Fairness in Information Access Systems
Other titles in Foundations and Trends® in Information Retrieval

Deep Learning for Dialogue Systems: Chit-Chat and Beyond
Rui Yan, Juntao Li and Zhou Yu
ISBN: 978-1-63828-022-4

Search Interface Design and Evaluation
Chang Liu, Ying-Hsang Liu, Jingjing Liu and Ralf Bierig

Psychology-informed Recommender Systems
Elisabeth Lex, Dominik Kowald, Paul Seitlinger, Thi Ngoc Trang Tran, Alexander Felfernig and Markus Schedl

Search and Discovery in Personal Email Collections
Michael Bendersky, Xuanhui Wang, Marc Najork and Donald Metzler
ISBN: 978-1-68083-838-1

Extracting, Mining and Predicting Users’ Interests from Social Media
Fattane Zarrinkalam, Stefano Faralli, Guangyuan Piao and Ebrahim Bagheri
ISBN: 978-1-68083-738-4
Fairness in Information Access Systems

Michael D. Ekstrand
People and Information Research Team (PIReT)
Boise State University
ekstrand@acm.org

Anubrata Das
School of Information
University of Texas at Austin
anubrata@utexas.edu

Robin Burke
Department of Information Science
University of Colorado
robin.burke@colorado.edu

Fernando Diaz
Mila - Quebec AI Institute
diazf@acm.org

Full text available at: http://dx.doi.org/10.1561/1500000079
Foundations and Trends® in Information Retrieval
Volume 16, Issue 1-2, 2022

Editorial Board

Editors-in-Chief
Yiqun Liu  Diane Kelly
Tsinghua University  University of Tennessee
China  USA

Editors
Barbara Poblete  Lynda Tamine
University of Chile  University of Toulouse
Claudia Hauff  Maarten de Rijke
Delft University of Technology  University of Amsterdam and Ahold Delhaize
Ellen M. Voorhees  Rodrygo Luis Teodoro Santos
National Institute of Standards and Technology  Universidade Federal de Minas Gerais
Hang Li  Ruihua Song
Bytedance Technology  Renmin University of China
Isabelle Moulinier  Shane Culpepper
Capital One  RMIT
Jaap Kamps  Soumen Chakrabarti
University of Amsterdam  Indian Institute of Technology
Jimmy Lin  Xiangnan He
University of Waterloo  University of Science and Technology of China
Leif Azzopardi  Xuanjing Huang
University of Glasgow  Fudan University
Lorraine Goeuriot  Yubin Kim
Université Grenoble Alpes  Etsy
Zi Helen Huang  University of Queensland

Full text available at: http://dx.doi.org/10.1561/1500000079
Editorial Scope

Topics

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, 2022, Volume 16, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.
# Contents

1 Introduction  
1.1 Abstracting Information Access .......................... 6  
1.2 A Brief History of Fairness .............................. 7  
1.3 Fairness and Bias ........................................ 9  
1.4 Fairness and Other Responsibility Concerns .............. 10  
1.5 Running Examples ....................................... 11  
1.6 How to Use This Monograph ............................... 14  
1.7 Our Perspective ........................................ 15  
1.8 Some Cautions ........................................... 16  

2 Information Access Fundamentals  
2.1 System Overview ......................................... 22  
2.2 Repository of Items ...................................... 23  
2.3 Users and Information Needs ............................. 25  
2.4 Presentation .............................................. 26  
2.5 Evaluation ................................................ 28  
2.6 Algorithmic Foundations ................................. 32  

3 Fairness Fundamentals  
3.1 Sources of Unfairness .................................. 41  
3.2 Problems and Concepts .................................. 47  
3.3 Mitigation Methods ..................................... 56  
3.4 Wrapping Up .............................................. 59
4 The Problem Space

4.1 What Breaks in Information Access? ........................................ 64
4.2 Kinds of Harms in Information Access ................................. 68
4.3 Fair for Who? .................................................................. 70
4.4 Fair How? ...................................................................... 78
4.5 Fair from What Vantage Point? ........................................... 84
4.6 Fair on What Time Scale? .................................................. 86
4.7 Fairness and the System Pipeline ........................................... 87
4.8 Fairness and Other Concerns ................................................ 89
4.9 Contributing Back to ML Fairness ........................................... 91
4.10 Navigating the Problem Space .............................................. 92

5 Consumer Fairness

5.1 Individual Fairness ............................................................ 94
5.2 Group Fairness through Disaggregated Utility ..................... 95
5.3 Disparate Effectiveness ...................................................... 97
5.4 Providing Fair Utility ........................................................ 99
5.5 Fairness Beyond Accuracy .................................................. 100
5.6 More Complex Scenarios ................................................... 102

