37th AAAI Conference on Artificial Intelligence (AAAI-23)

Volume 12: AAAI Technical Tracks

- AI for Social Impact
- Safe and Robust AI

Washington, DC, USA
7-14 February 2023

Part 1 of 2

ISBN: 978-1-7138-7822-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by Association for the Advancement of Artificial Intelligence
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact Association for the Advancement of Artificial Intelligence
at the address below.

Association for the Advancement of Artificial Intelligence
2275 East Bayshore Road
Suite 160
Palo Alto, California 94303
USA

Phone: 1-650-328-3123
Fax: 1-650-321-4457

https://aaai.org/Press/press.php

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
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

PART 1

AAAI SPECIAL TRACK ON AI FOR SOCIAL IMPACT

COSMOS: Catching Out-of-Context Image Misuse Using Self-Supervised Learning 14084
Shivangi Aneja, Chris Bregler, Matthias Niessner

Med-EASi: Finely Annotated Dataset and Models for Controllable Simplification of Medical Texts......... 14093
Chandrayee Basu, Rosni Vasu, Michihiro Yasunaga, Qian Yang

On the Challenges of Using Reinforcement Learning in Precision Drug Dosing: Delay and Prolongedness of Action Effects... 14102
Sumana Basu, Marc-André Legault, Adriana Romero-Soriano, Doina Precup

On the Cost of Demographic Parity in Influence Maximization 14110
Ruben Becker, Gianlorenzo D'Angelo, Sajjad Ghobadi

Improving Fairness in Information Exposure by Adding Links 14119
Ruben Becker, Gianlorenzo D'Angelo, Sajjad Ghobadi

A Fair Incentive Scheme for Community Health Workers 14127
Avinandan Bose, Tracey Li, Arunesh Sinha, Tien Mai

Rehabilitating Homeless: Dataset and Key Insights.. 14136
Anna Bykova, Nikolay Filippov, Ivan P. Yamshchikov

Counterfactuals for the Future.. 14144
Lucius E. J. Bynum, Joshua R. Loftus, Julia Stoyanovich

Towards Learning to Discover Money Laundering Sub-network in Massive Transaction Network............ 14153
Ziwei Chai, Yang Yang, Jiawang Dan, Sheng Tian, Changhua Meng, Weiqiang Wang, Yifei Sun

Estimating Geographic Spillover Effects of COVID-19 Policies from Large-Scale Mobility Networks 14161
Serina Chang, Damir Vrabac, Jure Leskovec, Johan Ugander

Seq2Seq Surrogates of Epidemic Models to Facilitate Bayesian Inference 14170
Giovanni Charles, Timothy M. Wolock, Peter Winskill, Azra Ghani, Samir Bhatt, Seth Flaxman

Leveraging Old Knowledge to Continually Learn New Classes in Medical Images.................................. 14178
Evelyn Chee, Mong Li Lee, Wynne Hsu

SARAS-Net: Scale and Relation Aware Siamese Network for Change Detection...................................... 14187
Chao-Peng Chen, Jun-Wei Hsieh, Ping-Yang Chen, YI-Kuan Hsieh, Bor-Shiun Wang

Improving Interpretability of Deep Sequential Knowledge Tracing Models with Question-centric Cognitive Representations... 14196
Jiahao Chen, Zitao Liu, Shuyan Huang, Qiongqiong Liu, Weiqi Luo

Critical Firms Prediction for Stemming Contagion Risk in Networked-Loans through Graph-Based Deep Reinforcement Learning 14205
Dawei Cheng, Zhibin Niu, Jianfu Zhang, Yiyi Zhang, Changjun Jiang

GAN-Based Domain Inference Attack 14214
Yuechun Gu, Keke Chen

Physics Guided Neural Networks for Time-Aware Fairness: An Application in Crop Yield Prediction 14223
Erhu He, Yiqun Xie, Licheng Liu, Weiye Chen, Zhenong Jin, Xiaowei Jia

“Nothing Abnormal”: Disambiguating Medical Reports via Contrastive Knowledge Infusion 14232
Zexue He, An Yan, Amilcare Gentili, Julian McAuley, Chun-Nan Hsu

