceedings of the ACM Web Conference 2023*. 3182–3192.

Perez, E., S. Huang, F. Song, T. Cai, R. Ring, J. Aslanides, A. Glaese, N. McAleese, and G. Irving. (2022). "Red Teaming Language Models with Language Models". In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Ed. by Y. Goldberg, Z. Kozareva, and Y. Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. 3419–3448.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018). "Deep Contextualized Word Representations". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Ed. by M. Walker, H. Ji, and A. Stent. New Orleans, Louisiana: Association for Computational Linguistics. 2227–2237. DOI: 10.18653/v1/N18-1202.

Pinney, C., A. Raj, A. Hanna, and M. D. Ekstrand. (2023). "Much Ado about gender: Current practices and future recommendations for appropriate gender-aware information access". In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 269–279.

Pitoura, E., K. Stefanidis, and G. Koutrika. (2021). "Fairness in Rankings and Recommendations: An Overview". *The VLDB Journal*. 31(3): 431–458. DOI: 10.1007/s00778-021-00697-y.

Pleiss, G., M. Raghavan, F. Wu, J. Kleinberg, and K. Q. Weinberger. (2017). "On fairness and calibration". *Advances in neural information processing systems*. 30.

Pradeep, R., S. Sharifymoghaddam, and J. Lin. (2023). "Rankvicuna: Zero-shot listwise document reranking with open-source large language models". *arXiv preprint arXiv:2309.15088*.

Pradel, F., F. Haak, S.-O. Proksch, and P. Schaer. (2024). "Dynamics in Search Engine Query Suggestions for European Politicians". In: *Proceedings of the 16th ACM Web Science Conference*. 279–289.

Precup, D., R. S. Sutton, and S. P. Singh. (2000). "Eligibility Traces for Off-Policy Policy Evaluation". In: *Proceedings of the Seventeenth International Conference on Machine Learning*. *ICML '00*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 759–766.

Qiao, Y., C. Xiong, Z. Liu, and Z. Liu. (2019). "Understanding the Behaviors of BERT in Ranking". *CoRR*. abs/1904.07531. arXiv: 1904.07531. URL: http://arxiv.org/abs/1904.07531.

Qin, Z., R. Jagerman, K. Hui, H. Zhuang, J. Wu, J. Shen, T. Liu, J. Liu, D. Metzler, X. Wang, and M. Bendersky. (2023). "Large Language Models are Effective Text Rankers with Pairwise Ranking Prompting". arXiv: 2306.17563 [cs.IR].

Quiñonero Candela, J., Y. Wu, B. Hsu, S. Jain, J. Ramos, J. Adams, R. Hallman, and K. Basu. (2023). "Disentangling and Operationalizing AI Fairness at LinkedIn". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1213–1228.

Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. (2021). "Learning Transferable Visual Models From Natural Language Supervision". In: *Proceedings of the 38th International Conference on Machine Learning*. 8748–8763.

Radford, A. and K. Narasimhan. (2018). "Improving Language Understanding by Generative Pre-Training".

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et al.* (2019). "Language models are unsupervised multitask learners". *OpenAI blog*. 1(8): 9.

Radlinski, F., P. N. Bennett, B. Carterette, and T. Joachims. (2009). "Redundancy, diversity and interdependent document relevance". In: *ACM SIGIR Forum*. Vol. 43. 46–52.

Radlinski, F., R. Kleinberg, and T. Joachims. (2008). "Learning diverse rankings with multi-armed bandits". In: *ICML*. ACM. 784–791.

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". *J. Mach. Learn. Res.* 21(1).

Rahmani, H. A., C. Siro, M. Aliannejadi, N. Craswell, C. L. Clarke, G. Faggioli, B. Mitra, P. Thomas, and E. Yilmaz. (2024). "Llm4eval: Large language model for evaluation in ir". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3040–3043.

Raj, A. and M. D. Ekstrand. (2022). "Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '22*. Madrid, Spain: Association for Computing Machinery. 726–736.

Raj, A., B. Mitra, N. Craswell, and M. D. Ekstrand. (2023). "Patterns of gender-specializing query reformulation". *SIGIR*.