6 Provider Fairness

6.1 Provider Representation ..................................................... 105
6.2 Provider Exposure and Utility ............................................. 112
6.3 Fair Accuracy and Pairwise Fairness ................................... 119
6.4 Related Problem: Subject Fairness ....................................... 121

7 Dynamic Fairness

7.1 Feedback Loops ............................................................... 124
7.2 Dynamic Evaluation ........................................................ 126
7.3 Opportunities in Feedback Loops ........................................ 126

8 Next Steps for Fair Information Access

8.1 Directions for Future Research ........................................... 128
8.2 Recommendations for Studying Fairness ............................ 131
8.3 Concluding Remarks ....................................................... 133

Acknowledgements ................................................................. 134
Fairness in Information Access Systems

Michael D. Ekstrand¹, Anubrata Das², Robin Burke³ and Fernando Diaz⁴

¹People and Information Research Team (PIReT), Boise State University, USA; ekstrand@acm.org
²School of Information, University of Texas at Austin, USA; anubrata@utexas.edu
³That Recommender Systems Lab, Department of Information Science, University of Colorado, USA; robin.burke@colorado.edu
⁴Mila - Quebec AI Institute, Canada; diazf@acm.org

ABSTRACT

Recommendation, information retrieval, and other information access systems pose unique challenges for investigating and applying the fairness and non-discrimination concepts that have been developed for studying other machine learning systems. While fair information access shares many commonalities with fair classification, there are important differences: the multistakeholder nature of information access applications, the rank-based problem setting, the centrality of personalization in many cases, and the role of user response all complicate the problem of identifying precisely what types and operationalizations of fairness may be relevant.

In this monograph, we present a taxonomy of the various dimensions of fair information access and survey the literature to date on this new and rapidly-growing topic. We
preface this with brief introductions to information access and algorithmic fairness to facilitate the use of this work by scholars with experience in one (or neither) of these fields who wish to study their intersection. We conclude with several open problems in fair information access, along with some suggestions for how to approach research in this space.
# List of Key Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Defined in</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias</td>
<td>1.3</td>
<td>9</td>
</tr>
<tr>
<td>disparate impact</td>
<td>3.2.3</td>
<td>51</td>
</tr>
<tr>
<td>disparate mistreatment</td>
<td>3.2.3</td>
<td>52</td>
</tr>
<tr>
<td>disparate treatment</td>
<td>3.2.3</td>
<td>50</td>
</tr>
<tr>
<td>fairness</td>
<td>1.3</td>
<td>9</td>
</tr>
<tr>
<td>group fairness</td>
<td>3.2.3</td>
<td>50</td>
</tr>
<tr>
<td>individual fairness</td>
<td>3.2.2</td>
<td>48</td>
</tr>
<tr>
<td>information access</td>
<td>1.1</td>
<td>7</td>
</tr>
<tr>
<td>information need</td>
<td>2.3</td>
<td>25</td>
</tr>
</tbody>
</table>

The index provides a more comprehensive cross-reference of terms used in this monograph.
As long as humans have recorded information in durable form, they have needed tools to access it: to locate the information they seek, review it, and consume it. Digitally, tools to facilitate information access take a variety of forms, including information retrieval and recommendation systems; these tools have been powered by technologies built on various paradigms, from heuristic metrics and expert systems to deep neural networks with sophisticated rank-based objective functions. Fundamentally, these technologies take a user’s information need (an explicit and/or implicit need for information for some purpose (Kuhlthau, 1993), such as filling in knowledge or selecting a product) and locate documents or items that are relevant (that is, will meet the user’s need).

Throughout the history of these technologies — which we treat under the integrated banner of information access systems — both research and development have been concerned with a range of effects beyond a system’s ability to locate individual items that are relevant to a user’s information need. Research has examined the diversity and novelty of results (Santos et al., 2015; Hurley and Zhang, 2011) and the coverage of the system, among other concerns. In recent years, this concern has extended to the fairness of an information access system: are
the benefits and resources it provides fairly allocated between different people or organizations it affects? Does it introduce or reproduce harms, particularly harms distributed in an unfair or unjust way? This challenge is connected to the broader set of research on fairness in sociotechnical systems generally and AI systems more particularly (Mitchell et al., 2020; Barocas et al., 2019), but information access systems have their own set of particular challenges and possibilities.