MTDiag: An Effective Multi-Task Framework for Automatic Diagnosis 14241
Zhenyu Hou, Yukuo Cen, Ziding Liu, Dongxue Wu, Baoyan Wang, Xuanhe Li, Lei Hong, Jie Tang

Walkability Optimization: Formulations, Algorithms, and a Case Study of Toronto 14249
Weimin Huang, Elias B. Khalil

Low Emission Building Control with Zero-Shot Reinforcement Learning 14259
Scott Jeen, Alessandro Abate, Jonathan M. Cullen

Spatio-Temporal Graph Neural Point Process for Traffic Congestion Event Prediction 14268
Guangyin Jin, Lingbo Liu, Fuxian Li, Jincai Huang

Taxonomizing and Measuring Representational Harms: A Look at Image Tagging 14277
Jared Katzman, Angelina Wang, Morgan Scheuerman, Su Lin Blodgett, Kristen Laird, Hanna Wallach, Solon Barocas

Winning the CityLearn Challenge: Adaptive Optimization with Evolutionary Search under Trajectory-Based Guidance 14286
Vanshaj Khattar, Ming Jin

Robust Planning over Restless Groups: Engagement Interventions for a Large-Scale Maternal Telehealth Program 14295
Jackson A. Killian, Arpita Biswas, Lily Xu, Shresth Verma, Vineet Nair, Aparna Taneja, Aparna Hegde, Neha Madhiwalla, Paula Rodriguez Diaz, Sonja Johnson-Yu, Milind Tambe

Equivariant Message Passing Neural Network for Crystal Material Discovery 14304
Astrid Klipfel, Zied Bouraoui, Olivier Peltre, Yaël Fregier, Najwa Harrati, Adlane Sayede

Accurate Fairness: Improving Individual Fairness without Trading Accuracy 14312
Xuran Li, Peng Wu, Jing Su

Point-to-Region Co-learning for Poverty Mapping at High Resolution Using Satellite Imagery 14321
Zhili Li, Yiqun Xie, Xiaowei Jia, Kara Stuart, Caroline Delaire, Sergii Skakun

AirFormer: Predicting Nationwide Air Quality in China with Transformers 14329
Yuxuan Liang, Yutong Xia, Songyu Ke, Yiwei Wang, Qingsong Wen, Junbo Zhang, Yu Zheng, Roger Zimmermann

SimFair: A Unified Framework for Fairness-Aware Multi-Label Classification 14338
Tianci Liu, Haoyu Wang, Yaqing Wang, Xiaoqian Wang, Lu Su, Jing Gao

Human Mobility Modeling during the COVID-19 Pandemic via Deep Graph Diffusion Infomax 14347
Yang Liu, Yu Rong, Zhuoning Guo, Nuo Chen, Tingyang Xu, Fugee Tsung, Jia Li

Interpretable Chirality-Aware Graph Neural Network for Quantitative Structure Activity Relationship Modeling in Drug Discovery 14356
Yunchao (Lance) Liu, Yu Wang, Oanh Vu, Rocco Moretti, Bobby Bodenheimer, Jens Meiler, Tyler Derr

Task-Adaptive Meta-Learning Framework for Advancing Spatial Generalizability 14365
Zhexiong Liu, Licheng Liu, Yiqun Xie, Zhenong Jin, Xiaowei Jia

A Composite Multi-Attention Framework for Intraoperative Hypotension Early Warning 14374
Feng Lu, Wei Li, Zhiqiang Zhou, Cheng Song, Yifei Sun, Yuwei Zhang, Yufei Ren, Xiaofei Liao, Hai Jin, Ailin Luo, Albert Y. Zomaya

Bugs in the Data: How ImageNet Misrepresents Biodiversity 14382
Alexandra Sasha Luccioni, David Rolnick

LUCID: Exposing Algorithmic Bias through Inverse Design 14391
Carmen Mazijn, Carina Prunkl, Andres Algaba, Jan Danckaert, Vincent Ginis

