"End to End"

Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems

Avital Shulner Tal
The University of Haifa
Haifa, Israel
avitalshulner@gmail.com

Khuyagbaatar Batsuren
The University of Trento
Trento, Italy
k.batsuren@unitn.it

Veronika Bogina
The University of Haifa
Haifa, Israel
sveron@gmail.com

Fausto Giunchiglia
The University of Trento
Trento, Italy
fausto.giunchiglia@unitn.it

Alan Hartman
The University of Haifa
Haifa, Israel
alan.hartman.gm@gmail.com

Styliani Kleanthous Loizou
Open University of Cyprus
Nicosia, Cyprus
styliani.kleanthous@gmail.com

Tsvi Kuflik
The University of Haifa
Haifa, Israel
tsvikak@is.haifa.ac.il

Jahna Otterbacher
Open University of Cyprus
Nicosia, Cyprus
jahna.otterbacher@ouc.ac.cy

***Abstract*— Algorithms play an increasing role in our everyday lives. Recently, the harmful potential of biased algorithms has been recognized by researchers and practitioners. We have also witnessed a growing interest in ensuring the fairness and transparency of algorithmic systems. However, so far there is no agreed upon solution and not even an agreed terminology. The proposed research defines the problem space, solution space and a prototype of comprehensive framework for the detection and reducing biases in algorithmic systems.**

Keywords*— *Algorithmic Systems; Transparency; Bias; Diversity; Fairness

I. Introduction

Personal diversity, biases and discrimination are not a new phenomenon, they have been the subject of numerous studies. With the evolution of information and communication technologies and especially the Internet that has globalized the world, a new set of biases and discrimination has emerged. This phenomenon is referred to as algorithmic bias [5]. When designing a system, we have some implicit assumptions about its users and purposes, regardless of the development process we follow [24]. When we deploy a system online, we cannot anticipate who will use it and how. Thus, the users of our system will be diverse, and most likely different from the audience we initially had in mind [25].

The increasing diversity of users of online systems has raised the issue of algorithmic biases. A given system may show behaviors that deviate from what users expect, or what they consider to be normal or "fair" with respect to their own context. Detection of such deviations in a system's behavior leads to discussions of whether the system is behaving in a manner that is "fair". However, there is no single baseline or standard to which we can compare the behaviors of a given system with a globalized user base. What is "normal" depends on many contextual factors, including one's socio-cultural environment and the prevailing values in a given society [6]. Furthermore, Giunchiglia et al. defined diversity as the co-existence of contradictory statements, some of which may be non-factual or referring to opposing beliefs/opinions [11]. By bringing diversity into the discussion, we aim to create a framework for detecting and reducing algorithmic biases and its application in real world domains.

In our connected world, the data, information and knowledge sources available for exploitation by information systems have become increasingly complex and dynamic [12]. Given these challenges, we juxtapose the notions of diversity, bias, fairness and transparency, to achieve a holistic understanding of the algorithmic problems and create an "End to End" framework for "fair", bias-minimized and transparent algorithmic systems.

In this research, we characterize the problem (sources of bias) and solution space (tools for avoiding/detecting and addressing the problem). We also propose general prototype for future work in the area of Algorithmic Transparency (AT). Transparency is fundamental to algorithmic systems [26]. However, we argue diversity in knowledge and context cannot be neglected when addressing problems of algorithmic bias.

The rest of this paper is organized as follows. The next section presents the motivation for this research and a brief compilation of the relevant background (e.g. definitions) and related work in different domains. Subsequently, we present our main contributions: the problem and solution spaces of algorithmic systems and a framework prototype for reducing biases and promoting transparency.

II. Background and Related Work

Although algorithms simply present results of calculations, the training data and definitions that they use may be provided by humans, machines, or both, and can mistakenly pick up human biases during the process (e.g. when the algorithm is programmed or in the cases of biased datasets), or when humans interact with the algorithm. Moreover, developers need to ensure proper context and application of the results while the users influence how the results are presented to them [16].