Fairness is not an entirely new concern for information access; various fairness problems can be connected to topics with long precedent in the information retrieval and recommender systems literature. In the context of information retrieval, Friedman and Nissenbaum (1996) and Introna and Nissenbaum (2000) recognized the potential for search engines to embed social, political, and moral values in their ranking functions. In order to assess the impact of such values, Mowshowitz and Kawaguchi (2002) developed a metric to measure a search engine’s deviation from an ideal exposure of content. Although conversations often focus on bias in algorithmic ranking, Vaughan and Zhang (2007) and Vaughan and Thelwall (2004) note that bias can be introduced because of biased crawling and indexing; in particular, they describe, writing in the 2000s, how Chinese webpages were under-indexed by search engines. These observations led to discussion amongst legal scholars about the regulation of search engines (Goldman, 2005; Pasquale, 2006). Azzopardi and Vinay (2008) proposed the notion of document retrievability and investigated the skew in this distribution for different retrieval systems. Work on popularity bias (Celma and Cano, 2008; Zhao et al., 2013; Cañamares and Castells, 2018) and rich-get-richer effects (Cho et al., 2005), along with attempts to ensure quality and equity in long-tail recommendations (Ferraro, 2019), can be viewed as a type of fairness problem: the system should not inordinately favor popular, well-known, and possibly well-funded content creators. In a group recommendation, one common objective is to ensure that the various members of a group are treated fairly (Kaya et al., 2020).

The work on fair information access that we present here goes beyond these problems to examine how various forms of unfairness — particularly those that arise from social biases (Olteanu et al., 2019) — can make their way in to the data, algorithms, and outputs of informa-
Introduction

In this monograph, we provide an introduction to fairness in information access, aiming to give students, researchers, and practitioners a starting point for understanding the problem space, the research to date, and a foundation for their further study. Fairness in information access draws heavily from the fair machine learning literature, which we summarize in Section 3; researchers and practitioners looking to study or improve the fairness of information access will do well to pay attention to a broad set of research results. For reasons of scope, we are primarily concerned here with the fairness of the information access transaction itself: providing results in response to a request encoding an information need. Fairness concerns can also arise in other aspects of the system, such as the representation and presentation of documents themselves, or in support facilities such as query suggestions (Noble, 2018). We provide brief pointers on these topics, but a detailed treatment is left for future synthesis, noting that they have not yet received as much attention in the research literature. We are also specifically concerned with fairness-related harms, and not the broader set of harms that may arise in information access such as the amplification of disinformation.

Throughout this work, we use the term system to describe an algorithmic system that performs some task: retrieving information, recommending items, classifying or scoring people based on their data. These systems are embedded in social contexts, operating on human-provided inputs and producing results acted upon by humans. The technical system forms one part of a broader socio-technical system.

1.1 Abstracting Information Access

Our choice to title this monograph “Fairness in Information Access” is quite deliberate. While there is significant technical and social overlap between information retrieval, recommender systems, and related fields, they are distinct communities with differences in terminology, problem
1.2 A Brief History of Fairness

definitions, and evaluation practices. However, there are fundamental
commonalities, and they present many of the same problems that
complicate notions of fairness, including ranked outputs, personalized
relevance, repeated decision-making, and multistakeholder structure.
We therefore refer to them together as information access systems —
algorithmic systems that mediate the interaction between a repository
of documents or items and a user’s information need.

This information access umbrella includes information retrieval,
recommender systems, information filtering, and some applications of
natural language processing. In Section 2, we present a fuller treatment
of this integration and reviews the fundamentals of information access,
both to introduce the concepts to readers who come to this paper from
a general fairness background and to lay out consistent terminology
for our readers from information retrieval or recommender systems
backgrounds.

1.2 A Brief History of Fairness

In the pursuit of fairness in algorithmic systems and the society more
generally, the authority of Aristotle’s citation of Plato “treat like cases
alike” is a key touchstone: a normative requirement that those who
are equal before the law should receive equal treatment (Gosepath,
2011). In more recent scholarship, the study of distributive welfare
extends these concepts considerably, recognizing four distinct concepts of
fairness: “exogenous rights, compensation, reward, and fitness.” (Moulin,
2004). Exogenous rights, as the term suggests, relate to external claims
that a system must satisfy: equal shares in property as defined by
contract, for example, or equality of political rights in democratic
societies. Compensation recognizes that fairness may require extra
consideration for parties where costs are unequal — affirmative action
in hiring and college admissions are well-known examples. Reward
justifies inequality on the basis of differing contributions: for example,
increased bonuses to employees with greater contribution to the bottom
line. Finally, we have fitness, the most nebulous category, and the one
that many information access systems inhabit. The fitness principle
holds that goods be distributed to those most fit to use, appreciate, or
derive benefit from them. It is an efficiency principle, where the fairest use is the one that allocates goods where the distribution achieves the maximum utility. Fitness has a natural application to information access, as we seek to locate documents and make them visible based on their utility to the user’s information need.

