ML-Fair Investigation

Fair Machine Learning Systems
An empirical investigation of the state of the practice

Tutor: Dott.ssa. Giulia Sellitto

https://github.com/CFerrara98/TesiFairness

"Fairness is giving all people the treatment they earn and deserve. It doesn't mean treating everyone alike."

About Me

CARMINE FERRARA

Master Student DI Unisa
Curriculum Software Engineering & IT Management

https://github.com/CFerrara98

Bachelor's degree in
Computer Science (2020)

Some interesting projects

IFY - Internship For You (Software Engineering 2020)

Digital Donation (Software Projects Management 2022)

Monumental Tree Finder (Cloud Computing 2022)

TABLE OF CONTENTS

01 Introduction

Fairness in a ML System

A new big problem for a ML Intensive Solutions

INPUT

ALGORITHM

ML Intensive Module

OUTPUT

DISCRIMINATION

DISCRIMINATION
Unequal treatment, unequal outcomes obtained by reinforcing patterns in predicting policing with feedback loops.

In 2015 an **Amazon Recruting Software** made discriminations based on Gender
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

In 2014 **COMPAS**, an American AI tool for crime recidivism judgment made discrimination against black people
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

The famous case of Google Translate & wrong gender translation
https://scroll.in/article/991275/google-translate-is-sexist-and-it-needs-a-little-gender-sensitivity-training

Some studies have demonstrated that Some airlines may be using algorithms to split up families during flights
https://www.vox.com/the-goods/2018/11/27/18115164/airline-flying-seat-assignment-ryanair

And many many other note cases...

BIAS as ML discrimination Cause

"When you think of A.I., it's forward-looking. But A.I. is based on data, and data is a reflection of our history." *- Coded Bias 2020*

Kinds of Biases

Selection bias Disproportionate weight introduced by the selection of data for analysis;

Disproportionate weight introduced by the suppression of information, e.g., when using surveys to collect data;   **Reporting bias**

Bias of an estimator Difference between the predictions made and the actual values;

Set of wrong assumptions that a learner uses to predict outputs.   **Learning bias**

ML solutions are designed and developed from Human, and Humans naturally reproduce their judgments in a ML Solutions
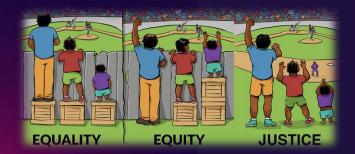
Fairness in a ML System

So, what is fairness?

In general...

> **FAIRNESS**
> Set of requirements, methods, and techniques to let an artificial intelligence solutions act "fairly".

A very **HOT topic...**

... that has started to attract the attention of researchers in **AI**, **Software Engineering** and **Law communities**

... with more than **twenty different notions & metrics** offairness proposed in the last few years.

https://fairware.cs.umass.edu/papers/Verma.pdf

Many levels of action to treat it

EQUALITY   EQUITY   JUSTICE

Fairness in a ML System

Define and generalize the concept of Fairness...

...for research community it's impossible, because any ML system is different from another

So it's possible **measure and monitor** fairness in **different way**

Probability Approaches

Similarity Approaches

Casual Reasoning Approaches

Statistical parity
Outcome test
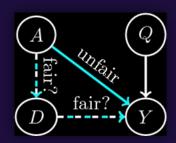
Casual discrimination
Fairness through unawareness

Counterfactual fairness

Fairness as Research Topic

Fair Data balancing & Feature Selection – M.Zhang e Mark Harmon
https://ieeexplore.ieee.org/document/9402057

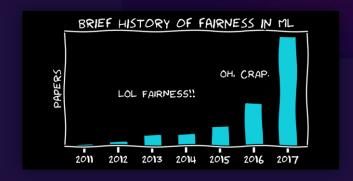*Fair Bias Identification – Chakraborty et al.*
https://arxiv.org/abs/2105.12195

Diversity Data Selection under Fairness Constrain – Moumoulidou at al.
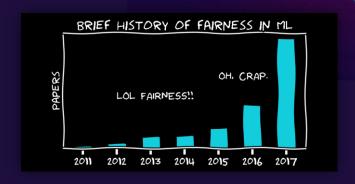*https://drops.dagstuhl.de/opus/volltexte/2021/13721/pdf/LIPIcs-ICDT-2021-13.pdf*

BRIEF HISTORY OF FAIRNESS IN ML

PAPERS

OH. CRAP.

LOL FAIRNESS!!