Rastegarpanah, B., K. Gummadi, and M. Crovella. (2021). "Auditing black-box prediction models for data minimization compliance". *Advances in Neural Information Processing Systems*. 34: 20621–20632.

Ratz, L., M. Schedl, S. Kopeinik, and N. Rekabsaz. (2024). "Measuring Bias in Search Results Through Retrieval List Comparison". In: *Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings, Part V*. Glasgow, United Kingdom: Springer-Verlag. 20–34. DOI: 10.1007/978-3-031-56069-9_2.

Reimers, N. and I. Gurevych. (2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 3982–3992.

Rekabsaz, N., S. Kopeinik, and M. Schedl. (2021). "Societal Biases in Retrieved Contents: Measurement Framework and Adversarial Mitigation of BERT Rankers". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '21*. New York, NY, USA: Association for Computing Machinery. 306–316.

Rekabsaz, N. and M. Schedl. (2020). "Do Neural Ranking Models Intensify Gender Bias?" In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2065–2068.

Ren, Y., H. Tang, and S. Zhu. (2022). "Unbiased Learning to Rank with Biased Continuous Feedback". In: *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*. 1716–1725.

Richardson, M., E. Dominowska, and R. Ragno. (2007). "Predicting Clicks: Estimating the Click-through Rate for New Ads". In: *Proceedings of the 16th International Conference on World Wide Web. WWW '07.* Banff, Alberta, Canada: Association for Computing Machinery. 521–530.

Robertson, S. E. (1977). "The probability ranking principle in IR". *Journal of documentation.*

Ross, C., B. Katz, and A. Barbu. (2021). "Measuring Social Biases in Grounded Vision and Language Embeddings". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Ed. by K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou. Online: Association for Computational Linguistics. 998–1008. DOI: 10.18653/v1/2021.naacl-main.78.

Roy, D., D. Ganguly, M. Mitra, and G. Jones. (2016). "Representing Documents and Queries as Sets of Word Embedded Vectors for Information Retrieval". *ArXiv.* abs/1606.07869. URL: https://api.semanticscholar.org/CorpusID:15987308.

Saito, Y. (2020a). "Asymmetric Tri-Training for Debiasing Missing-Not-At-Random Explicit Feedback". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '20.* Virtual Event, China: Association for Computing Machinery. 309–318.

Saito, Y. (2020b). "Doubly Robust Estimator for Ranking Metrics with Post-Click Conversions". In: *Proceedings of the 14th ACM Conference on Recommender Systems. RecSys '20.* Virtual Event, Brazil: Association for Computing Machinery. 92–100.

Saito, Y. and T. Joachims. (2021). "Counterfactual Learning and Evaluation for Recommender Systems: Foundations, Implementations, and Recent Advances". In: *Proceedings of the 15th ACM Conference on Recommender Systems. RecSys '21.* Amsterdam, Netherlands: Association for Computing Machinery. 828–830.

Saito, Y. and T. Joachims. (2022). "Off-Policy Evaluation for Large Action Spaces via Embeddings". In: *Proceedings of the 39th International Conference on Machine Learning.* PMLR. 19089–19122.

Samar, T., M. C. Traub, J. van Ossenbruggen, L. Hardman, and A. P. de Vries. (2018). "Quantifying retrieval bias in Web archive search". *International Journal on Digital Libraries*. 19: 57–75.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf. (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". *CoRR*. abs/1910.01108. URL: http://arxiv.org/abs/1910.01108.

Sap, M., D. Card, S. Gabriel, Y. Choi, and N. A. Smith. (2019). "The Risk of Racial Bias in Hate Speech Detection". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Ed. by A. Korhonen, D. Traum, and L. Màrquez. Florence, Italy: Association for Computational Linguistics. 1668–1678. DOI: 10.18653/v1/P19-1163.

Sapiezynski, P., W. Zeng, R. E Robertson, A. Mislove, and C. Wilson. (2019). "Quantifying the Impact of User Attentionon Fair Group Representation in Ranked Lists". In: *Companion Proceedings of The 2019 World Wide Web Conference. WWW '19*. San Francisco, USA: Association for Computing Machinery. 553–562.

Schick, T., S. Udupa, and H. Schütze. (2021). "Self-Diagnosis and Self-Debiasing: A Proposal for Reducing Corpus-Based Bias in NLP". *Computing Research Repository*. arXiv:2103.00453. URL: http://arxiv.org/abs/2103.00453.