U.S. legal theory has developed a rich tradition of anti-discrimination law, aimed at ensuring that people are not denied certain benefits (housing, work, education, financial services, etc.) on the basis of protected characteristics (race, color, religion, gender, disability, age, and in many jurisdictions, sexual orientation). It has given rise to several important concepts, such as the disparate impact standard (the idea that an allegedly discriminatory practice can be legally challenged on the grounds that it has disproportionate adverse impact on a protected group, without needing to show intent to discriminate\(^1\)). Crenshaw (1989) points out some of the limitations of this legal framework; in particular, it has often focused on discrimination on the basis of individual protected characteristics, and people who have suffered harm as a result of combinations of protected characteristics (e.g. Black women being denied promotions given to both Black men and White women) have difficulty proving their case and obtaining relief. This theory of particular harms deriving from combinations of characteristics is called intersectionality.

Questions of fairness and discrimination have been the subject of significant discussion in many other communities as well. Educational testing, for example, has several decades of research on the fairness of various testing and assessment instruments; this history is summarized for computer scientists by Hutchinson and Mitchell (2019). Friedman and Nissenbaum (1996) provide one of the earlier examples of addressing questions of bias in computer science, pointing out how even seemingly-innocuous technical decisions may result in biased effects when a computing system is used in its social context. The last ten years have seen significant new activity on fairness in machine learning that\(^1\)

\(^1\)Disparate impact is not sufficient basis to win a discrimination lawsuit; rather, it is the first step in a multi-stage burden-shifting framework used to decide discrimination cases under the standard. Barocas and Selbst (2016) provide an overview of the process.
forms the primary stream of algorithmic fairness research; in Section 3 we provide an introduction to this literature.

1.3 Fairness and Bias

There are many overlapping terms used to discuss issues of fairness, bias, and discrimination. While we give a fuller treatment of the vocabulary in Section 3, we will here introduce how we use these terms in this monograph. Work we cite may use them differently.

When we refer to **fairness**, we are talking about the ways a system treats people, or groups of people, in a way that is considered “unfair” by some moral, legal, or ethical standard. This is typically through effects or impacts that are not experienced in an equitable way, but can sometimes arise through the system’s internal operation or representations. This definition is similar to how Friedman and Nissenbaum (1996) use the term “bias”. There is not one particular definition of what constitutes fairness, as Selbst et al. (2019) and many others have noted; for the purpose of terminology, the important point is that we use the term to refer to normative ideas of what it means to treat people “fairly”, no matter their source.

When we talk about **bias**, we are using the term in something closer to its statistical sense: we mean properties of estimators, models, measurements, and data that systematically deviate from their intended ideal target. As detailed in Section 3.1, we share an expansive view of bias with Mitchell et al. (2020, Section 2.2.1), noting that these biases can be the kinds of statistical biases familiar to science (systematic discrepancies between data or outputs and the underlying observable world), but they can also be societal biases in the form of systematic discrepancies between the observable world and the arguable ideal world that would arise if society eliminated all forms of illegitimate discrimination.

The key distinction in our work is that we use the term “bias” to refer to a fact of the system without making any inherently normative judgment, and “fairness” to discuss the normative aspects of the system and its effects. Some biases are themselves fairness problems; some biases cause fairness problems; some have no effect with regards to
the concerns of fairness; and some may be intentionally introduced to address a fairness problem, often by correcting for another bias. Most fairness problems arise from biases somewhere in the system, its data, or its evaluation, but we find it useful to distinguish between the technical fact and the moral, ethical, or legal concern.

1.4 Fairness and Other Responsibility Concerns

Fairness is commonly grouped together with other concerns under the banner of responsibility in computing systems. These concerns include:

**Accountability** Research on accountability examines the legal, social, and technical mechanisms by which computing systems and their operators, developers, and providers may be held accountable, usually for the human effects of their systems. This can connect directly to fairness when considering how to hold organizations accountable for ensuring their systems uphold societal goals to be fair. Such accountability can be through formal structures, such as applying anti-discrimination law to computing systems, or through informal structures such as applying pressure through publicizing the results of third-party audits.

**Transparency** Transparency (and its close cousin explainability) seeks to make the operation and results of algorithmic systems scrutable to users, developers, auditors, and other stakeholders so that it can be understood, reviewed, and contested. This relates to long-standing concern in information access on explanation (Tintarev and Masthoff, 2007), as well as ideas such as scrutable user models (Kay et al., 2002).

