Debugging Machine Learning for Fairness

Gary Tan, Professor of EECS, Penn State
Nov 30\textsuperscript{th}, 2023, Binghamton University
Collaborators

• UT El Paso: Saeid Tizpaz-Niari, Verya Monjezi
• Penn State: Vishnu Dasu, Ashish Kumar
• UC Boulder: Ashutosh Trivedi
Data driven software

Parole decision

Credit score

Is ML software fair?

Hiring
Fairness Defects

Racial Bias in Amazon
Same-day Delivery

Black Americans Face More Audit Scrutiny, IRS Acknowledges

Racial Bias in IRS Tax Audits
Debugging ML Software for Fairness

Q1: Identifying fairness issues

Q2: Isolating reasons of unfairness and providing mitigation
Fairness Notions

- A **protected feature** (e.g., age, sex, race)
  - A feature ML should not discriminate against
  - A **protected group** is a group of individuals with the same protected feature

- **Group fairness**
  - Similar statistics for different protected groups

- **Individual fairness**
  - Similar individuals receive similar treatments
Group Fairness Definitions

• **Equal Opportunity Difference (EOD)**
  • The difference between the true positive rates (TPR) of two protected groups
    \[
    EOD = |TPR(G1) - TPR(G2)|
    \]
  • E.g., hiring; G1 = the male group; G2 = the female group
  • When EOD=0%, both groups have about equal chances of being hired for qualified applicants
  • The smaller the EOD is, the more fair the system is

• **Average Odd Difference (AOD)**
  • Consider both the true positive and the false positive rates
    \[
    AOD = \frac{1}{2} (|TPR(G1) - TPR(G2)| + |FPR(G1) - FPR(G2)|)
    \]
Individual Fairness

- **Counterfactual**: a pair of individuals with the same non-protected features, but differ in one protected feature
Testing for Fairness

Fairness Testing: Testing Software for Discrimination
Sainyam Galhotra  Yuriy Brun  Alexandra Meliou

Black-box Fairness Testing of Machine Learning Models
Aniya Aggarwal  Pranay Lohia  Seema Nagar  Kuntal Dey  Diptikalyan Saha

Efficient White-Box Fairness Testing through Gradient Search
Lingfeng Zhang  Yueling Zhang*  Min Zhang*  East China Normal University  Singapore Management University  East China Normal University
Shanghai, China  Singapore, Singapore  Shanghai, China
lanford217@gmail.com  vlahang.ecnu@gmail.com  mzhang@sei.ecnu.edu.cn

Automated Directed Fairness Testing
Sakshi Udeshi  Pryanshu Arora  Sudipta Chattopadhyay

White-box Fairness Testing through Adversarial Sampling
Peixin Zhang  Jingyi Wang  Jun Sun

NeuronFair: Interpretable White-Box Fairness Testing through Biased Neuron Identification
Haibin Zheng  Zhiqing Chen  Tianyu Du  Zhejiang University
zhjzhang320@gmail.com  zhiqin@zju.edu.cn  hver2020@zju.edu.cn
Xuhong Zhang  Yao Cheng  Shouling Ji  zhangxuhong@zju.edu.cn  Huawei International Pte. Ltd.
cpeng0911@huawei.com  sj@ecs.ee.zju.edu.cn
Jingyi Wang  Yue Yu  Jinyin Chen  Zhejiang University
wangjinyi@zju.edu.cn  yueyu@nudt.edu.cn  zju.edu.cn

EIDIG
ISSTA’21

AEQUITAS
ASE’18

ADT
ICSE’20

ICSE’22
Parfait-ML: Fairness-Aware Configuration of Machine Learning Libraries
In 2022 International Conference on Software Engineering (ICSE)
Building ML Models

• Pick an ML library (TensorFlow, scikit-learn, ...) and an ML algorithm
• “Programming”: select hyperparameters
• Train an ML model
Parfait-ML: Do hyperparameters affect the fairness of resulting ML models?

