

APPENDIX I

LIST OF RESEARCH STUDIES ABOUT SOFTWARE FAIRNESS

Table-A I-1 Research Studies Analyzed in the Systematic Literature Mapping (1-42)

#	Studies Selected for the Systematic Literature Mapping
S1	'Ignorance and Prejudice' in software fairness
S2	Astraea: Grammar-Based Fairness Testing
S3	Bias in machine learning software: Why? how? what to do?
S4	Fairness in Design: A Framework for Facilitating Ethical Artificial Intelligence Designs
S5	FairMask: Better Fairness via Model-Based Rebalancing of Protected Attributes
S6	Fix Fairness, Don't Ruin Accuracy: Performance Aware Fairness Repair using AutoML
S7	Information-Theoretic Testing and Debugging of Fairness Defects in Deep Neural Networks
S8	Fairness Improvement with Multiple Protected Attributes: How Far Are We?
S9	An Empirical Study on Correlations between Deep Neural Network Fairness and Neuron Coverage Criteria
S10	Fair Enough: Searching for Sufficient Measures of Fairness
S11	Perfectly parallel fairness certification of neural networks
S12	RUNNER: Responsible UNfair NEuron Repair for Enhancing Deep Neural Network Fairness
S13	Automatic Fairness Testing of Machine Learning Models
S14	Fairway: A way to build fair ML software
S15	Fairify: Fairness Verification of Neural Networks
S16	MAAT: a novel ensemble approach to addressing fairness and performance bugs for machine learning software
S17	MBFair: a model-based verification methodology for detecting violations of individual fairness
S18	Metamorphic testing and certified mitigation of fairness violations in NLP models
S19	ReFAIR: Toward a Context-Aware Recommender for Fairness Requirements Engineering
S20	Fairness testing: Testing software for discrimination
S21	Training Data Debugging for the Fairness of Machine Learning Software
S22	Automated directed fairness testing
S23	Search-based fairness testing for regression-based machine learning systems
S24	Search-based Automatic Repair for Fairness and Accuracy in Decision-making Software
S25	The Impact of Data Preparation on the Fairness of Software Systems
S26	Improving Fairness in Speaker Recognition
S27	Identifying Imbalance Thresholds in Input Data to Achieve Desired Levels of Algorithmic Fairness
S28	A Maturity Model for Trustworthy AI Software Development
S29	Fairness-aware Configuration of Machine Learning Libraries
S30	Diversity-aware fairness testing of machine learning classifiers through hashing-based sampling
S31	Exploring and Repairing Gender Fairness Violations in Word Embedding-based Sentiment Analysis Model through Adversarial Patches
S32	Efficient white-box fairness testing through gradient search
S33	Coverage-Guided Fairness Testing
S34	Individual Fairness for Graph Neural Networks: A Ranking based Approach
S35	MAFT: Efficient Model-Agnostic Fairness Testing for Deep Neural Networks via Zero-Order Gradient Search
S36	A Comprehensive Empirical Study of Bias Mitigation Methods for Machine Learning Classifiers
S37	Supporting Many-Objective Software Requirements Decision: An Exploratory Study on the Next Release Problem
S38	Multi-objective search for gender-fair and semantically correct word embeddings
S39	Software doping analysis for human oversight
S40	FairBalance: How to Achieve Equalized Odds With Data Pre-processing
S41	The landscape and gaps in open source fairness toolkits
S42	Measuring Imbalance on Intersectional Protected Attributes and on Target Variable to Forecast Unfair Classifications

Table-A I-2 Research Studies Analyzed in the Systematic Literature Mapping (43-95)

#	Studies Selected for the Systematic Literature Mapping
S43	Fair-SSL: Building fair ML Software with less data
S44	Fairea: A model behaviour mutation approach to benchmarking bias mitigation methods
S45	White-box fairness testing through adversarial sampling
S46	Are My Deep Learning Systems Fair? An Empirical Study of Fixed-Seed Training
S47	Algorithmic bias and the value sensitive design approach
S48	FairnessLab: A Consequence-Sensitive Bias Audit and Mitigation Toolkit
S49	Balanced Fair K-Means Clustering
S50	Fair near neighbor search via sampling
S51	Testing machine learning algorithms for balanced data usage
S52	A responsible machine learning workflow with focus on interpretable models, post-hoc explanation, and discrimination testing
S53	Bias in human data: A feedback from social sciences
S54	Fairness without harm: Decoupled classifiers with preference guarantees
S55	AI Extension of SQuaRE Data Quality Model
S56	Data collection and quality challenges in deep learning: a data-centric AI perspective
S57	Detecting Discrimination Risk in Automated Decision-Making Systems with Balance Measures on Input Data
S58	QuoTe: Quality-oriented Testing for Deep Learning Systems
S59	An Algorithmic Framework for Bias Bounties
S60	Tiny, Always-on, and Fragile: Bias Propagation through Design Choices in On-device Machine Learning Workflows
S61	Experimental study on generating multi-modal explanations of black-box classifiers in terms of gray-box classifiers
S62	A Multidimensional Analysis of Social Biases in Vision Transformers
S63	De-biasing "bias" measurement
S64	Evaluating impact of race in facial recognition across machine learning and deep learning algorithms
S65	Evaluating impact of race in facial recognition across machine learning and deep learning algorithms
S66	Integrating Fairness in the Software Design Process: An Interview Study With HCI and ML Experts
S67	A qualitative exploration of perceptions of algorithmic fairness
S68	Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support
S69	Causal Intervention for Fairness in Multibehavior Recommendation
S70	A step toward building a unified framework for managing AI bias
S71	Software documentation is not enough! Requirements for the documentation of AI
S72	RELIANT: Fair Knowledge Distillation for Graph Neural Networks
S73	Certified Machine-Learning Models
S74	Abstract Interpretation-Based Feature Importance for Support Vector Machines
S75	Slice Tuner: A Selective Data Acquisition Framework for Accurate and Fair Machine Learning Models
S76	Proportionally fair clustering
S77	The what-if tool: Interactive probing of machine learning models
S78	Pros and cons of GAN evaluation measures: New developments
S79	Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry
S80	CO*IR: A Greedy and Individually Fair Re-ranker
S81	Towards model-based bias mitigation in machine learning
S82	Imbalanced data as risk factor of discriminating automated decisions: A measurement-based approach
S83	MLCHECK-property-driven testing of machine learning classifiers
S84	Multi-objective swarm intelligence approach for bias mitigation in decision-making software
S85	Experience: Bridging Data Measurement and Ethical Challenges with Extended Data Briefs
S86	FairPlay: A Collaborative Approach to Mitigate Bias in Datasets for Improved AI Fairness
S87	How fair are we? From conceptualization to automated assessment of fairness definitions
S88	Method for analysis and formation of representative text datasets
S89	Diversity Drives Fairness: Ensemble of Higher Order Mutants for Intersectional Fairness of Machine Learning Software
S90	Fairness Testing Through Extreme Value Theory
S91	Fairness Mediator: Neutralize Stereotype Associations to Mitigate Bias in Large Language Models
S92	Building Bridges, Not Walls: Fairness-Aware and Accurate Recommendation of Code Reviewers via LLM-Based Agents Collaboration
S93	FairPreprocessor: Better Fairness via Addressing Imbalanced Data through Synthetic Data Generation and Mitigating Biased Labels
S94	Beyond the Seeds: Fairness Testing via Counterfactual Analysis of Non-Seed Instances
S95	RSUTT: Robust Search Using T-Way Testing