**Safety** Information access systems can be harmful. They can distribute false information, promote fake or dangerous products, and provide support for illegal or malicious activities. These problems have received attention in the research literature, often under the general heading of adversarial information retrieval. See related workshops AIRWeb (Fetterly and Gyöngyi, 2009) and WebQuality (Nielek et al., 2016).
Privacy  Aspects of users’ profiles including queries, interaction history, and usage patterns may be highly revealing of sensitive personal information: consider queries about medical symptoms or clicks on web pages for addiction counseling. It follows that information access systems have a duty to protect such information from harmful disclosure. Research on privacy-preserving recommendation seeks technical solutions to this challenge. Friedman et al. (2015) provide a survey of this area.

Ethics  Computing ethics is concerned broadly with ensuring that the practice and products of computing adhere to appropriate ethical principles. The ACM Code of Ethics (ACM Council, 2018) specifically calls out non-discrimination, along with attention to potential harms, as an ethical obligation for computing professionals.

The report on the FACTS-IR Workshop on Fairness, Accountability, Confidentiality, Transparency, and Safety in Information Retrieval (Roegiest et al., 2019) discusses how many of these concepts play out in information retrieval. In this work we are concerned with fairness, but bring in other concerns as well when they relate to fairness.

1.5  Running Examples

Throughout this monograph, we will use several examples to motivate and explain the various concepts we discuss.

Job and Candidate Search  Many online platforms attempt to connect job-seekers and employment opportunities in some way. Some of these are dedicated employment-seeking platforms, while others, such as LinkedIn and Xing, are more general-purpose professional networking platforms for which job-seeking is one important component.

Job-seeking is a multisided problem — people need good employment and employers need good candidates — and also has significant fairness requirements that are often subject to regulation in various jurisdictions. Some of the specific fairness concerns for this application include:

• Do users receive a fair set of job opportunities in the recommendations or ads in their feed?
Introduction

- If the system assesses a match or fit score for a candidate and a job, is this score fair, or does it under- or over-estimate scores for particular candidates or groups of candidates?

- Do users have a fair opportunity to appear in search lists when recruiters are looking for candidates for a job opening (Geyik and Kenthapadi, 2018)?

- Do employers in protected groups (minority-owned businesses, for example) have their jobs fairly promoted to qualified candidates?

- What fairness concerns come from regulatory requirements?

Music Discovery  The search and recommendation systems in music platforms, such as Spotify, Pandora, and BandCamp, connect listeners with artists. These discovery tools have a significant impact not only on a user’s listening experience and musical enjoyment, but also on artists’ financial and career prospects, due both to direct revenue from listening and the commercial and reputational effects of visibility. Some specific fairness concerns include:

- Do artists receive fair exposure in the system’s search results, recommendation lists, or streamed programming?

- Does the system systematically over- or under-promote particular groups of artists or songwriters through recommendations, search results, and other discovery surfaces (Epps-Darling et al., 2020)?

- Do users receive fair quality of service, or does the system systematically do a better job of modeling some users’ tastes and preferences than others?

- Do recommendations reflect well a user’s preferences and if not, are there systematic errors due to stereotypes of gender, ethnicity, location, or other attributes?

News  News search and recommendation influences user exposure to news articles on social media, news aggregation applications, and
1.5. Running Examples

Search engines. Such influence extends to social and political choices users might make (Kulshrestha et al., 2017; Epstein and Robertson, 2015). Additionally, the filter bubble effect (Pariser, 2011; Alstyne and Brynjolfsson, 2005) may cause users to be exposed primarily to news items that reinforce their beliefs and increase polarization. Depending on the journalistic policy of the provider, news platforms may want to facilitate balanced exposure to news from across the social, political, and cultural spectrum, but this may need to be balanced with the need to de-rank malicious and low-credibility sources.

Specific fairness concerns in news discovery include:

- Does the system provide fair exposure to news on different topics or affected groups?
- Do journalists from different perspectives receive fair visibility or exposure for their content?
- Does the system reward original investigators or primarily direct readers to tertiary sources?
- Do users receive a balanced set of news content?
- Are users in different demographics or locations equally well-served by their news recommendations?

Philanthropic Giving  

Online platforms are increasingly a site for philanthropic giving (Goecks et al., 2008), and therefore recommendation is expected to be an increasing driver of donations. Sites may take an explicitly “peer-to-peer” approach to such giving, as in the educational charity site DonorsChoose.org; this results in many possible donation opportunities for donors to select from, requiring recommendation or sophisticated search to help match donors and opportunities. As many philanthropic organizations have a social justice focus, fairness concerns are essential in developing and evaluating their information access solutions, in particular to avoid potential positive feedback loops in which a subset of causes comes to dominate results and rankings.

In philanthropic settings, we would expect fairness issues to include:
Introduction

- Does the system provide fair opportunities for the various recipients / causes to have their needs supported?
- Are specific groups of recipients under- or over-represented in the recommendation results?