Schnabel, T., A. Swaminathan, A. Singh, N. Chandak, and T. Joachims. (2016). "Recommendations as Treatments: Debiasing Learning and Evaluation". In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. ICML'16*. JMLR.org. 1670–1679.

Scholer, F., D. Kelly, W.-C. Wu, H. S. Lee, and W. Webber. (2013). "The Effect of Threshold Priming and Need for Cognition on Relevance Calibration and Assessment". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '13*. Dublin, Ireland: Association for Computing Machinery. 623–632.

Schramowski, P., C. Turan, N. Andersen, C. A. Rothkopf, and K. Kersting. (2022). "Large pre-trained language models contain human-like biases of what is right and wrong to do". *Nature Machine Intelligence*. 4(3): 258–268. DOI: 10.1038/s42256-022-00458-8.

Segev, E. (2010). *Google and the digital divide: The bias of online knowledge*. Elsevier.

Sesari, E., M. Hort, and F. Sarro. (2022). "An Empirical Study on the Fairness of Pre-trained Word Embeddings". In: *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*. Ed. by C. Hardmeier, C. Basta, M. R. Costa-jussà, G. Stanovsky, and H. Gonen. Seattle, Washington: Association for Computational Linguistics. 129–144. DOI: 10.18653/v1/2022.gebnlp-1.15.

Severyn, A. and A. Moschitti. (2015). "Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 373–382.

Seyedsalehi, S., A. Bigdeli, N. Arabzadeh, B. Mitra, M. Zihayat, and E. Bagheri. (2022). "Bias-aware Fair Neural Ranking for Addressing Stereotypical Gender Biases". In: *Proceedings of the 25th International Conference on Extending Database Technology, EDBT 2022, Edinburgh, UK, March 29 - April 1, 2022*. OpenProceedings.org. 2:435–2:439.

Seymour, W., X. Zhan, M. Cote, and J. Such. (2023). "A Systematic Review of Ethical Concerns with Voice Assistants". *AAAI/ACM Conference on AI, Ethics, and Society*.

Shi, J.-C., Y. Yu, Q. Da, S.-Y. Chen, and A.-X. Zeng. (2019). "Virtualtaobao: Virtualizing real-world online retail environment for reinforcement learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 4902–4909.

Shokouhi, M., R. White, and E. Yilmaz. (2015). "Anchoring and Adjustment in Relevance Estimation". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '15*. Santiago, Chile: Association for Computing Machinery. 963–966.

Silva, A., P. Tambwekar, and M. Gombolay. (2021). "Towards a Comprehensive Understanding and Accurate Evaluation of Societal Biases in Pre-Trained Transformers". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics. 2383–2389. DOI: 10.18653/v1/2021.naacl-main.189.

Singh, A. and T. Joachims. (2017). "Equality of Opportunity in Rankings". In: *NIPS Workshop on Prioritizing Online Content (WPOC)*.

Singh, A. and T. Joachims. (2018). "Fairness of exposure in rankings". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2219–2228.

Singh, A. and T. Joachims. (2019). "Policy learning for fairness in ranking". *Advances in Neural Information Processing Systems*. 32.

Singh, A., D. Kempe, and T. Joachims. (2021). "Fairness in ranking under uncertainty". *Advances in Neural Information Processing Systems*. 34: 11896–11908.

Steed, R., S. Panda, A. Kobren, and M. Wick. (2022). "Upstream Mitigation Is *Not* All You Need: Testing the Bias Transfer Hypothesis in Pre-Trained Language Models". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by S. Muresan, P. Nakov, and A. Villavicencio. Dublin, Ireland: Association for Computational Linguistics. 3524–3542. DOI: 10.18653/v1/2022.acl-long.247.

Strehl, A. L., J. Langford, L. Li, and S. M. Kakade. (2010). "Learning from Logged Implicit Exploration Data". In: *Proceedings of the 23rd International Conference on Neural Information Processing Systems. NIPS'10*. Vancouver, British Columbia, Canada: Curran Associates Inc. 2217–2225.

