

2025 IEEE Symposium on Trustworthy, Explainable and Responsible Computational Intelligence (CITREx 2025)

**Trondheim, Norway
17-20 March 2025**



**IEEE Catalog Number: CFP254B7-POD
ISBN: 979-8-3315-2016-8**

**Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP254B7-POD
ISBN (Print-On-Demand):	979-8-3315-2016-8
ISBN (Online):	979-8-3315-2015-1

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Contents

Enhancing patient stratification and interpretability through class-contrastive and feature attribution techniques **1**

*Sharday Olowu (University of Cambridge); Neil Lawrence (University of Cambridge); Soumya Banerjee (University of Cambridge)**

XAI NT: eXplainable AI through Iterative Network Truncation **8**

Andrea Coppari (Spindox S.p.a.); Luca Capra (Spindox S.p.a.); Cristiano Carlevaro (Spindox S.p.a.)*

Decomposing Deep Learning Predictions to Uncover Beta-2 Adrenergic Conformational Landscapes **15**

Mario Gutierrez Mondragon (Universitat Polytechnica de Catalunya); Caroline König (Universitat Politècnica de Catalunya); Alfredo Vellido (Universitat Politècnica de Catalunya)*

Pool-based Active Learning with Decision Trees: Incorporate the Tree Structure to Explore and Exploit **22**

Marvin Schöne (University of Applied Sciences and Arts Bielefeld); Bjarne Jaster (University of Applied Sciences and Art Bielefeld); Julian Bültemeier (Technische Hochschule Ostwestfalen-Lippe); Justus Kösters (University of Applied Sciences and Arts Bielefeld); Christoph-Alexander Holst (Technische Hochschule Ostwestfalen-Lippe); Martin Kohlhasse (University of Applied Sciences and Arts Bielefeld)*

Fairness and Trust in Data-Driven Decisions: Analyzing Discrepancies in Ordinal Decisions **31**

Edward Abel (University of Southern Denmark); Sajid Siraj (University of Leeds)*

Improving interpretability for fault diagnosis in complex chemical processes using combined regularization and linearization in neural network modeling **38**

Pedro Silva (Federal University of Minas Gerais); Yan Ferreira (Federal University of Minas Gerais); Igor Lima (Federal University of Minas Gerais); Esa Vakkilainen (LUT University); Gustavo Almeida (LUT University); Antônio Braga (Federal University of Minas Gerais)*

RO-FIGS: Efficient and Expressive Tree-Based Ensembles for Tabular Data **45**

Urška Matjašec (University of Cambridge); Nikola Simidjievski (University of Cambridge); Mateja Jamnik (University of Cambridge)*

Can LLMs Identify Event Causality More Accurately through Debate? A Systematic Assessment of LLMs' Factuality and Reasoning **52**

*Yiheng Zhao (Concordia); Jun Yan (Concordia University)**

Can LIME Make Interactive Multiobjective Optimization Methods Explainable? **60**

Bekir Afsar (University of Jyväskylä); Giomara Lárraga Maldonado (University of*

Jyväskylä); Kaisa Miettinen (University of Jyväskylä)

AI Literacy and Governance of Explainable AI: A Cross-Jurisdictional Review of National AI Strategies **67**

Ayaz Karimov (University of Jyväskylä); Mirka Saarela (University of Jyväskylä)*

ENHANCING EXPLAINABILITY, ROBUSTNESS, AND AUTONOMY: A COMPREHENSIVE APPROACH IN TRUSTWORTHY AI **74**

Mobyen Uddin Ahmed (Mälardalen University); Shahina Begum (Mälardalen University); Shaibal Barua (Mälardalen University); Abu Naser Masud (Mälardalen University); Gianluca Di Flumeri (Sapienza University of Rome); Nicolò Navarin (Università di Padova)*

Impacts of Anthropogenic Noise on the House Wren's Song: An XAI Approach to Bioacoustic Insights **81**

Larissa Silva (Universidade Federal do Amazonas - ICOMP); Juan G. Colonna (Federal University of Amazonas); Bernardo Gatto (MTI Ltd.); Marcelo Protázio (ICEN - UFPA)*

Algorithmic Bias Detection: A Focus on Skin Tone and Gender Fairness in Computer Vision Models **88**

Peter Taiwo (Morgan State University); Happy Iguare (Morgan State University)*

Explainability of Deep Neural Networks using the Constructive Axiom of Choice Model **94**
*Steeph ECKOUBILI (Devoteam)**