With the increasing prevalence of algorithms, many studies and projects dealing with AT have emerged in different domains, such as search engine biases, recommender systems, context aware systems, decision-making algorithms, text

978-1-7281-3634-9/19/$31.00 ©2019 IEEE

classification and others. Many conferences focusing on information and communication technology and its application include related topics/tracks. Some examples include UMAP 2019 that has Privacy and Transparency track as well as a workshop on algorithmic transparency [1], IUI 2019 features workshops on the explainability of smart systems [2] and intelligent user interfaces for algorithmic transparency in emerging technologies[3], as well as RecSys 2019[4], IJCAI 2019[5], WWW 2019[6], CSCW 2019[7] who showed interest in algorithmic systems transparency explainability and fairness in different domains. Due to the constraints of space, we present only a representative sample of studies which relate to these issues.

One example is the OPAL (Open Algorithms) project, which aims to unlock the potential of private data for public good in a privacy-conscientious, scalable, socially and economically sustainable manner in order to enhance fairness, accountability, and transparency in algorithmic decision-making [17]. Lepri et al. also highlighted the criticality and urgency of multi-disciplinary research for co-developing, deploying, and evaluating real-world algorithmic decision-making processes design in order to maximize fairness and transparency.

Bellotti and Edwards addressed the issue of intelligible and accountable design of context-aware systems for both users and the system itself by considering the following questions "how will we know when information is captured, accessed, and used, by whom, and for what purposes in context-aware settings? How will this phenomenon make us feel? What measures can we take to ensure that we are aware of the implications of this and are also able to deal with them?" [3]

Another domain which is studied is recommender systems. There are differences in the effectiveness of recommender systems and there is a need to understand the demographic distribution of the underlying data. This is necessary since the largest subgroup of users usually dominate overall statistics. The effectiveness distribution of recommender systems across diverse groups of varying sizes needs to be demonstrated and considered (e.g. rebalancing data sets, exploring user profile size influences on recommendation quality, and the interaction between demographic biases and recommender evaluation [8]. Eslami et al. present a methodology for studying users' biases awareness and interactions in the area of hotel recommendation systems. They suggest intra-platform and inter-platform schemes for auditing algorithms in order to detect biases that can affect the results of the recommendations [9]. Beside recommendations for individuals, there is also the issue of group recommendations. Multiple users can have different preferences, which can increase the challenge in providing fair recommendations for the entire group [27]. Zemel et al. propose a fair classification learning algorithm which can achieve both individual and group fairness, by removing information about individuals' groups membership with respect to the protected class group [28].

Furthermore, many examples of AT research can be found in the information retrieval literature. Mowshowitz and Kawaguchi present a method for measuring search engines biases which uses the user queries as an input and produces a set of most relevant documents. According to their method, a "fair" results set is created by sampling results, for a given query across a collection of search engines. Their bias measure quantifies the extent to which a given results set, from a given engine, deviates from this ideal/fair distribution for the query [20].

III. Initial Results and Challenges

As a first step in this study, an initial literature review was conducted, related to the following areas of research: Recommender systems, HCI, Computer vision, and Information retrieval (search, text classification). We examined some relevant studies addressing problems related to algorithmic systems, AI, big data, machine learning, etc. Given the growing literature in the area of AT, we plan to continue to review additional work, which in practice, will extend our initial indentations and validate our initial model (or refine it).

A. Definitions

The first challenge we face is the different understanding of the terms diversity, bias, fairness and transparency in relation to algorithms. Here we provide a definition of the terms that we will use in this research. They are based on the Cambridge dictionary [8], and modified to be more cognizant of the algorithmic context.

Algorithm. a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem.

Algorithmic Systems. Figure 1 shows a general architecture of an algorithmic system. We define an algorithm (or a decision-making system) to have four main components:

1) Algorithmic Model (M), the algorithmic core, which is modulated by the three inputs below.

2) Training Data (D), used to train the model in the supervised learning algorithms.

3) Third Parties (T), are the meta parameters, given by third parties (not necessarily set by developers), to influence the design and performance of the model (M).

4) Fairness (F), methods to estimate fairness and to reduce biases from the components described above.

When the system receives Input (I) for a particular instance of the algorithm's operation, the model (M) performs computation based on these inputs and produces an Output (O).