Neighbor Auto-Grouping Graph Neural Networks for Handover Parameter Configuration in Cellular Network 14400
Mehrtash Mehrabi, Walid Masoudimansour, Yingxue Zhang, Jie Chuai, Zhitang Chen, Mark Coates, Jianye Hao, Yanhui Geng

Help Me Heal: A Reinforced Polite and Empathetic Mental Health and Legal Counseling Dialogue System for Crime Victims 14408
Kshitij Mishra, Priyanshu Priya, Asif Ekbal

Carburacy: Summarization Models Tuning and Comparison in Eco-Sustainable Regimes with a Novel Carbon-Aware Accuracy 14417
Gianluca Moro, Luca Ragazzi, Lorenzo Valgimigli

Joint Self-Supervised Image-Volume Representation Learning with Intra-inter Contrastive Clustering 14426
Duy M. H. Nguyen, Hoang Nguyen, Truong T. N. Mai, Tri Cao, Binh T. Nguyen, Nhat Ho, Paul Swoboda, Shadi Albarqouni, Pengtao Xie, Daniel Sonntag

For the Underrepresented in Gender Bias Research: Chinese Name Gender Prediction with Heterogeneous Graph Attention Network 14436
Zihao Pan, Kai Peng, Shuai Ling, Haipeng Zhang

FakeSV: A Multimodal Benchmark with Rich Social Context for Fake News Detection on Short Video Platforms 14444
Peng Qi, Yuyan Bu, Juan Cao, Wei Ji, Ruihao Shui, Junbin Xiao, Danding Wang, Tat-Seng Chua

EINNs: Epidemiologically-Informed Neural Networks 14453
Alexander Rodríguez, Jiaming Cui, Naren Ramakrishnan, Bijaya Adhikari, B. Aditya Prakash

Counterfactual Fairness Is Basically Demographic Parity 14461
Lucas Rosenblatt, R. Teal Witter

Detecting Anomalous Networks of Opioid Prescribers and Dispensers in Prescription Drug Data 14470
Katie Rosman, Daniel B. Neill

Practical Disruption of Image Translation Deepfake Networks 14478
Nataniel Ruiz, Sarah Adel Bargal, Cihang Xie, Stan Sclaroff

Daycare Matching in Japan: Transfers and Siblings 14487
Zhaohong Sun, Yoshihiro Takenami, Daisuke Moriwaki, Yoji Tomita, Makoto Yokoo

City-Scale Pollution Aware Traffic Routing by Sampling Max Flows Using MCMC 14496
Shreevignesh Suriyanarayanan, Praveen Paruchuri, Girish Varma

Weather2vec: Representation Learning for Causal Inference with Non-local Confounding in Air Pollution and Climate Studies 14504
Mauricio Tec, James G. Scott, Corwin M. Zigler

Evaluating Digital Agriculture Recommendations with Causal Inference 14514
Ilias Tsoumas, Georgios Giannarakis, Vasileios Sitokonstantinou, Alkiviadis Koukos, Dimitra Loka, Nikolaos Bartsotas, Charalampos Kontoes, Ioannis Athanasiadis

Everyone's Voice Matters: Quantifying Annotation Disagreement Using Demographic Information 14523
Ruyuan Wan, Jaehyung Kim, Dongyeop Kang

MixFairFace: Towards Ultimate Fairness via MixFair Adapter in Face Recognition 14531
Fu-En Wang, Chien-Yi Wang, Min Sun, Shang-Hong Lai

PateGail: A Privacy-Preserving Mobility Trajectory Generator with Imitation Learning 14539
Huandong Wang, Changzheng Gao, Yuchen Wu, Depeng Jin, Lina Yao, Yong Li

Noise Based Deepfake Detection via Multi-Head Relative-Interaction 14548
Tianyi Wang, Kam Pui Chow

Semi-supervised Credit Card Fraud Detection via Attribute-Driven Graph Representation 14557
Sheng Xiang, Mingzhi Zhu, Dawei Cheng, Enxia Li, Ruihui Zhao, Yi Ouyang, Ling Chen, Yefeng Zheng

Privacy-Preserved Evolutionary Graph Modeling via Gromov-Wasserstein Autoregression 14566
Yue Xiang, Dixin Luo, Hongteng Xu