Su, W., X. Zhu, Y. Cao, B. Li, L. Lu, F. Wei, and J. Dai. (2020a). "VL-BERT: Pre-training of Generic Visual-Linguistic Representations". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=SygXPaEYvH.

Su, Y., M. Dimakopoulou, A. Krishnamurthy, and M. Dudik. (2020b). "Doubly robust off-policy evaluation with shrinkage". In: *Proceedings of the 37th International Conference on Machine Learning*. Ed. by H. D. III and A. Singh. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 9167–9176. URL: https://proceedings.mlr.press/v119/su20a.html.

Su, Y., L. Wang, M. Santacatterina, and T. Joachims. (2019). "CAB: Continuous Adaptive Blending for Policy Evaluation and Learning". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 6005–6014. URL: https://proceedings.mlr.press/v97/su19a.html.

Sun, G., Y. Bai, X. Yang, Y. Fang, Y. Fu, and Z. Tao. (2024a). "Aligning Out-of-Distribution Web Images and Caption Semantics via Evidential Learning". In: *Proceedings of the ACM on Web Conference 2024*.

Sun, G., C. Qin, J. Wang, Z. Chen, R. Xu, and Z. Tao. (2024b). "SQ-LLaVA: Self-Questioning for Large Vision-Language Assistant". In: *The European Conference on Computer Vision (ECCV)*.

Sun, W., L. Yan, X. Ma, P. Ren, D. Yin, and Z. Ren. (2023). "Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agent". *ArXiv*. abs/2304.09542.

Sutskever, I., O. Vinyals, and Q. V. Le. (2014). "Sequence to Sequence Learning with Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 27.

Swaminathan, A. and T. Joachims. (2015a). "Batch Learning from Logged Bandit Feedback through Counterfactual Risk Minimization". *Journal of Machine Learning Research*. 16(52): 1731–1755. URL: http://jmlr.org/papers/v16/swaminathan15a.html.

Swaminathan, A. and T. Joachims. (2015b). "The Self-Normalized Estimator for Counterfactual Learning". In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett. Vol. 28. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2015/file/39027dfad5138c9ca0c474d71db915c3-Paper.pdf.

Sweeney, L. (2013). "Discrimination in online ad delivery". *Communications of the ACM*. 56(5): 44–54.

Tam, W., X. Liu, K. Ji, L. Xue, J. Liu, T. Li, Y. Dong, and J. Tang. (2023). "Parameter-Efficient Prompt Tuning Makes Generalized and Calibrated Neural Text Retrievers". In: *Findings of the Association for Computational Linguistics: EMNLP 2023*. Ed. by H. Bouamor, J. Pino, and K. Bali. Singapore: Association for Computational Linguistics. 13117–13130. DOI: 10.18653/v1/2023.findings-emnlp.874.

Tan, H. and M. Bansal. (2019). "LXMERT: Learning Cross-Modality Encoder Representations from Transformers". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Ed. by K. Inui, J. Jiang, V. Ng, and X. Wan. Hong Kong, China: Association for Computational Linguistics. 5100–5111. DOI: 10.18653/v1/D19-1514.

Thomas, P., S. Spielman, N. Craswell, and B. Mitra. (2024). "Large language models can accurately predict searcher preferences". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1930–1940.

Touvron, H., T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, *et al.* (2023). "Llama: Open and efficient foundation language models". *arXiv preprint arXiv:2302.13971*.

Tversky, A. and D. Kahneman. (1974). "Judgment under Uncertainty: Heuristics and Biases". *Science*. 185(4157): 1124–1131.

Tversky, A. and D. Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty". *Journal of Risk and Uncertainty*. 5(4): 297–323.

Vardasbi, A., H. Oosterhuis, and M. de Rijke. (2020). "When Inverse Propensity Scoring Does Not Work: Affine Corrections for Unbiased Learning to Rank". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management. CIKM '20*. Virtual Event, Ireland: Association for Computing Machinery. 1475–1484.

Vassimon Manela, D. de, D. Errington, T. Fisher, B. van Breugel, and P. Minervini. (2021). "Stereotype and Skew: Quantifying Gender Bias in Pre-trained and Fine-tuned Language Models". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Online: Association for Computational Linguistics. 2232–2242. DOI: 10.18653/v1/2021.eacl-main.190.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30.