Example Hyperparameters

- # of layers/neurons;
- max depth of trees;
- # of clusters;
Characterizing Hyperparameter-Fairness Relationship

• **Problem statement**
  • Given an ML algorithm, what hyperparameter valuations can lead to the most fair/unfair ML models?
  • Accuracy constraint: close to the accuracy of the model with a default hyperparameter configuration
  • Focus on group fairness

• **Methodology**: search over the hyperparameter space
  • Random search: randomly generate hyperparameters
  • Evolutionary search: black-box fuzzing; grey-box fuzzing
Evolutionary Search

1. Evolutionary Search
2. Fuzzing Driver
3. Parse Hyper-Parameter
4. Training Data Set
5. Validation Data Set
6. ML Model

Accuracy and fairness results; path coverage
Statistical Fault Localization

• **Problem statement**: explain the dependence of hyperparameters toward fairness

• **Methodology**
  • k clusters of hyperparameters in terms of their accuracy and fairness results
  • Build a decision tree from hyperparameters to clusters
Adult Census Income

<table>
<thead>
<tr>
<th>age</th>
<th>workclass</th>
<th>race</th>
<th>sex</th>
<th>capital-gain</th>
<th>hours-per-week</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>Private</td>
<td>White</td>
<td>Female</td>
<td>0</td>
<td>45</td>
<td>&gt;50K</td>
</tr>
<tr>
<td>24</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>40</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>45</td>
<td>State-gov</td>
<td>Black</td>
<td>Female</td>
<td>0</td>
<td>65</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>30</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>65</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>20</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>3781</td>
<td>50</td>
<td>&gt;50K</td>
</tr>
<tr>
<td>17</td>
<td>?</td>
<td>White</td>
<td>Male</td>
<td>8</td>
<td>8</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>39</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>40</td>
<td>&gt;50K</td>
</tr>
<tr>
<td>40</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>30</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>46</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>7298</td>
<td>8</td>
<td>&gt;50K</td>
</tr>
<tr>
<td>19</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>28</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>20</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>20</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>20</td>
<td>Private</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>63</td>
<td>?</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>63</td>
<td>?</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
<tr>
<td>27</td>
<td>Local-gov</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>50</td>
<td>&lt;=50K</td>
</tr>
</tbody>
</table>

Random Forest Ensemble

18 Hyperparameters
- \(n\_estimators \in \mathbb{Z}_{\geq 1}\)
- \(max\_depth \in \mathbb{Z}_{\geq 1}\)
- \(criterion \in \{\text{gini, entropy}\}\)
- \(max\_features \in \{\text{auto, log, ...}\}\)
- \(ccp\_alpha \in \mathbb{R}_{\geq 0}\)

- **max\_feature**: the maximum number of features during training
- Setting it to be \(\sqrt{\text{total\_features}}\) or \(\log(\text{total\_features})\) increases bias
Evaluation

Q1: Hyperparameter can aggravate or suppress present biases in the dataset.
Q2: Mutation-based Evolutionary algorithms are effective search technique.
Q3: Some hyperparameters systematically introduce biases.
Q4: Parfait-ML outperforms Exp. Gradients and Fairway mitigation techniques.

Parfait-ML and all experimental subjects are publicly available: https://github.com/Tizpaz/Parfait-ML
Q1: Can hyperparameters selection affect ML algorithm fairness?

Tuning of hyperparameters significantly affects fairness. Within 1% of accuracy margins, fairness metric values can range from 1% to 23%.
Q2: Can evolutionary search help in finding hyperparameter valuations with large bias?

Evolutionary search is effective in finding hyperparameter configurations that lead to large bias.
Q3: Can statistical fault localization be helpful?

We analyze 900 decision tree models, 180 models per algorithm:

- **Random Forest**: max_features (170) and min_weight_fraction_leaf (160)
- **Decision Tree**: min_fraction_leaf (114) and max_features (114)
- **Logistic Regression**: solver (175), tol (53), and fit-intercept (50)
- **Support Vector Machine**: Degree (53)
- **Discriminant Analysis**: tol (141)
**Q4:** How does our approach compare to the state-of-the-art in terms of mitigation?

*Parfait-ML is effective to improve fairness by finding low-bias configurations of hyperparameters (vis-a-vis state-of-the-art).*