Auto-CM: Unsupervised Deep Learning for Satellite Imagery Composition and Cloud Masking Using Spatio-Temporal Dynamics 14575
Yiqun Xie, Zhili Li, Han Bao, Xiaowei Jia, Dongkuan Xu, Xun Zhou, Sergii Skakun

ERASER: AdvERsArial Sensitive Element Remover for Image Privacy Preservation 14584
Guang Yang, Juan Cao, Danding Wang, Peng Qi, Jintao Li

Deep Learning on a Healthy Data Diet: Finding Important Examples for Fairness 14593
Abdelrahman Zayed, Prasanna Parthasarathi, Gonçalo Mordido, Hamid Palangi, Samira Shabanian, Sarath Chandar

On the Effectiveness of Curriculum Learning in Educational Text Scoring 14602
Zijie Zeng, Dragan Gasevic, Guangliang Chen

Censored Fairness through Awareness 14611
Wenbin Zhang, Tina Hernandez-Boussard, Jeremy Weiss

A Continual Pre-training Approach to Tele-Triaging Pregnant Women in Kenya 14620
Wenbo Zhang, Hangzhi Guo, Prerna Ranganathan, Jay Patel, Sathyanath Rajasekharan, Nidhi Danayak, Manan Gupta, Amulya Yadav

Future Aware Pricing and Matching for Sustainable On-Demand Ride Pooling 14628
Xianjie Zhang, Pradeep Varakantham, Hao Jiang

A Crowd-AI Collaborative Duo Relational Graph Learning Framework towards Social Impact Aware Photo Classification 14637
Yang Zhang, Ziyi Kou, Lanyu Shang, Huimin Zeng, Zhenrui Yue, Dong Wang

People Taking Photos That Faces Never Share: Privacy Protection and Fairness Enhancement from Camera to User 14646
Junjie Zhu, Lin Gu, Xiaoxiao Wu, Zheng Li, Tatsuya Harada, Yingying Zhu

OpenMapFlow: A Library for Rapid Map Creation with Machine Learning and Remote Sensing Data 14655
Ivan Zvonkov, Gabriel Tseng, Catherine Nakalembe, Hannah Kerner

AAAI SPECIAL TRACK ON SAFE AND ROBUST AI

Formally Verified SAT-Based AI Planning 14665
Mohammad Abdulaziz, Friedrich Kurz

Shielding in Resource-Constrained Goal POMDPs 14674
Michal Ajdarów, Šimon Brlej, Petr Novotný

Implicit Bilevel Optimization: Differentiating through Bilevel Optimization Programming 14683
Francesco Alesiani

Query-Based Hard-Image Retrieval for Object Detection at Test Time 14692
Edward Ayers, Jonathan Sadeghi, John Redford, Romain Mueller, Puneet K. Dokania

Probabilities Are Not Enough: Formal Controller Synthesis for Stochastic Dynamical Models with Epistemic Uncertainty 14701
Thom Badings, Licio Romao, Alessandro Abate, Nils Jansen

PART 2

Accelerating Inverse Learning via Intelligent Localization with Exploratory Sampling 14711
Sirui Bi, Victor Fung, Jiaxin Zhang

Attention-Conditioned Augmentations for Self-Supervised Anomaly Detection and Localization 14720
Behzad Bozorgtabar, Dwarikanath Mahapatra

Robust-by-Design Classification via Unitary-Gradient Neural Networks 14729
Fabio Brau, Giulio Rossolini, Alessandro Biondi, Giorgio Buttazzo

Ensemble-in-One: Ensemble Learning within Random Gated Networks for Enhanced Adversarial Robustness 14738
Yi Cai, Xuefei Ning, Huazhong Yang, Yu Wang

Safe Reinforcement Learning via Shielding under Partial Observability 14748
Steven Carr, Nils Jansen, Sebastian Junges, Ufuk Topcu

PowRL: A Reinforcement Learning Framework for Robust Management of Power Networks 14757
Anandsingh Chauhan, Mayank Baranwal, Ansuma Basumatary