Algorithms (or systems) can be further categorized into two types: black-box and white-box. White-box algorithms have more transparent models than black-box algorithms. For example, the decision tree algorithm is white-box since its decision and outcome is explainable by the model itself while

[1] http://www.cyprusconferences.org/umap2019/pages/cfp.html
[2] http://explainablesystems.comp.nus.edu.sg/2019/
[3] https://iuiatec.wordpress.com/
[4] https://recsys.acm.org/recsys19/call/#content-tab-1-0-tab
[5] https://www.ijcai19.org/call-for-papers.html
[6] https://www2019.thewebconf.org/accepted-papers
[7] http://cscw.acm.org/2019/submit-papers.html
[8] https://dictionary.cambridge.org/dictionary/english/

black-box algorithms include the deep neural networks and random forest algorithms whose decisions have no easily comprehended explanation.

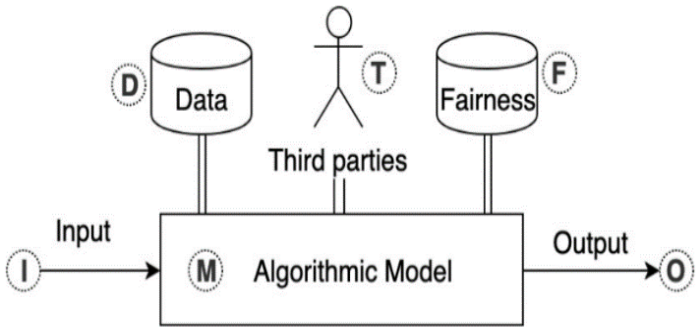

Fig 1. The general architecture of an algorithmic system.

Bias. The action of supporting or opposing a particular person or thing in an unfair way, by allowing personal opinions to influence judgment.

According to the Perception Institute[9] bias can be divided into two kinds:

- Explicit Bias (attitudes and beliefs we have about a person or group on a conscious level).
- Implicit Bias (attitudes or associated stereotypes towards people without conscious knowledge).

Algorithmic Bias. Algorithmic biases refer to biases exhibited by autonomous systems, which may have a computational origin, as well as biases arising from the inappropriate use of a system [5]. Danks and London also presented a taxonomy of biases and they detailed five sources of bias:

- training data.
- algorithmic focus (e.g. differential usage of attributes in the training data).
- algorithmic processing (e.g. use of a statistically biased estimator in a model).
- transfer context (e.g. application in a context differing from the one for which the system was developed).
- interpretation bias (e.g. user misinterpretation of the system output).

These biases can be classified into three main classes, as presented in Figure 2:

- *Data bias*, in which the input or training data is biased in some way (e.g. training data may contain information about sensitive attributes of people and such information is unbalanced and discriminatory to specific groups of people. Such biases can be fixed in some cases by simply removing the sensitive attributes or by employing fairness sampling or fairness learning).
- *Human bias*, in which the bias is caused by inappropriate system use by humans.
- *Algorithmic processing bias*, in which a system is biased in some way during algorithmic processing (e.g. biases that have occurred during the learning process of algorithms. Sometimes, mutual information of insensitive attributes can be representative of sensitive attributes so that the algorithms in practical cases catch such discriminatory rules unintentionally [18].

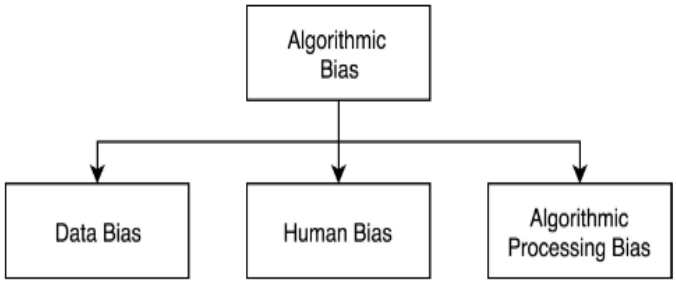

Fig 2. Classification of algorithmic biases.

Diversity. The fact of many different types of things or people being included in something; a range of different things or people.