<table>
<thead>
<tr>
<th>Alg.</th>
<th>Scenario</th>
<th>Time</th>
<th>FLASH* (%)</th>
<th>PARFAIT-ML (%)</th>
<th>EOD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>Census, Sex</td>
<td>0.0%</td>
<td>11.0% (+/- 1.0%)</td>
<td>4.3% (+/- 1.3%)</td>
<td>4.3% (+/- 1.3%)</td>
</tr>
<tr>
<td>LR</td>
<td>Census, Race</td>
<td>0.0%</td>
<td>0.8% (+/- 1.4%)</td>
<td>5.8% (+/- 1.4%)</td>
<td>5.8% (+/- 1.4%)</td>
</tr>
<tr>
<td>LR</td>
<td>Credit, Sex</td>
<td>0.0%</td>
<td>0.2% (+/- 0.0%)</td>
<td>0.4% (+/- 0.0%)</td>
<td>0.4% (+/- 0.0%)</td>
</tr>
<tr>
<td>LR</td>
<td>Bank, Age</td>
<td>0.0%</td>
<td>0.0% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
</tr>
<tr>
<td>LR</td>
<td>Compass, Sex</td>
<td>0.0%</td>
<td>0.0% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
</tr>
<tr>
<td>LR</td>
<td>Compass, Race</td>
<td>0.0%</td>
<td>0.0% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
</tr>
<tr>
<td>DT</td>
<td>Census, Sex</td>
<td>2.5</td>
<td>71.5% (+/- 1.3%)</td>
<td>71.1% (+/- 0.0%)</td>
<td>2.2% (+/- 0.5%)</td>
</tr>
<tr>
<td>DT</td>
<td>Census, Race</td>
<td>124.8</td>
<td>90.9% (+/- 0.1%)</td>
<td>88.9% (+/- 0.4%)</td>
<td>4.4% (+/- 1.6%)</td>
</tr>
<tr>
<td>DT</td>
<td>Credit, Sex</td>
<td>5.9</td>
<td>97.1% (+/- 0.0%)</td>
<td>97.1% (+/- 0.0%)</td>
<td>5.6% (+/- 0.0%)</td>
</tr>
<tr>
<td>DT</td>
<td>Bank, Age</td>
<td>11.4</td>
<td>97.1% (+/- 0.0%)</td>
<td>97.1% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
</tr>
<tr>
<td>DT</td>
<td>Compass, Sex</td>
<td>9.1</td>
<td>97.1% (+/- 0.0%)</td>
<td>97.1% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
</tr>
<tr>
<td>DT</td>
<td>Compass, Race</td>
<td>9.1</td>
<td>97.1% (+/- 0.0%)</td>
<td>97.1% (+/- 0.0%)</td>
<td>0.0% (+/- 0.0%)</td>
</tr>
</tbody>
</table>

* Joymally Chakraborty, Suvodeep Majumder, Zhe Yu, Tim Menzies:  
  **Fairway: a way to build fair ML software.** ESEC/SIGSOFT FSE 2020: 654-665
DICE: Information-Theoretic Debugging of Fairness Defects in DNNs

In 2023 International Conference on Software Engineering (ICSE)
Limitations of Individual Fairness Definitions

Limitation:
- Binary (yes/no)
- No Metric Characterizing Amounts of Discrimination
- Don’t Prioritize Test Cases
Quantitative Individual Discrimination (QID): an Information-Theoretic Approach

Extreme Cases:
- $K = 1$ (No Sensitivity)
- $K = |CF|$ (Strong Sensitivity)
Quantitative Individual Discrimination (QID): an Information-Theoretic Approach

Advantages:
+ Quantify sensitivity of DNN to the protected features.
+ Smooth Feedback during Search.
+ Enable Test-Case Prioritization.

Min entropy
\[ \log_2(K) \]

Shannon entropy
\[ \log_2(|CF|) - \sum_{j=1}^{k} \frac{|Pj|}{\#CF} \times \log_2(|Pj|) \]

K : # clusters
DICE Overview

Global phase
Goal: Find inputs that maximize the amount of discrimination

Local phase
Goal: Generate maximal number of discriminatory instances

Quantitative Individual Discrimination (QID)

Causal Effect of Neurons on QID

\[ E[QID|Do(Ne_3^3)] > 0, A \]

\[ E[QID|Do(Ne_3^3)] = 0, A \]
Debugging Phase: Layer Localizer

• Goal: detect a layer with the largest sensitivity to protected features

Distance between the outputs of layer i for an individual’s counterfactual sets

Rate of change: \[ R_i = \frac{\sigma_i - \max \sigma}{\max \sigma + \epsilon} \]

Debugging Phase: Neuron Localizer

• Goal: localize neurons in the layer that have significant positive or negative effects on fairness

\[ QID|Do(Ne_0^0 > 0) \]
\[ QID|Do(Ne_0^0 = 0) \]

\[ \text{Accuracy}_{\geq 0} \]
\[ \text{Accuracy}_{< 0} \]

\[ ACD_0 = QID > 0 - QID = 0 \]
\[ ACD_1 = QID > 0 - QID = 0 \]