1.6 How to Use This Monograph

We have written this monograph with two audiences in mind:

- Researchers, engineers, and students in information retrieval, recommender systems, and related fields who are looking to understand the literature on fairness, bias, and discrimination, and how it applies to their work.

- Researchers in algorithmic fairness who are looking to understand information access systems, how existing fairness concepts do or do not apply to this set of applications, and the things that information access brings to the research space that may differ from the application settings in which fairness is usually studied.

Due to our interest in serving both of these audiences, we do not expect our readers to have significant familiarity with either information retrieval or algorithmic fairness, although some background in machine learning will be helpful. We have organized the material as follows:

- **Section 2** rehearses the fundamentals of information access systems. This will be a review for most information retrieval and recommender systems researchers; such readers should read it for the terminology we use to integrate the fields, but may wish to focus their study energy elsewhere.

- **Section 3** provides an overview of research on fairness in machine learning generally, particularly in classification. Algorithmic fairness researchers will likely find this section to be a review.

- **Section 4** lays out the problem space of fair information access, providing a multi-faceted taxonomy of the problems in evaluating and removing discrimination and related harms in such systems.
1.7. Our Perspective

- **Sections 5 and 6** survey key literature to date (as of 2021) on fairness in information access, with pointers to research working on many of the problems identified in Section 4, focused on the two most commonly-studied stakeholders: consumers and providers (with discussion of subjects in Section 6.4).

- **Section 7** discusses the need to go beyond point-in-time views of fairness to understand fairness over time how the temporal dynamics of an information access system affect fairness.

- **Section 8** looks to future work and provides tips for research and engineering on fair information access.

Section 4 is the keystone of this work that ties the rest together; subsequent sections work out details in the form of a literature survey of several of the problems discussed in Section 4, and the preceding sections set up the background needed to understand it. For readers looking to budget their time, we recommend they ensure they have the necessary background from Sections 2 and 3, read Sections 4 and 8, and read the later sections that are relevant to their work.

1.7 Our Perspective

While we have written this monograph to be useful for researchers approaching the topic of fairness from a variety of perspectives, we think it is helpful to explicitly describe our own perspectives and motivations, as well as the position from which we approach this work and some limitations it may bring.

Information access systems need to meet a variety of objectives from multiple stakeholders. They need to deliver relevant results to their users, business value for their operators, and visibility to the creators of the documents they present; they often also need to meet a variety of other goals and constraints, such as diversity across subtopics, regulatory compliance, and reducing avoidable harm to users or society. Fairness, as we conceive of it and present it in this monograph, is not a be-all end goal, but rather another family of objectives to be considered in the design and evaluation of information access systems, and a
collection of techniques for enabling those objectives. It also does not encompass the totality of social or ethical objectives guiding a system’s design. Researchers and developers need to work with experts in ethics, policy, sociology, and other relevant fields to identify relevant harms and appropriate objectives for any particular application; the concepts we discuss will be relevant to some of those harms and objectives.

We also emphasize the importance of starting with a robust problem framing: Section 4 is intended to help readers think about the fairness problem they are trying to solve, and position it in a landscape of information access; we have then organized our survey in Sections 5-7 around aspects of problem definition, instead of underlying techniques. Metrics and mitigations are best developed and assessed in the context of a specific, well-defined problem.

Finally, all four authors work in North America and approach the topic primarily in that legal and moral context. A Western focus, and particularly concepts of bias and discrimination rooted in United States legal theory, currently dominates thinking and research on algorithmic fairness in general. This is a limitation of the field that others have noted and critiqued (Sambasivan et al., 2020); our present work acknowledges but does not correct this imbalance. While we attempt to engage with definitions and fairness objectives beyond the U.S., this work admittedly has a Western and especially U.S. focus in its treatment of the material. We look forward to seeing other scholars survey this topic from other perspectives.

1.8 Some Cautions

We hope that this monograph will help scholars from a variety of backgrounds to understand the emerging literature on fairness in information access and to advance the field in useful directions. In addition to the general concerns of careful, thoughtful science, work on fairness often engages with data and constructs that touch on fundamental aspects of human identity and experience. This work must also be done with great care and compassion to ensure that users, creators, and other stakeholders are treated with respect and dignity and to avoid various traps that result in overbroad or ungeneralizable claims.
1.8. Some Cautions

We argue that there is nothing particularly new about this, but that thinking about the fairness of information access brings to the surface issues that should be considered in all research and development on information systems.

1.8.1 Beware Abstraction Traps

Our first caution is to beware of the allure of abstraction. Selbst et al. (2019) describe several specific problems that arise from excessive or inappropriate abstraction in fairness research in general. Their core argument is that the tendency in computer science to seek general, abstract forms of problems, while useful for developing tools and results that can be applied to a wide range of tasks, can cause important social aspects of technology and its impacts to be obscured.