Diversity in algorithms. The diversity of knowledge is reflected in the data used to train algorithms systems as biased data which can create an "unfair" algorithm. Furthermore, a system user base may be global, such that it serves individuals who perceive the world differently and will not interpret system behaviors in the same way [12]. The training data as well as the results of an algorithmic model can be influenced by the following dimensions of diversity [19]:

- Diversity of sources (multiplicity of sources).
- Diversity of resources (e.g., images, text).
- Diversity of topics.
- Diversity of speakers/actors/opinion holders (e.g., variety of political affiliation of opinion holders).
- Diversity of opinions.
- Diversity of genre (e.g., blogs, news, comments).
- Diversity of languages.
- Geographical/spatial diversity.
- Temporal diversity.

Fairness. The quality of treating people equally or in a way that is right or reasonable.

There is no consensus on the definition of fairness but there are 21 different definitions to fairness in the literature [21]. Fairness is subjective, what will be considered "fair" to one may be "un-fair" to others. According to Chiu et al. "Fairness is concerned with an individual's perceptions about the output/input ratio, the procedure that produces the outcome and

[9] https://perception.org/

the quality of interpersonal treatment" [4]. Therefore, we refer to it as *Perceived Fairness.*

Fairness can be classified into two classes [28]:

- *Individual Fairness* (similar individuals should be treated similarly).
- *Group Fairness* (the proportion of members in a protected group receiving positive classification is identical to the proportion in the population as a whole).

Fairness in algorithms. Dwork et al. presented a framework for characterizing fairness in classification. The main claim of this framework is that an algorithmic system can be considered as "fair" if similar people are being treated equally in the classification while still allowing a preferential treatment of individuals in the group. This approach can be used for certify fairness or for detecting unfairness in the system [7].

Transparency. The quality of being done in an open way without secrets.

Algorithmic Transparency. Algorithmic Transparency can serve multiple purposes (see Figure 3):

- *Discrimination Discovery*, which refers to the ability to identify discrimination against sensitive groups in the population, caused by biases in an algorithmic system.
- *Explainability Promotion*, which is the ability to explain the decisions made by algorithmic systems to users.
- *Fairness Managing*, which refers to the ability to ensure fairness with regard to sensitive groups in the population.
- *Auditing*, which refers to the ability to audit the results of the algorithm (e.g. study correlation between inputs/outputs [9])

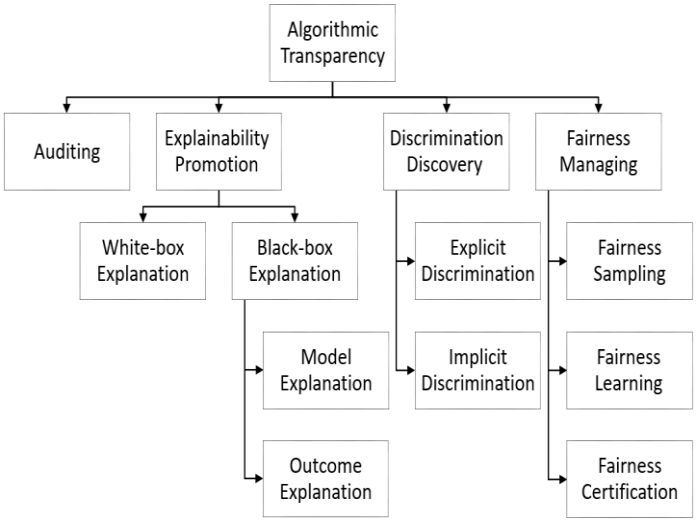

Fig 3. General classification of algorithmic transparency.

Regarding discrimination discovery, there are two types of discrimination: direct (explicit) discrimination [14] and indirect (implicit) discrimination [22]. Explicit discrimination is often caused by both: data bias and inappropriate use of sensitive attributes in algorithms, while implicit discrimination is caused by algorithmic processing bias and human bias due to the fact that some insensitive attributes are very informative about sensitive attributes.

White-box algorithms reveal their structure, making them transparent and easy to explain, while black-box algorithms hide their implementation details [1]. Therefore, Explainability Promotion, can refer both “white-box” and “black-box” explanations. These "black-box" explanations fill an intention gap between user's needs and interests and the system’s goals [23]. Guidotti et al. present a comprehensive survey with the aim of explaining black-box models [13].