Two Wrongs Don't Make a Right: Combating Confirmation Bias in Learning with Label Noise 14765
Mingcai Chen, Hao Cheng, Yuntao Du, Ming Xu, Wenyu Jiang, Chongjun Wang

Testing the Channels of Convolutional Neural Networks 14774
Kang Choi, Donghyun Son, Younghoon Kim, Jiwon Seo

Feature-Space Bayesian Adversarial Learning Improved Malware Detector Robustness 14783
Bao Gia Doan, Shuiqiao Yang, Paul Montague, Olivier De Vel, Tamas Abraham, Seyit Camtepe, Salil S. Kanhere, Ehsan Abbasnejad, Damith C. Ranashinghe

Correct-by-Construction Reinforcement Learning of Cardiac Pacemakers from Duration Calculus Requirements 14792
Kalyani Dole, Ashutosh Gupta, John Komp, Shankaranarayanan Krishna, Ashutosh Trivedi

SafeLight: A Reinforcement Learning Method toward Collision-Free Traffic Signal Control 14801
Wenlu Du, Junyi Ye, Jingyi Gu, Jing Li, Hua Wei, Guiling Wang

PatchNAS: Repairing DNNs in Deployment with Patched Network Architecture Search 14811
Yuchu Fang, Wenzhong Li, Yao Zeng, Yang Zheng, Zheng Hu, Sanglu Lu

Similarity Distribution Based Membership Inference Attack on Person Re-identification 14820
Junyao Gao, Xinyang Jiang, Huishuai Zhang, Yifan Yang, Shuguang Dou, Dongsheng Li, Duoqian Miao, Cheng Deng, Cairong Zhao

Out-of-Distribution Detection Is Not All You Need 14829
Joris Guerin, Kevin Delmas, Raul Ferreira, Jérémie Guiochet

Contrastive Self-Supervised Learning Leads to Higher Adversarial Susceptibility 14838
Rohit Gupta, Naveed Akhtar, Ajmal Mian, Mubarak Shah

AutoCost: Evolving Intrinsic Cost for Zero-Violation Reinforcement Learning 14847
Tairan He, Weiye Zhao, Changliu Liu

Test Time Augmentation Meets Post-hoc Calibration: Uncertainty Quantification under Real-World Conditions 14856
Achim Hekler, Titus J. Brinker, Florian Buettner

Robust Training of Neural Networks against Bias Field Perturbations 14865
Patrick Henriksen, Alessio Lomuscio

Redactor: A Data-Centric and Individualized Defense against Inference Attacks 14874
Geon Heo, Steven Euijong Whang

Improving Adversarial Robustness with Self-Paced Hard-Class Pair Reweighting 14883
Pengyue Hou, Jie Han, Xingyu Li

CodeAttack: Code-Based Adversarial Attacks for Pre-trained Programming Language Models 14892
Akshita Jha, Chandan K. Reddy

Formalising the Robustness of Counterfactual Explanations for Neural Networks 14901
Junqi Jiang, Francesco Leofante, Antonio Rago, Francesca Toni

READ: Aggregating Reconstruction Error into Out-of-Distribution Detection 14910
Wenyu Jiang, Yuxin Ge, Hao Cheng, Mingcai Chen, Shuai Feng, Chongjun Wang

Sample-Dependent Adaptive Temperature Scaling for Improved Calibration 14919
Tom Joy, Francesco Pinto, Ser-Nam Lim, Philip H.S. Torr, Puneet K. Dokania

Heuristic Search in Dual Space for Constrained Fixed-Horizon POMDPs with Durative Actions 14927
Majid Khonji, Duoaa Khalifa

Iteratively Enhanced Semidefinite Relaxations for Efficient Neural Network Verification 14937
Jianglin Lan, Yang Zheng, Alessio Lomuscio

A Semidefinite Relaxation Based Branch-and-Bound Method for Tight Neural Network Verification 14946
Jianglin Lan, Benedikt Brückner, Alessio Lomuscio

Robust Image Steganography: Hiding Messages in Frequency Coefficients 14955
Yuhang Lan, Fei Shang, Jianhua Yang, Xiangui Kang, Enping Li