One reason for this is that social problems that appear to be structurally similar arise from distinct (though possibly intertwined) causes and mechanisms, and may require different solutions. Sexism and anti-Black racism, for example, are both types of discrimination and fall into the “group fairness” category of algorithmic fairness, but they are not the same problem and have not been reinforced by the same sets of legal and social processes. Discrimination also varies by culture and jurisdiction, and oppression of what appears to be same group may arise from different causes and through different mechanisms in the different places in which it appears. Kohler-Hausmann (2019) argues that social constructivist frameworks for understanding group identities and experiences imply that even understanding what constitutes a group, let alone the discrimination it experiences, is inextricably linked with understanding how that group is constructed and treated in a particular society — an understanding that is inherently bound to the society in question, although there may be similarities in group construction in different contexts.

The result is that unfairness needs to be measured and addressed in each specific way in which it may appear. While general solutions for detecting and mitigating fairness-related harms may arise and be very useful, their effectiveness needs to be re-validated in context for the harms they are meant to address, a point reiterated by Dwork and Ilvento (2018).
Hoffmann (2019) similarly provides several warnings against overly simple ideas of the harms that can arise from discrimination and bias. Computational fairness inherits some of these limitations from its reference material, such as limitations of anti-discrimination law; others arise from what Hoffmann, Selbst et al. (2019), and others argue are reductionistic operationalizations of rich concepts. Hoffmann (2019) notes in particular—and we agree—that treating categories of personal identity as objective features in a multi-dimensional space (a natural move for computer scientists) obfuscates the role of technical and social systems in enacting and producing such categories. This move also has the effect of reducing intersectionality concerns to what can be captured by a subspace projection or similar formal operation, whether or not that corresponds to individual’s lived experience.

We believe computing systems in general, and information access systems in particular, have the opportunity to advance the discussion of emancipation and justice, not just bring existing constructs into a new domain. Information professionals have long been concerned about issues of ethics and justice. Just as two examples, we note that Edmund Berkeley, one of the founders of the Association for Computing Machinery, was an outspoken advocate for the ethical responsibilities of computer scientists as far back as the 1960s (Longo, 2015), and the creation of Computer Professionals for Social Responsibility in the mid-1980s (Finn and DuPont, 2020). The call here is to realize that vision fully and for all people affected by information access systems.

1.8.2 Beware Limits

It is crucial to be clear about the limitations of particular fairness studies and methods. Any work will be limited, if for no other reason than the impossibility of completely solving the problem of discrimination. Those limitations should not paralyze the research community or keep researchers from doing the most they can to advance equity and justice with the resources available to them; rather, work in this space needs to be forthright and thorough about the limitations of its approach, data, and findings. Some limitations common to this space include:
1.8. Some Cautions

- Single-dimensional attributes for which fairness is considered, when in reality people experience discrimination and oppression along multiple simultaneous dimensions.

- Binary definitions of attributes, when in reality many social dimensions have more than two categories or exist on a continuum.

- Taking attributes as fixed and exogenous, when social categories are complex and socially constructed (Hanna et al., 2020).

- Incomplete, erroneous, and/or biased data (Olteanu et al., 2019; Ekstrand and Kluver, 2021).

This is not to say that work on single binary attributes is not useful; research must start somewhere. But it should not stop there, and authors need to be clear about the relationship their work in its broader context and provide a careful accounting of its known limitations.

Some methods are so limited that we advise against their use. For example, some work on fair information access has used statistical gender recognition based on names or computer vision techniques for gender recognition based on profile pictures. This source of data is error-prone, subject to systemic biases (Buolamwini and Gebru, 2018), reductionistic (Hamidi et al., 2018), and fundamentally denies subjects control over their identities, so we do not consider it good practice.

1.8.3 Beware Convenience

Researchers working in this problem space also need to be careful to do the best research possible with available resources, and work to expand those resources to increase the quality and social fidelity of their work, and not take the path of least resistance.

One particular application pertains to this monograph itself and to its proper use and citation. It is convenient and common practice to cite survey papers to quickly summarize a topic or substantiate its

---

\(^2\)We do not provide citations to support the claim that this is in use because our purpose in this paragraph is to critique a general trend, not to focus on any specific paper. Elsewhere in this monograph, we cite work making use of these techniques where it makes a relevant contribution.
relevance. While we naturally welcome citations of our work, we would prefer to be cited specifically for our contributions to the organization and synthesis of fair information access research. The purpose of much of this monograph is to point our readers to the work that others have done, and we specifically ask that you cite those papers, instead of — or in addition to — this one when that work is relevant to your writing and research.
Acknowledgements

Michael Ekstrand’s contributions are based upon work supported by the National Science Foundation under Grant No. IIS 17-51278. Anubrata Das is supported in part by the Micron Foundation, Wipro, and Good Systems¹, a University of Texas at Austin Grand Challenge to develop responsible AI technologies. Robin Burke’s work was supported by the National Science Foundation under Grant No. IIS 19-11025. We also thank James Atwood, Asia Biega, Yoni Halpern, Matt Lease, Hansa Srinivasan, and the anonymous reviewers for providing additional feedback and suggestions.

