

TABLE OF CONTENTS

Special Section on Bias and Fairness in AI

- 1 Introduction to the Special Section on Bias and Fairness in AI: *Toon Calders, Eirini Ntoutsi, Mykola Pechenizkiy, Bodo Rosenhahn, and Salvatore Ruggieri*
- 4 Two Kinds of Discrimination in AI-Based Penal Decision-Making: *Dietmar Hübner*
- 14 On the Applicability of Machine Learning Fairness Notions: *Karima Makhoul, Sami Zhioua and Catuscia Palamidessi*
- 24 Gendering algorithms in social media: *Eduard Fosch-Villaronga, Adam Poulsen, Roger A. Søraa and Bart Custers*
- 32 Ethical Adversaries: Towards Mitigating Unfairness with Adversarial Machine Learning: *Pieter Delobelle, Paul Temple, Gilles Perrouin, Benoit Frénay, Patrick Heymans and Bettina Berendt*
- 42 Blind Spots in AI: the Role of Serendipity and Equity in Algorithm-Based Decision-Making: *Cora van Leeuwen, Annelien Smets and An Jacobs*
- 50 Assessing Viewpoint Diversity in Search Results Using Ranking Fairness Metrics: *Tim Draws, Nava Tintarev and Ujwal Gadiraju*
- Contributed Articles**
- 59 Generative Counterfactuals for Neural Networks via Attribute-Informed Perturbation: *Fan Yang, Ninghao Liu, Mengnan Du, and Xia Hu*
- 69 An Empirical Comparison of Bias Reduction Methods on Real-World Problems in High-Stakes Policy Settings: *Hemank Lamba, Kit T. Rodolfa and Rayid Ghani*
- 86 Adversarial Attacks and Defenses: An Interpretation Perspective: *Ninghao Liu, Mengnan Du, Ruocheng Guo, Huan Liu, and Xia Hu*

Editor-in-Chief:
Hanghang Tong

Associate Editors:
Xin Luna Dong
Ankur Teredesai
Reza Zafarani
<http://www.kdd.org/explorations/>



**Association for
Computing Machinery**

Advancing Computing as a Science & Profession