

# AIEB 2024: Workshop on Implementing AI Ethics through a Behavioural Lens

Luca Nannini<sup>1</sup>, Annabel Gillard<sup>2</sup>, Caroline Friedman Levy<sup>3</sup>, Ali Ozkes<sup>4</sup> and Marija Slavkovik<sup>5</sup>

<sup>1</sup>*CITiUS USC and Minsait by Indra Sistemas, Spain*

<sup>2</sup>*Institute of Business Ethics, United Kingdom*

<sup>3</sup>*Consultative Network, UN Secretary General's Envoy on Technology, United States*

<sup>4</sup>*SKEMA Business School, France*

<sup>5</sup>*University of Bergen, Norway*

## Preface

The First Workshop on Implementing AI Ethics through a Behavioural Lens (AIEB 2024) was held on October 19, 2024, in Santiago de Compostela, Spain, co-located with the 26th European Conference on Artificial Intelligence (ECAI 2024). This workshop aimed to address a critical gap in the field of AI ethics: translating high-level ethical principles into tangible practices through the lens of behavioral science.

As artificial intelligence continues to advance rapidly, ensuring ethical development and deployment has become increasingly crucial. While numerous organizations have established AI ethics guidelines and principles, implementing these effectively remains a significant challenge. AIEB 2024 brought together researchers and practitioners from diverse disciplines, including AI, ethics, philosophy, psychology, and social sciences, to explore how insights from behavioral science can be leveraged to promote ethical AI development and governance.

The workshop received submissions addressing various aspects of ethical AI implementation through behavioral perspectives. After a thorough peer review process, we accepted 7 papers. The accepted papers cover a wide range of topics, including:

- Explanation methods for AI systems using counterfactual approaches
- Fairness-sensitive active learning techniques
- Novel procedures for fair data augmentation
- Ethical considerations in synthetic data generation
- Analysis of AI regulation and governance
- Frameworks for ethical human-AI collaboration
- Methods for evaluating AI trustworthiness

The workshop program featured paper presentations, expert panel discussions on ethical AI standards in organizations, and interactive sessions focused on practical interventions for ethical AI implementation. The panel brought together experts from academia, industry, and standardization bodies to discuss behavioral barriers to AI ethics standards adoption and implementation, ways to nudge ethical AI development, and approaches to measuring and incentivizing ethical behavior in AI development.

We would like to thank all authors for their valuable contributions, the program committee members for their reviews, and the invited speakers and panelists for sharing their expertise. We also express our gratitude to the ECAI 2024 workshop chairs and organizers for their support in hosting this event.

*AIEB 2024: Workshop on Implementing AI Ethics through a Behavioural Lens, Co-located with the 26th European Conference on Artificial Intelligence (ECAI 2024), October 19, 2024, Santiago de Compostela, Spain*

✉ [lucanannini2019@gmail.com](mailto:lucanannini2019@gmail.com) (L. Nannini)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## **Program Committee**

- Jose Maria Alonso, CiTIUS, Universidade de Santiago de Compostela
- Agathe Balayn, TU Delft; ServiceNow
- Senén Barro, CiTIUS, Universidade de Santiago de Compostela
- Leopoldo Bertossi, SKEMA Business School, Carleton University
- Alejandro Català Bolos, CiTIUS, Universidade de Santiago de Compostela
- Virginie Do, PRAIRIE Institute
- Marianna Ganapini, Union College; Notre Dame University; NYU
- Matthieu Hervouin, LAMSADE, Paris-Dauphine University
- Diletta Husykes, Università di Milano Statale; Immanence; Privacy Network
- Michael Klenk, TU Delft
- Stuart Mills, University of Leeds
- Anne-Marie Nussberger, Center for Humans and Machines, Max Planck Institute for Human Development
- Enrico Panai, Università Cattolica del Sacro Cuore; Emlyon Business School
- Giada Pistilli, Sorbonne Université; HuggingFace
- Adam Leon Smith, CTO Dragonfly
- Clemens Stachl, University of St. Gallen
- Alessio Tartaro, University of Sassari
- Jurgen Willems, Institute for Public Management and Governance, WU Vienna

## **Workshop Chairs**

- Luca Nannini, CITiUS USC, Spain
- Annabel Gillard, Institute of Business Ethics, United Kingdom
- Caroline Friedman Levy, Consultative Network, UN Secretary General's Envoy on Technology, United States
- Ali Ozkes, SKEMA Business School, France
- Marija Slavkovik, University of Bergen, Norway

## **Declaration on Generative AI**

The author(s) have not employed any Generative AI tools.