<table>
<thead>
<tr>
<th>Neuron Index</th>
<th>Average Causal Differences (ACD)</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( ACD_0 )</td>
<td>✓</td>
</tr>
<tr>
<td>1</td>
<td>( ACD_1 )</td>
<td>✓</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( j )</td>
<td>( ACD_j )</td>
<td>✓</td>
</tr>
</tbody>
</table>

\( Ne + \) \hspace{2cm} \( Ne - \)
## Evaluation

### Datasets:
- Census
- Compas
- German Credit
- Default Credit
- Heart Health
- Bank Marketing
- Diabetes
- Student Performance
- MEPS 15
- MEPS 16

### State-of-the-art:
- AEQUITAS (ASE’18)
- ADF (ICSE’20)
- NEURONFAIR (ICSE’22)

**DICE** is an open-source tool publicly available in [https://github.com/armanunix/Fairness-testing](https://github.com/armanunix/Fairness-testing)
Effectiveness of QID-Based Search

Increased QID by 3.4× on average.
Enabled us to prioritize test cases efficiently: focus on 50 or less test cases.
Found discriminating instances that used up to 75% of protected information.

TABLE II: DICE characterizes QID for 10 datasets and DNNs in 1 hour run (results are the average of 10 runs).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$m$</th>
<th>$Q_I$</th>
<th>$K_I$</th>
<th>$#I$</th>
<th>$K_F$</th>
<th>$T_{K_F}$</th>
<th>$Q_\infty$</th>
<th>$Q_1$</th>
<th>$#I_{K_F}^1$</th>
<th>$#I_{K_F}^2$</th>
<th>$#I_{K_F}^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census</td>
<td>6</td>
<td>3.3</td>
<td>2.39</td>
<td>504,414</td>
<td>7.90</td>
<td>0.016</td>
<td>1.40</td>
<td>1.11</td>
<td>89.7</td>
<td>609.6</td>
<td>2,310.1</td>
</tr>
<tr>
<td>Compass</td>
<td>6</td>
<td>3.3</td>
<td>2.39</td>
<td>504,414</td>
<td>7.90</td>
<td>0.016</td>
<td>1.40</td>
<td>1.11</td>
<td>89.7</td>
<td>609.6</td>
<td>2,310.1</td>
</tr>
<tr>
<td>German</td>
<td>6</td>
<td>3.3</td>
<td>2.39</td>
<td>504,414</td>
<td>7.90</td>
<td>0.016</td>
<td>1.40</td>
<td>1.11</td>
<td>89.7</td>
<td>609.6</td>
<td>2,310.1</td>
</tr>
<tr>
<td>Default</td>
<td>6</td>
<td>3.3</td>
<td>2.39</td>
<td>504,414</td>
<td>7.90</td>
<td>0.016</td>
<td>1.40</td>
<td>1.11</td>
<td>89.7</td>
<td>609.6</td>
<td>2,310.1</td>
</tr>
<tr>
<td>Heart</td>
<td>6</td>
<td>3.3</td>
<td>2.39</td>
<td>504,414</td>
<td>7.90</td>
<td>0.016</td>
<td>1.40</td>
<td>1.11</td>
<td>89.7</td>
<td>609.6</td>
<td>2,310.1</td>
</tr>
<tr>
<td>Bank</td>
<td>6</td>
<td>3.3</td>
<td>2.39</td>
<td>504,414</td>
<td>7.90</td>
<td>0.016</td>
<td>1.40</td>
<td>1.11</td>
<td>89.7</td>
<td>609.6</td>
<td>2,310.1</td>
</tr>
<tr>
<td>Diabetes</td>
<td>10</td>
<td>5.2</td>
<td>7.03</td>
<td>19,673</td>
<td>18.52</td>
<td>31.62</td>
<td>2.61</td>
<td>1.62</td>
<td>2.6</td>
<td>3.5</td>
<td>6.0</td>
</tr>
<tr>
<td>Students</td>
<td>16</td>
<td>4</td>
<td>1.90</td>
<td>133,221</td>
<td>10.90</td>
<td>14</td>
<td>1.93</td>
<td>1.35</td>
<td>16.0</td>
<td>130.7</td>
<td>128.7</td>
</tr>
<tr>
<td>MEPS15</td>
<td>36</td>
<td>5.2</td>
<td>7.03</td>
<td>19,673</td>
<td>18.52</td>
<td>31.62</td>
<td>2.61</td>
<td>1.62</td>
<td>2.6</td>
<td>3.5</td>
<td>6.0</td>
</tr>
<tr>
<td>MEPS16</td>
<td>36</td>
<td>5.2</td>
<td>7.03</td>
<td>19,673</td>
<td>18.52</td>
<td>31.62</td>
<td>2.61</td>
<td>1.62</td>
<td>2.6</td>
<td>3.5</td>
<td>6.0</td>
</tr>
</tbody>
</table>
### Comparison with the state-of-the-art