Vaughan, L. and M. Thelwall. (2004). "Search engine coverage bias: evidence and possible causes". *Information processing & management*. 40(4): 693–707.

Vaughan, L. and Y. Zhang. (2007). "Equal representation by search engines? A comparison of websites across countries and domains". *Journal of computer-mediated communication*. 12(3): 888–909.

Vig, J., S. Gehrmann, Y. Belinkov, S. Qian, D. Nevo, Y. Singer, and S. Shieber. (2020). "Investigating Gender Bias in Language Models Using Causal Mediation Analysis". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc. 12388–12401. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/92650b2e92217715fe312e6fa7b90d82-Paper.pdf.

Voorhees, E. M. (2002). "The Philosophy of Information Retrieval Evaluation". In: *Evaluation of Cross-Language Information Retrieval Systems*. Ed. by C. Peters, M. Braschler, J. Gonzalo, and M. Kluck. Berlin, Heidelberg: Springer Berlin Heidelberg. 355–370.

Wang, J., Y. Liu, and X. E. Wang. (2021a). "Are gender-neutral queries really gender-neutral? mitigating gender bias in image search". *arXiv preprint arXiv:2109.05433*.

Wang, J., X. Hu, W. Hou, H. Chen, R. Zheng, Y. Wang, L. Yang, H. Huang, W. Ye, X. Geng, B. Jiao, Y. Zhang, and X. Xie. (2023a). "On the Robustness of ChatGPT: An Adversarial and Out-of-distribution Perspective". *CoRR*. abs/2302.12095. DOI: 10.48550/arXiv.2302.12095. arXiv: 2302.12095.

Wang, L. and T. Joachims. (2021). "User fairness, item fairness, and diversity for rankings in two-sided markets". In: *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*. 23–41.

Wang, N., Z. Qin, X. Wang, and H. Wang. (2021b). "Non-Clicks Mean Irrelevant? Propensity Ratio Scoring As a Correction". In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. *WSDM '21*. Virtual Event, Israel: Association for Computing Machinery. 481–489.

Wang, P., X. Mi, X. Liao, X. Wang, K. Yuan, F. Qian, and R. A. Beyah. (2018). "Game of Missuggestions: Semantic Analysis of Search-Autocomplete Manipulations." In: *NDSS*.

Wang, Y.-X., A. Agarwal, and M. Dudik. (2017). "Optimal and Adaptive Off-Policy Evaluation in Contextual Bandits". In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. *ICML'17*. Sydney, NSW, Australia: JMLR.org. 3589–3597.

Wang, X., M. Bendersky, D. Metzler, and M. Najork. (2016). "Learning to Rank with Selection Bias in Personal Search". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 115–124.

Wang, Y., W. Ma, M. Zhang, Y. Liu, and S. Ma. (2023b). "A survey on the fairness of recommender systems". *ACM Transactions on Information Systems*. 41(3): 1–43.

Wang, Y., Z. Tao, and Y. Fang. (2022a). "A Meta-learning Approach to Fair Ranking". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2539–2544.

Wang, Y., Z. Tao, and Y. Fang. (2024a). "A Unified Meta-learning Framework for Fair Ranking with Curriculum Learning". *IEEE Transactions on Knowledge and Data Engineering*.

Wang, Y., X. Wu, H.-T. Wu, Z. Tao, and Y. Fang. (2024b). “Do Large Language Models Rank Fairly? An Empirical Study on the Fairness of LLMs as Rankers”. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Ed. by K. Duh, H. Gomez, and S. Bethard. Mexico City, Mexico: Association for Computational Linguistics. 5712–5724. DOI: 10.18653/v1/2024.naacl-long.319.

Wang, Y., P. Yin, Z. Tao, H. Venkatesan, J. Lai, Y. Fang, and P. Xiao. (2023c). “An Empirical Study of Selection Bias in Pinterest Ads Retrieval”. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5174–5183.

Wang, Y., L. Lyu, and A. Anand. (2022b). “BERT Rankers Are Brittle: A Study Using Adversarial Document Perturbations”. In: *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval*. 115–120.

