Preface: Fairness and Bias in AI: Insights from AEQUITAS 2023

Artificial Intelligence (AI) decision support systems are being used more widely across industries and sectors, including hiring, admissions, loans, healthcare, and crime prediction. However, with rising societal inequalities and discrimination, it's crucial to ensure that AI doesn't reiterate these issues. AI systems offer an opportunity to improve processes and repair injustices by identifying and addressing biases. Building trust in these systems requires understanding bias in AI and determining practical and ethically justified ways to mitigate it, which remains an ongoing challenge despite increased efforts in recent years.

In this context, the first edition of the AEQUITAS served as a workshop for the discussion of ideas, presentation of research findings, and sharing of preliminary work encompassing various facets of fairness and bias in AI. This preface sets the stage for the insightful invited talks and papers that emerged from the AEQUITAS 2023 workshop, co-located with the ECAI conference. It introduces the topics and the discussions that arose during the workshop, providing insight into the complex and multifaceted realm of fairness and bias in AI.

The 1st edition of the conference featured 12 presentations of high-quality papers. Accepted contributions ranged from foundational and theoretical results to practical experiences, case studies, and applications, and they covered a wide range of topics in the scope of technical, social and legal aspects of fairness and bias in AI.

In the opening talk, "Are we being fair about fairness in machine learning?" the first invited speaker provocatively questioned whether we, as scholars, researchers and innovators, are truly being fair in our pursuit of fairness itself. The audience was prompted to consider the broader, systemic aspects of fairness, challenging the conventional understanding that fairness can be neatly defined and achieved through mathematical algorithms alone. The second invited talk, "#breakthebias - towards an inclusive and equitable AI ecosystem" encouraged us to recognize the fundamental biases that pervade society and the necessity of mitigating these biases in AI systems.

The following papers presented at AEQUITAS 2023 showcased a remarkable diversity of perspectives and approaches to addressing fairness and bias in AI. In "Causal Fair Machine Learning via Rank-Preserving Interventional Distributions" the authors challenged us to think deeply about fairness, introducing a concept of normative equality based on causality. In "Visualizing Bias in Activations of Deep Neural Networks as Topographic Maps" the researchers invited us to see inside the 'black boxes' of deep neural networks, offering a visual means to identify and address biases in learned representations. The paper "Recommendations for Bias Mitigation Methods: Applicability and Legality" served as a reminder that fairness in AI is not solely an academic pursuit but a practical necessity, with real-world applicability and legal implications. Further, "Reasoning With Bias" presented a logical perspective on fairness, discussing the basic characteristics of biased systems and fairness metrics from a logical standpoint. "FAiRDAS: Fairness-Aware Ranking as Dynamic Abstract System" proposed a dynamic system approach to ensure long-term fairness in ranking systems, emphasizing the critical role of contrasting fairness metrics in these dynamic systems. In "Impact based fairness framework for socio-technical decision making" a novel framework was introduced to model the information flow in socio-technical decision systems, shedding light on the impact of biases. “EXTRACT: Explainable Transparent Control of Bias in Embeddings" delved into the realm of knowledge graphs and introduced methods to control and assess the presence of biases in embeddings. The paper “Fairness in job recommendations: estimating, explaining, and reducing gender gaps" tackled the specific issue of gender bias in job recommendation systems, providing valuable insights into algorithmic fairness. The paper “Gender Bias in Multimodal Models: A Transnational Feminist Approach" took a transnational feminist perspective to uncover and address gender biases in visual-linguistic multimodal models. In “A geometric framework for fairness" the researchers offered a novel, intuitive geometric framework to comprehend and address fairness in AI, making the complex world of fairness more accessible. Finally, “Symbolic AI (LFIT) for XAI to handle biases" and “Mitigating Bias in Language Models using Adversarial Learning" showcased innovative approaches to understanding and mitigating biases in AI models.

AEQUITAS 2023 was a journey of exploration, reflection, and collaboration. It underscored that fairness in AI is not a destination but a continuous search—a journey that requires research and innovation, and above all, the commitment to ensuring that AI benefits all of humanity equitably. Through the pages of this book, the reader will learn about insights, challenges, and solutions presented in the papers with the hope of further progress in making AI not just intelligent but fair and ethical.

The Editors,

Roberta Calegari
Andrea Aler Tubella
Gabriel González Castañe
Virginia Dignum
Michela Milano