Quantization-Aware Interval Bound Propagation for Training Certifiably Robust Quantized Neural Networks 14964
Mathias Lechner, Đorđe Žikelić, Krishnendu Chatterjee, Thomas A. Henzinger, Daniela Rus

Revisiting the Importance of Amplifying Bias for Debiasing 14974
Jungsoo Lee, Jeonghoon Park, Daeyoung Kim, Juyoung Lee, Edward Choi, Jaegul Choo

WAT: Improve the Worst-Class Robustness in Adversarial Training 14982
Boqi Li, Weiwei Liu

PLMmark: A Secure and Robust Black-Box Watermarking Framework for Pre-trained Language Models.. 14991
Peixuan Li, Pengzhou Cheng, Fangqi Li, Wei Du, Haodong Zhao, Gongshen Liu

Rethinking Label Refurbishment: Model Robustness under Label Noise.. 15000
Yangdi Lu, Zhiwei Xu, Wenbo He

A Holistic Approach to Undesired Content Detection in the Real World... 15009
Todor Markov, Chong Zhang, Sandhini Agarwal, Florentine Eloundou Nekoul, Theodore Lee, Steven Adler, Angela Jiang, Lilian Weng

A Risk-Sensitive Approach to Policy Optimization 15019
Jared Markowitz, Ryan W. Gardner, Ashley Llorens, Raman Arora, I-Jeng Wang

Anonymization for Skeleton Action Recognition... 15028
Saemi Moon, Myeonghyeon Kim, Zhenyue Qin, Yang Liu, Dongwoo Kim

Monitoring Model Deterioration with Explainable Uncertainty Estimation via Non-parametric Bootstrap 15037
Carlos Mougan, Dan Saattrup Nielsen

Certified Policy Smoothing for Cooperative Multi-Agent Reinforcement Learning 15046
Ronghui Mu, Wenjie Ruan, Leandro Soriano Marcolino, Gaojie Jin, Qiang Ni

Constrained Reinforcement Learning in Hard Exploration Problems... 15055
Pathmanathan Pankayaraj, Pradeep Varakantham

Defending from Physically-Realizable Adversarial Attacks through Internal Over-Activation Analysis.. 15064
Giulio Rossolini, Federico Nesti, Fabio Brau, Alessandro Biondi, Giorgio Buttazzo

Formally Verified Solution Methods for Markov Decision Processes... 15073
Maximilian Schäffeler, Mohammad Abdulaziz

Improving Training and Inference of Face Recognition Models via Random Temperature Scaling............ 15082
Lei Shang, Mouxiao Huang, Wu Shi, Yuchen Liu, Yang Liu, Wang Steven, Baigui Sun, Xuansong Xie, Yu Qiao

Task and Model Agnostic Adversarial Attack on Graph Neural Networks 15091
Kartik Sharma, Samidha Verma, Sourav Medya, Arnab Bhattacharya, Sayan Ranu

Robust Sequence Networked Submodular Maximization 15100
Qihao Shi, Bingyang Fu, Can Wang, Jiawei Chen, Sheng Zhou, Yan Feng, Chun Chen

Safe Policy Improvement for POMDPs via Finite-State Controllers 15109
Thiago D. Simão, Marnix Suilen, Nils Jansen

STL-Based Synthesis of Feedback Controllers Using Reinforcement Learning..15118
Nikhil Kumar Singh, Indranil Saha

Understanding and Enhancing Robustness of Concept-Based Models 15127
Sanchit Sinha, Mengdi Huai, Jianhui Sun, Aidong Zhang

Misspecification in Inverse Reinforcement Learning... 15136
Joar Skalse, Alessandro Abate

Planning and Learning for Non-markovian Negative Side Effects Using Finite State Controllers.............. 15144
Aishwarya Srivastava, Sandhya Saisubramanian, Praveen Paruchuri, Akshat Kumar, Shlomo Zilberstein

Toward Robust Uncertainty Estimation with Random Activation Functions.. 15152
Yana Stoyanova, Soroush Ghandi, Maryam Tavakol