Wang, Z., Z. Wu, J. Zhang, N. Jain, X. Guan, and A. Koshiyama. (2024c). “Bias Amplification: Language Models as Increasingly Biased Media”. *arXiv preprint arXiv:2410.15234*.

Weidinger, L., J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh, *et al.* (2021). “Ethical and social risks of harm from language models”. *arXiv preprint arXiv:2112.04359*.

Wilkie, C. and L. Azzopardi. (2013). “Relating retrievability, performance and length”. In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 937–940.

Wilkie, C. and L. Azzopardi. (2014a). “A retrievability analysis: Exploring the relationship between retrieval bias and retrieval performance”. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. 81–90.

Wilkie, C. and L. Azzopardi. (2014b). “Best and fairest: An empirical analysis of retrieval system bias”. In: *European Conference on Information Retrieval*. Springer. 13–25.

Wu, H., C. Ma, B. Mitra, F. Diaz, and X. Liu. (2022a). "A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation". *ACM Transactions on Information Systems*. 41(2): 1–29.

Wu, H., B. Mitra, and N. Craswell. (2024a). "Towards Group-aware Search Success". In: *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval*. 123–131.

Wu, H., B. Mitra, C. Ma, F. Diaz, and X. Liu. (2022b). "Joint multisided exposure fairness for recommendation". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 703–714.

Wu, X., S. Li, H.-T. Wu, Z. Tao, and Y. Fang. (2024b). "Does RAG Introduce Unfairness in LLMs? Evaluating Fairness in Retrieval-Augmented Generation Systems". *arXiv preprint arXiv:2409.19804*.

Wu, X., Y. Wang, H. Wu, Z. Tao, and Y. Fang. (2024c). "Evaluating Fairness in Large Vision-Language Models Across Diverse Demographic Attributes and Prompts". *CoRR*. abs/2406.17974. DOI: 10.48550/ARXIV.2406.17974. arXiv: 2406.17974.

Xiao, L., Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, and M. Shaoping. (2017). "Fairness-aware group recommendation with pareto-efficiency". In: *RecSys*.

Xiong, L., C. Xiong, Y. Li, K.-F. Tang, J. Liu, P. N. Bennett, J. Ahmed, and A. Overwijk. (2021). "Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=zeFrfgyZln.

Xu, S., D. Hou, L. Pang, J. Deng, J. Xu, H. Shen, and X. Cheng. (2023). "AI-Generated Images Introduce Invisible Relevance Bias to Text-Image Retrieval". *arXiv preprint arXiv:2311.14084*.

Yadav, H., Z. Du, and T. Joachims. (2019). "Fair Learning-to-Rank from Implicit Feedback". *CoRR*. abs/1911.08054.

Yadav, H., Z. Du, and T. Joachims. (2021). "Policy-Gradient Training of Fair and Unbiased Ranking Functions". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. *SIGIR '21*. New York, NY, USA: Association for Computing Machinery. 1044–1053.

Yan, L., Z. Qin, H. Zhuang, X. Wang, M. Bendersky, and M. Najork. (2022). “Revisiting Two-Tower Models for Unbiased Learning to Rank”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’22*. New York, NY, USA: Association for Computing Machinery. 2410–2414.

Yanagi, R., R. Togo, T. Ogawa, and M. Haseyama. (2021). “Database-Adaptive Re-Ranking for Enhancing Cross-Modal Image Retrieval”. In: *Proceedings of the 29th ACM International Conference on Multimedia. MM ’21*. Virtual Event, China: Association for Computing Machinery. 3816–3825.

Yang, J.-H., C. Lassance, R. Sampaio De Rezende, K. Srinivasan, M. Redi, S. Clinchant, and J. Lin. (2023a). “AToMiC: An Image/Text Retrieval Test Collection to Support Multimedia Content Creation”. In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’23*. New York, NY, USA: Association for Computing Machinery. 2975–2984.

Yang, K., J. R. Loftus, and J. Stoyanovich. (2020). “Causal intersectionality for fair ranking”. *arXiv preprint arXiv:2006.08688*.

Yang, K. and J. Stoyanovich. (2017). “Measuring fairness in ranked outputs”. *SSDBM*.

Yang, L., Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang. (2023b). “Diffusion models: A comprehensive survey of methods and applications”. *ACM Computing Surveys*. 56(4): 1–39.