To ensure fairness, three steps are required: sampling a subset by reducing the data bias (Fairness Sampling), learn an algorithm to be fair with the given fairness constraints (Fairness Learning) and, finally, verify whether algorithm satisfies fairness constraints (Fairness Certification) [15].

To manage fairness, Gajane and Pechenizkiy discussed whether fairness is considered as achieving parity or satisfying preferences and whether fairness needs to be measured in the treatment or in the impact (results). They suggest a formalization of fairness in the domain of machine learning algorithms [10].

In order to promote algorithmic transparency, the ACM created a list of the following seven principles: *Awareness, Access and Redress, Accountability, Explanation, Data Provenance, Auditability and Validation and Testing* [2], which are intended to support algorithmic decision-making in order to minimize potential biases and harms that can occur when using an algorithmic system. Despite the high level definition of such principles, there is still ambiguity regarding how to relate to them.

IV. Proposed Prototype

A. Research Goal and Questions

A range of stakeholders are affected by the behaviors of algorithmic processes, including developers who take part in the development of an algorithm and its use, as well as users, who may be affected by the consequences of the algorithm’s behaviors. Taking into account the ACM abstract principles as well as the large body of existing work, we would like to take a step further and propose a prototype for integrative framework for reducing biases, ensuring fairness and transparency in algorithmic systems.

Hence in order to answer the question: **“How we can ensure fairness in an algorithmic system?”**, we present the prototype of the system that can ensure fairness in the algorithmic system. Therefore, the main contribution of this research is to present a prototype towards a comprehensive "End to End" framework for detecting and reducing biases and promoting transparency in process: the regulator and the developer. The regulator is responsible for defining specs and requirements for transparent systems, according to which the system can be audited. The developer should follow such specs and make sure that according to both such predefined specs and to his own perceived fairness the system is fair. According to figure 4 the process is iterative and in the end of each iteration the discrimination discovery is performed. In case no discrimination

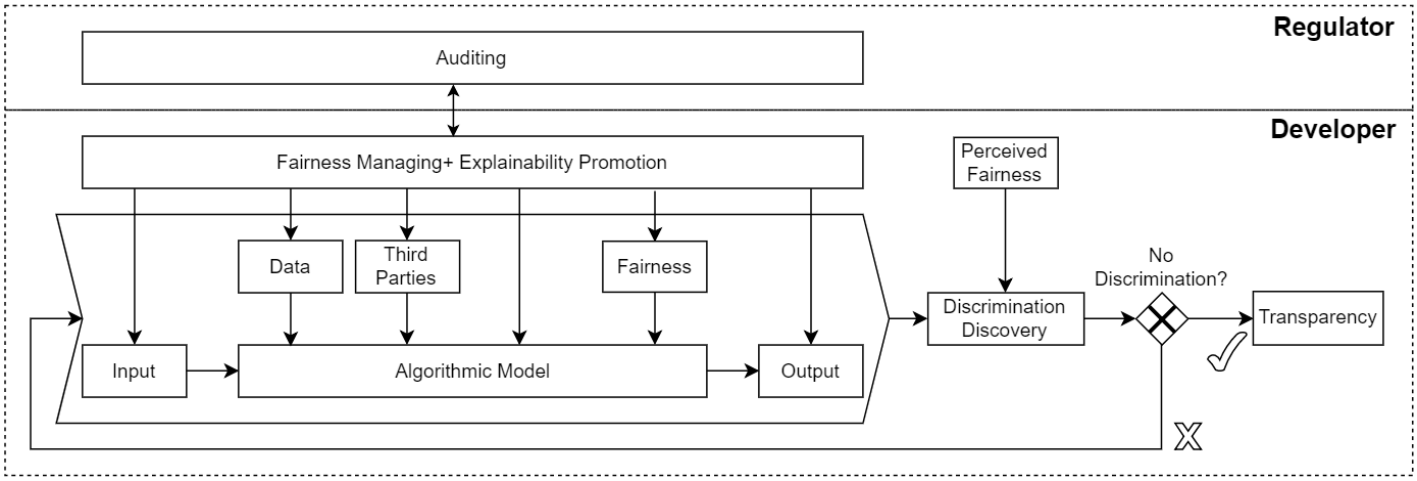

Fig 4. Framework Prototype.

is detected, the system is certificated as transparent. Otherwise, all process of discovering discrimination upon output on different parts of the system is repeated.

algorithmic systems. Two stakeholders are involved in this In order to address fairness managing and explainability promotion in the prototype, the following challenges are presented for each component of the algorithmic system (see table 1). Dealing with these challenges requires the creation of guidelines for developers, measurements for assessing fairness, tools for helping developers to ensure their systems are fair and even certification mechanisms as well as creating a convenient and understandable (accessible to users with variable levels of algorithmic maturity) method for the explanation of an algorithm and its results.