Improving Robust Fariness via Balance Adversarial Training 15161
Chunyu Sun, Chenye Xu, Chengyuan Yao, Siyuan Liang, Yichao Wu, Ding Liang, Xianglong Liu, Aishan Liu

DPAUC: Differentially Private AUC Computation in Federated Learning... 15170
Jiankai Sun, Xin Yang, Yuanshun Yao, Junyuan Xie, Di Wu, Chong Wang

Conflicting Interactions among Protection Mechanisms for Machine Learning Models 15179
Sebastian Szyller, N. Asokan

Neural Policy Safety Verification via Predicate Abstraction: CEGAR... 15188
Marcel Vinzent, Siddhant Sharma, Jöerg Hoffmann

Towards Verifying the Geometric Robustness of Large-Scale Neural Networks 15197
Fu Wang, Peipei Xu, Wenjie Ruan, Xiaowei Huang

Revisiting Item Promotion in GNN-Based Collaborative Filtering: A Masked Targeted Topological Attack Perspective.. 15206
Yongwei Wang, Yong Liu, Zhiqi Shen

Robust Average-Reward Markov Decision Processes... 15215
Yue Wang, Alvaro Velasquez, George Atia, Ashley Prater-Bennette, Shaofeng Zou

Robust Graph Meta-Learning via Manifold Calibration with Proxy Subgraphs 15224
Zhenzhong Wang, Lulu Cao, Wanyu Lin, Min Jiang, Kay Chen Tan

HOTCOLD Block: Fooling Thermal Infrared Detectors with a Novel Wearable Design 15233
Hui Wei, Zhixiang Wang, Xuemei Jia, Yinqiang Zheng, Hao Tang, Shin'ichi Satoh, Zheng Wang

Beyond NaN: Resiliency of Optimization Layers in the Face of Infeasibility 15242
Wai Tuck Wong, Sarah Kinsey, Ramesha Karunasena, Thanh H. Nguyen, Arunesh Sinha

DeepGemini: Verifying Dependency Fairness for Deep Neural Network.. 15251
Xuan Xie, Fuyuan Zhang, Xinwen Hu, Lei Ma

Auditing and Robustifying COVID-19 Misinformation Datasets via Anticontent Sampling...................... 15260
Clay H. Yoo, Ashiqur R. KhudaBukhsh

User-Oriented Robust Reinforcement Learning 15269
Haoyi You, Beichen Yu, Haiming Jin, Zhaoxing Yang, Jiahui Sun

Safety Verification of Nonlinear Systems with Bayesian Neural Network Controllers............................... 15278
Xia Zeng, Zhengfeng Yang, Li Zhang, Xiaochao Tang, Zhenbing Zeng, Zhiming Liu

Reachability Analysis of Neural Network Control Systems 15287
Chi Zhang, Wenjie Ruan, Peipei Xu

BIFRNet: A Brain-Inspired Feature Restoration DNN for Partially Occluded Image Recognition 15296
Jiahong Zhang, Lihong Cao, Qiuxia Lai, Bingyao Li, Yunxiao Qin

Robustness to Spurious Correlations Improves Semantic Out-of-Distribution Detection........................... 15305
Lily H. Zhang, Rajesh Ranganath

Evaluating Model-Free Reinforcement Learning toward Safety-Critical Tasks.. 15313
Linrui Zhang, Qin Zhang, Li Shen, Bo Yuan, Xueqian Wang, Dacheng Tao

Video-Audio Domain Generalization via Confounder Disentanglement 15322
Shengyu Zhang, Xusheng Feng, Wenyan Fan, Wenjing Fang, Fuli Feng, Wei Ji, Shuo Li, Li Wang, Shanshan Zhao, Zhou Zhao, Tat-Seng Chua, Fei Wu

Rethinking Safe Control in the Presence of Self-Seeking Humans 15331
Zixuan Zhang, Maitham AL-Sunni, Haoming Jing, Hirokazu Shirado, Yorie Nakahira

Towards Safe AI: Sandboxing DNNs-Based Controllers in Stochastic Games... 15340
Bingzhuo Zhong, Hongpeng Cao, Majid Zamani, Marco Caccamo

Author Index