Yang, T. and Q. Ai. (2021). “Maximizing marginal fairness for dynamic learning to rank”. In: *Proceedings of the Web Conference 2021*. 137–145.

Yang, Z., Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le. (2019). “XLNet: Generalized Autoregressive Pretraining for Language Understanding”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*.

Yao, L., W. Chen, and Q. Jin. (2023). “CapEnrich: Enriching Caption Semantics for Web Images via Cross-Modal Pre-Trained Knowledge”. In: *Proceedings of the ACM Web Conference 2023. WWW '23*. Austin, TX, USA: Association for Computing Machinery. 2392–2401.

Yao, S. and B. Huang. (2017). “Beyond Parity: Fairness Objectives for Collaborative Filtering”. In: *NeurIPS*.

Yates, A., R. Nogueira, and J. Lin. (2021). “Pretrained Transformers for Text Ranking: BERT and Beyond”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorials*. Ed. by G. Kondrak, K. Bontcheva, and D. Gillick. Online: Association for Computational Linguistics. 1–4. DOI: 10.18653/v1/2021.naacl-tutorials.1.

Yilmaz, E. and J. A. Aslam. (2006). “Estimating Average Precision with Incomplete and Imperfect Judgments”. In: *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*. 102–111.

Yu, L., J. Chen, A. Sinha, M. Wang, Y. Chen, T. L. Berg, and N. Zhang. (2022). “CommerceMM: Large-Scale Commerce MultiModal Representation Learning with Omni Retrieval”. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4433–4442.

Yue, Y., R. Patel, and H. Roehrig. (2010). “Beyond Position Bias: Examining Result Attractiveness as a Source of Presentation Bias in Clickthrough Data”. In: *Proceedings of the 19th International Conference on World Wide Web. WWW '10*. Raleigh, North Carolina, USA: Association for Computing Machinery. 1011–1018.

Zehlike, M., F. Bonchi, C. Castillo, S. Hajian, M. Megahed, and R. Baeza-Yates. (2017). “FA* IR: A Fair Top-k Ranking Algorithm”. *CIKM*.

Zehlike, M. and C. Castillo. (2020). “Reducing disparate exposure in ranking: A learning to rank approach”. In: *Proceedings of the web conference 2020*. 2849–2855.

Zehlike, M., T. Sühr, C. Castillo, and I. Kitanovski. (2020). "FairSearch: A Tool For Fairness in Ranked Search Results". In: *Companion Proceedings of the Web Conference 2020. WWW '20*. Taipei, Taiwan: Association for Computing Machinery. 172–175.

Zehlike, M., K. Yang, and J. Stoyanovich. (2022). "Fairness in ranking: A survey". *ACM Computing Surveys*. 6: 1–36.

Zellers, R., X. Lu, J. Hessel, Y. Yu, J. S. Park, J. Cao, A. Farhadi, and Y. Choi. (2021). "MERLOT: Multimodal Neural Script Knowledge Models". In: *Advances in Neural Information Processing Systems 34*.

Zemel, R., Y. Wu, K. Swersky, T. Pitassi, and C. Dwork. (2013). "Learning fair representations". In: *International conference on machine learning*. PMLR. 325–333.

Zerveas, G., N. Rekabsaz, D. Cohen, and C. Eickhoff. (2022a). "CODER: An efficient framework for improving retrieval through COntextual Document Embedding Reranking". In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Ed. by Y. Goldberg, Z. Kozareva, and Y. Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. 10626–10644. DOI: 10.18653/v1/2022.emnlp-main.727.

Zerveas, G., N. Rekabsaz, D. Cohen, and C. Eickhoff. (2022b). "Mitigating Bias in Search Results Through Contextual Document Reranking and Neutrality Regularization". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '22*. New York, NY, USA: Association for Computing Machinery. 2532–2538.

Zhan, J., J. Mao, Y. Liu, M. Zhang, and S. Ma. (2020). "An Analysis of BERT in Document Ranking". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '20*. Virtual Event, China: Association for Computing Machinery. 1941–1944.

Zhang, H., R. Zhang, J. Guo, M. de Rijke, Y. Fan, and X. Cheng. (2024a). "Are Large Language Models Good at Utility Judgments?" In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '24*. Washington DC, USA: Association for Computing Machinery. 1941–1951.