B. Addressing Challenges

We aim to address the challenges described above as follows:

Phase 1: Standardization of definitions. The first phase in the proposed research will be to gather, analyze and characterize different definitions (from different domains) that refer to algorithmic systems. In practice it will extend our initial results and validate our initial model (or refine it)

Phase 2: "Back End". In this phase, we will create guidelines and assess different measurements for the developer's side regarding the challenges (e.g. guidelines for development of bias-minimized algorithms in terms of data, third party, fairness and algorithmic model). The focus of this phase will be on developing methods and measurements for detecting and reducing bias in all the algorithmic model components, as well as methods for ensuring fairness and transparency.

Phase 3: "Front End". In this phase, we will characterize and analyze transparency reports and statements in order to create a transparency model. The focus of this phase will be on creating standardization for transparency reports and making them available and understandable to users.

Phase 4: Experiments and Evaluation. The last phase of the research will be to experiment with the usage of the whole framework and to analyze the results. Analysis will include comparisons of the whole "End to End" framework across different domains and diverse populations. Our intention is to compare algorithmic system fairness and transparency, from both the developers and public viewpoint, between system that were developed according to this framework and not developed according to this framework.

TABLE I. FAIRNESS MANAGING AND EXPLAINABILITY PROMOTION CHALLENGES.

System Component	Fairness Managing and Explainability Promotion Challenges
Input	Ensure fair input (e.g. sensitive parameters reduction).
Data	Ensure ethical principles on the storage, collection and use of personal data (e.g. GDPR). Verification of the data fairness (e.g. tools and techniques). Handling biased data (e.g. enrichment of missing data, rebalancing data).
Third Parties	Ensure "transparency" of that the system (e.g. meaningful algorithm description, transparency reports). Reducing third party's biases (e.g. personality questionnaire)
Algorithmic Model	Selection and implementation of the algorithmic model (e.g. different applications for different populations). Provide explanations that, on the one hand do not disclose trade secrets, but on the other hand explain the process (e.g. the accuracy versus transparency trade-off).
Fairness	Methods for identifying unfair systems (e.g. accuracy, fairness and bias testing and validation). Tools and measures for fairness.
Output	Minimize differences between developer and user perceived fairness (e.g. understanding politics, effectiveness, fairness, accuracy and transparency reports and models of the algorithmic system). Presentation of explanations

V. SUMMARY

As complex algorithmic systems become part of our daily lives, the issue of algorithmic biases, fairness and transparency increases. Many studies have discussed different definitions, measurements and methods for identifying and preventing biases and creating fair algorithms, but there is still no standardization in those areas. Our intention in this research is to take a step further in the standardization of the definitions and to propose a comprehensive and integrative "End to End" framework for detecting and avoiding biases and ensuring fairness in algorithmic systems by promoting algorithmic transparency.

ACKNOWLEDGMENT

This research has been partly supported by the Cyprus Center for Algorithmic Transparency, which has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 810105 (CyCAT – Call: H2020-WIDESPREAD-05-2017-Twinning).

REFERENCES

[1] Adler, P., Falk, C., Friedler, S. A., Nix, T., Rybeck, G., Scheidegger, C., ... & Venkatasubramanian, S. (2018). Auditing black-box models for indirect influence. Knowledge and Information Systems, 54(1), 95-122.

[2] https://www.acm.org/binaries/content/assets/publicpolicy/2017_usacm_statement_algorithms.pdf[

[3] Bellotti, V., & Edwards, K. (2001). Intelligibility and accountability: human considerations in context-aware systems. Human–Computer Interaction, 16(2-4), 193-212.