¹https://goodsystems.utexas.edu/
Appendix
A

Resources for Fair Information Access

In this appendix, we collect pointers to several resources for studying and working on fair information access. We have made every effort to ensure these links are current as of the time of publication, but they may degrade more quickly than the references in the rest of the publication.

A.1 Data Sets


- The PIReT Book Data Tools at https://bookdata.piret.info provide tools to integrate book recommendation data sets (including from BookCrossing, Amazon, and GoodReads) with publicly-available book and author metadata to study provider fairness in book recommendation, as used by Ekstrand and Kluver (2021).

- Ghosh et al. (2021) develop a number of data sets for fair ranking, using various methods and studying the errors of demographic inference for data augmentation.
A.2 Software

There are not yet widely-distributed open-source software for fair recommendation and retrieval; the available code is mostly embedded in published experiment scripts, or general-purpose systems repurposed for fair information access.

- Terrier (http://terrierteam.dcs.gla.ac.uk/research.html) provides xQuAD, a diversification technique that has been successfully applied for fair search ranking (Mcdonald and Ounis, 2020).


- librec-auto (https://librec-auto.readthedocs.io/en/latest/) provides automated support for running recommender systems experiments, including fairness metrics.


References


References


References


References


References


References


References


References


accountability, 10, 69
anti-classification, 55
anti-subordination, 55, 112
bias, 9
  defined for this work, 9
  societal, 45
  statistical, 45
browsing model, 30, 113
calibrated fairness, 109, 110
candidate search, 11
Chouldechova-Kleinberg theorem, 54
cold start, 83, 95
collaborative filter, 32
collection, 23
customer, 21, 22, 25, 66, 67, 71–73
customer fairness, 71, 93
corpus, see repository creation, 23
discount function, 31, 114
discovery
  job candidates, 11
  jobs, 11
  music, 12
  news, 12
disparate effectiveness, 97
disparate impact, 8, 51, 117
disparate treatment, 50, 117
diversity, 37, 75, 90, 105, 121
document, see item
dominant group, see group, dominant
embedding
  item, 34
  user, 36
ethics, 11
evaluation, 28, 88, 95
exposure, 103, 112
fairness, 9, 15
defined for this work, 9
feedback loop, 86, 124–125
four-fifths rule, 51

group, 50
dominant, 50
majority, see dominant
protected, 50, 77, 107
unprotected, see dominant
group fairness, 41, 50, 78, 95, 116

harm
distributional, 39, 48, 78, 106
representational, 39, 48, 78, 106

individual fairness, 41, 48, 55, 78, 94, 113
information access system, 7, 21, 23, 61
information need, 21, 22, 25, 65
intersectionality, 8, 18, 102
item, 21, 22, 23, 64, 114

job search, 11, 77

limitations, 18

majority group, see group, dominant
matching, 98
maximum marginal relevance, 37, 122
measurement invariance, 98
meritocratic fairness, 53, 118
metadata, 24

multisided platform, 70
music discovery, 12
news discovery, 12

parity, 51–54
calibration, 53
error, 52, 55
predictive value, 53
recall, 53
statistical, 51, 55, 116
philanthropic giving, 13
policy, 32
privacy, 10
process, 100
protected group, see group, protected
provider, 23, 67, 73–74, 103
provider fairness, 73, 75, 90, 103

query, 25, 65

ranking, 27, 32, 64, 83, 105
re-ranking, 37, 99, 111, 118
repository, 21, 23
representation
item, 24, 87
user, 26, 88
results, 21, 26
rivalrous, see subtractable

safety, 10
satisfaction, 21, 28
scoring function, 32, 82
sensitive attribute, 8, 41
separate, simultaneous, and symmetric, 64
session, 25, 26, 27
social constructionism, 17
societal bias, see bias, societal
stakeholders, 70–78
statistical bias, see bias, statistical
stochastic ranking, 113
subject, 15, 75, 121
subtractability, 83
subtractable, 64, 100
system, 6
time, 86
transparency, 10
turn, 26
usage, 24
user model, 35
utility, 28, 64
we’re all equal (WAE), 49, 51, 54, 55
what you see is what you get (WYSIWYG), 49, 52, 54, 112