- Found 20× more individual discrimination (ID) instances than the state-of-the-art.
Efficacy and efficiency of layer and neuron localization

- Activation mitigation can reduce QID discrimination by 5 to 64.3% (with a 2-3% loss of accuracy).
- Deactivation can improve the fairness by 6 to 27% (with 1-2% loss of accuracy).

<table>
<thead>
<tr>
<th></th>
<th>Students</th>
<th>MEPS15</th>
<th>MEPS16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>0.898</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>6.42</td>
<td>9.35</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>N/A</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td>5.79</td>
<td>N/A</td>
<td>6.98</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.866</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td>2.29</td>
<td>6.93</td>
<td>6.27</td>
</tr>
<tr>
<td></td>
<td>1,206</td>
<td>1,348</td>
<td>1,368</td>
</tr>
</tbody>
</table>
Ongoing Work: Neuron Repair for Fairness Mitigation
Observation

• A subset of neurons disparately contributes to unfairness
• Dropping them can result in fairer predictions
• Challenge: How do you identify an optimal subset in the exponential search space?
Solution

• Optimal subset is found using Simulated Annealing (SA)
• States in SA are a subset of neurons that are dropped
• The neighbors of a state are all states that differ in 1 neuron
• Cost Function: Unfairness of state + Penalty * (Initial unfairness if F1 of state is less than a threshold)
• Find state that minimizes cost function

\[ C(s) = EOD_s + p \cdot EOD_{s_0} \cdot 1(F1_s < F1_{s_0}) \]
Workflow Diagram

Original Unfair Model

Simulated Annealing Repair

1. Sample initial set of neurons $S_0$
2. Set best cost $C^* = \text{Cost}(S_0)$

Unfair Model $M$

Simulated Annealing Repair

1. Sample neighbor $S'$ of $S$
2. Compute cost $C = \text{Cost}(S)$, $C' = \text{Cost}(S')$

Validation Data

1. Set $S = S'$ if $C' \leq C$
2. Update best state $S^* = S$

Training Data

Test Data

Repaired Fair Model
Some Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Vanilla Network</th>
<th></th>
<th></th>
<th>NeuronRepair Network</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EOD</td>
<td>F1</td>
<td>Accuracy</td>
<td>EOD</td>
<td>F1</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Test</td>
<td>Train</td>
<td>Val</td>
<td>Test</td>
<td>Train</td>
</tr>
<tr>
<td>Adult</td>
<td>13.2%</td>
<td>13.18%</td>
<td>12.92%</td>
<td>0.684</td>
<td>0.653</td>
<td>0.667</td>
</tr>
<tr>
<td>Bank</td>
<td>10.2%</td>
<td>7.73%</td>
<td>15.81%</td>
<td>0.565</td>
<td>0.553</td>
<td>0.609</td>
</tr>
<tr>
<td>Default</td>
<td>10.62%</td>
<td>9.78%</td>
<td>9.51%</td>
<td>0.539</td>
<td>0.543</td>
<td>0.538</td>
</tr>
<tr>
<td>MEPS16</td>
<td>21.29%</td>
<td>18.69%</td>
<td>17.78%</td>
<td>0.542</td>
<td>0.535</td>
<td>0.547</td>
</tr>
</tbody>
</table>

- Across all datasets, the loss in F1 score is significantly lesser than the improvement in fairness
- Dropping neurons is a viable technique to improve fairness
Conclusions

• Non-functional properties such as fairness of ML software become increasingly important

• Parfait-ML and DICE
  • Applying software-engineering techniques, including fuzzing, fault localization, and repairs to improve ML fairness