Zhang, J., J. Mao, Y. Liu, R. Zhang, M. Zhang, S. Ma, J. Xu, and Q. Tian. (2019). "Context-Aware Ranking by Constructing a Virtual Environment for Reinforcement Learning". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1603–1612.

Zhang, R., J. Han, C. Liu, A. Zhou, P. Lu, Y. Qiao, H. Li, and P. Gao. (2024b). "LLaMA-Adapter: Efficient Fine-tuning of Large Language Models with Zero-initialized Attention". In: *The Twelfth International Conference on Learning Representations*. URL: https://openreview.net/forum?id=d4UiXAHN2W.

Zhang, Y., L. Yan, Z. Qin, H. Zhuang, J. Shen, X. Wang, M. Bendersky, and M. Najork. (2023). "Towards Disentangling Relevance and Bias in Unbiased Learning to Rank". In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '23*. New York, NY, USA: Association for Computing Machinery. 5618–5627.

Zhao, J., T. Wang, M. Yatskar, R. Cotterell, V. Ordonez, and K.-W. Chang. (2019a). "Gender Bias in Contextualized Word Embeddings". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Ed. by J. Burstein, C. Doran, and T. Solorio. Minneapolis, Minnesota: Association for Computational Linguistics. 629–634. DOI: 10.18653/v1/N19-1064.

Zhao, Z., L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, and E. Chi. (2019b). "Recommending What Video to Watch next: A Multitask Ranking System". In: *Proceedings of the 13th ACM Conference on Recommender Systems. RecSys '19*. Copenhagen, Denmark: Association for Computing Machinery. 43–51.

Zhou, Y., Z. Dou, Y. Zhu, and J.-R. Wen. (2021). "PSSL: Self-Supervised Learning for Personalized Search with Contrastive Sampling". In: *Proceedings of the 30th ACM International Conference on Information and Knowledge Management. CIKM '21.* Virtual Event, Queensland, Australia. 2749–2758.

Zhu, Y., H. Yuan, S. Wang, J. Liu, W. Liu, C. Deng, Z. Dou, and J.-R. Wen. (2023). "Large language models for information retrieval: A survey". *arXiv preprint arXiv:2308.07107.*

Zhuang, H., Z. Qin, K. Hui, J. Wu, L. Yan, X. Wang, and M. Bendersky. (2024). "Beyond Yes and No: Improving Zero-Shot LLM Rankers via Scoring Fine-Grained Relevance Labels". In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers).* Ed. by K. Duh, H. Gomez, and S. Bethard. Mexico City, Mexico: Association for Computational Linguistics. 358–370. DOI: 10.18653/v1/2024.naacl-short.31.

Zhuang, H., Z. Qin, R. Jagerman, K. Hui, J. Ma, J. Lu, J. Ni, X. Wang, and M. Bendersky. (2023). "RankT5: Fine-Tuning T5 for Text Ranking with Ranking Losses". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '23.* New York, NY, USA: Association for Computing Machinery. 2308–2313.

Zhuang, H., Z. Qin, X. Wang, M. Bendersky, X. Qian, P. Hu, and D. C. Chen. (2021). "Cross-Positional Attention for Debiasing Clicks". In: *Proceedings of the Web Conference 2021. WWW '21.* Ljubljana, Slovenia: Association for Computing Machinery. 788–797.

Zou, L., H. M. andXiaokai Chu, J. Tang, W. Ye, S. Wang, and D. Yin. (2022a). "A Large Scale Search Dataset for Unbiased Learning to Rank". In: *NeurIPS 2022.*

Zou, L., C. Hao, H. Cai, S. Wang, S. Cheng, Z. Cheng, W. Ye, S. Gu, and D. Yin. (2022b). "Approximated Doubly Robust Search Relevance Estimation". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management. CIKM '22.* Atlanta, GA, USA: Association for Computing Machinery. 3756–3765.

Zou, L., S. Zhang, H. Cai, D. Ma, S. Cheng, S. Wang, D. Shi, Z. Cheng, and D. Yin. (2021). "Pre-Trained Language Model Based Ranking in Baidu Search". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4014–4022.