[4] Chiu, C. M., Lin, H. Y., Sun, S. Y., & Hsu, M. H. (2009). Understanding customers' loyalty intentions towards online shopping: an integration of technology acceptance model and fairness theory. Behaviour & Information Technology, 28(4), 347-360.

[5] David Danks and Alex John London. Algorithmic bias in autonomous systems. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pages 4691–4697. AAAI Press, 2017.

[6] Virginia Dignum. Responsible autonomy. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pages 4698–4704. AAAI Press, 2017.

[7] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012, January). Fairness through awareness. In Proceedings of the 3rd innovations in theoretical computer science conference (pp. 214-226). ACM.

[8] Ekstrand, M. D., Tian, M., Azpiazu, I. M., Ekstrand, J. D., Anuyah, O., McNeill, D., & Pera, M. S. (2018, January). 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 (pp. 172-186).

[9] Eslami, M., Vaccaro, K., Karahalios, K., & Hamilton, K. (2017, May). "Be careful; things can be worse than they appear": Understanding Biased Algorithms and Users' Behavior around Them in Rating Platforms. In Eleventh International AAAI Conference on Web and Social Media.

[10] Gajane, P., & Pechenizkiy, M. (2017). On formalizing fairness in prediction with machine learning. arXiv preprint arXiv:1710.03184.

[11] Fausto Giunchiglia, Vincenzo Maltese, and Biswanath Dutta. Domains and context: first steps towards managing diversity in knowledge. Web Semantics: Science, Services and Agents on the World Wide Web, 12:53–63, 2012.

[12] Fausto Giunchiglia. Managing diversity in knowledge. In ECAI 2006: 17th European Conference on Artificial Intelligence, volume 141, page 4. IOS Press, 2006.

[13] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 93.

[14] Hannák, A., Wagner, C., Garcia, D., Mislove, A., Strohmaier, M., & Wilson, C. (2017, February). Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (pp. 1914-1933). ACM.

[15] Kilbertus, N., Gascón, A., Kusner, M. J., Veale, M., Gummadi, K. P., & Weller, A. (2018). Blind justice: fairness with encrypted sensitive attributes. arXiv preprint arXiv:1806.03281.

[16] Kirkpatrick, K. (2016). Battling algorithmic bias: How do we ensure algorithms treat us fairly?. Communications of the ACM, 59(10), 16-17.

[17] Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. Philosophy & Technology, 31(4), 611-627.

[18] Madaan, N., Mehta, S., Agrawaal, T., Malhotra, V., Aggarwal, A., Gupta, Y., & Saxena, M. (2018, January). Analyze, detect and remove gender stereotyping from bollywood movies. In Conference on Fairness, Accountability and Transparency (pp. 92-105).

[19] Maltese, V., Giunchiglia, F., Denecke, K., Lewis, P., Wallner, C., Baldry, A., & Madalli, D. (2009). On the interdisciplinary foundations of diversity. University of Trento.

[20] Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. Information processing & management, 41(5), 1193-1205

[21] Narayanan, A. (2018, February). Translation tutorial: 21 fairness definitions and their politics. In Proc. Conf. Fairness Accountability Transp., New York, USA.

[22] Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F., Arvanitakis, G., Benevenuto, F., ... & Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising Till Speicher MPI-SWS MPI-SWS MPI-SWS. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*) (Vol. 81, pp. 1-15).

[23] Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. Journal of Management Information Systems, 23(4), 217-246

[24] Wasson, C. S. (2015). System engineering analysis, design, and development: Concepts, principles, and practices. John Wiley & Sons.

[25] Woolliams, P., & Gee, D. (1992). Accounting for user diversity in configuring online systems. Online Review, 16(5), 303-311.

[26] Robert H Wortham, Andreas Theodorou, and Joanna J Bryson. What does the robot think? Transparency as a fundamental design requirement for intelligent systems. In IJCAI-2016 Ethics for Artificial Intelligence Workshop, 2016.

[27] Yu, Z., Zhou, X., Hao, Y., & Gu, J. (2006). TV program recommendation for multiple viewers based on user profile merging. User modeling and user-adapted interaction, 16(1), 63-82.

[28] Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013, February). Learning fair representations. In International Conference on Machine Learning (pp. 325-333).