2011 2012 2013 2014 2015 2016 2017

From Artificial Intelligence point of view…

Fairness in a ML System

Fairness as Research Topic

Fairness as non Functional Aspect - Yuriy Brun e Alexandra Meliou
https://people.cs.umass.edu/~brun/pubs/pubs/Brun18fse-nier.pdf

Fairness Testing: Testing Software for Discrimination - Brun at al.
https://dl.acm.org/doi/10.1145/3106237.3106277

Fairness Testing: Testing Software for Discrimination- Finkelstein at al.
https://www.researchgate.net/publication/221222533_Fairness_Analysis_in_Requirements_Assignments

From Software Engineering point of view...

02 Research Questions

Working with fairness!

Research questions

As we seen, ML Fairness is grown up as Research topic in last years...

but Fairness is still an Hot topic to explore and systematize...

For instance, it's possible analyze how fairness is approached in a working context

Two main advantages

Try to suggest at researchers what are the strong and week points of fairness dealing in a working context...

Bring ML workers closer to fair-oriented development

Research questions

RQ1   *How fairness is approached in a AI-Intensive Working Context?*

The main research goal of this research question is provide an analytic overview about the actual work practices adopt by expert (like Data Scientists or Software Engineering) during ML-Intensive development.

To provide an exaustive overview about working state of practices, this main goal has been devided in 5 research sub goals that will drive the next empirical phase of the project.

Empirical Investigation

Research questions

RQ1.1 — *What are the best definition or approaches to deal with fairness in a working context?*

The main goal of this sub-research question is try to understand what are the kind of metrics and approaches that worker mostly adopt to develope fair-ML Solutions

Probability Definitios

Similarity Definitions

Casual Reasoning Definitions

Equity

Sensitive Features

Minority Groups

Research questions

RQ1.2 — *How is generally composed an a working team for the development of ML-Intensive Fair Critical System?*

Data Scientists

Software Engineers

Project Managers

Analysts and Architects

Fairness Expert

The main goal of this sub-research question is try to what kind of professional roles are critical to preserve fairness in a ML Fair Critical Development Process.

Understand who is the «accountable» of fairness during the development of a ML Intensive Solution, could be helpfull for research community in order to define specific Management standard for this kind of projects

Research questions

RQ1.3 *How much is fairness important compared to other Non-functional aspect ?*

Research community, in particolar Software Engineering community, define Fairness as a new kind of software Non Functional Requirement

To deal fairness in an appropriate manner in a ML System, they specify that is necessary treat fairness like a first class NFR like Security, Accuracy...

But how do workers consider important fairness respect to other commons requirement?

Research questions

RQ1.4 *In witch phases of a tipical ML Pipeline is important adopt strategy to preserve fairness in a ML Solution?*

MlOps and ML Pipelines are two of most common development approaches to realize modern ML Intensive Solutions

An ML pipeline is also the best way to describe a tipical life cycle of an ML Module...

... And some tecniques could be taken to improve solution quality during pipeline phases

Is it also valid for fairness,, if yes, in witch phases is usefull adopt strategies to improve fairness level of the ML Solution?

- Development Process
- Operational Process

Requirement Elicitation & Analysis → Data Retrieval

Dataset

ML Algorithm

Data Preparation: Data Processing & Wrangling, Feature Extraction & Engineering, Feature Scaling & Selection, Data Balancing

Build Model

Model Training & Testing

Model Validation & Verification

Model Deployment & Monitoring

Analysis & Statistics

Feedback Reporting

Research questions

RQ1.5 *How mutch are companies "mature" to treat fairness as a non functional Requirement?*

Finally, to complete the overview of fairness working context, it could be useful evaluate the level of maturity of companies in fairness treatment

Software engineering state of art, suggests the CMM – Capability Maturity Model to evaluate the degree of "maturity" of a company

Is it possible specialize CMM, to evaluate the level of Fairness treatment in a Company?

If yes, what is the most frequent maturity level for companies?

LEVEL 5: OPTIMIZATION
- Defect Prevention
- Test Process Optimization
- Quality Control

LEVEL 4: MEASURED
- Test Measurement
- Product Quality Evaluation
- Advanced Reviews

LEVEL 3: DEFINED
- Test Organization
- Test Training Program
- Test Lifecycle and Integration
- Non-Functional Testing
- Peer Reviews

LEVEL 2: MANAGED
- Test Policy and Strategy
- Test Planning
- Test Monitoring and Control
- Test Design and Execution
- Test Environment

LEVEL 1: INITIAL

03 Research Metodology

The Survey

The survey

To collect data and extract informations in order to answare at the research goals, it's obviusly necessary involve some expert of the context...

In general Software Engineerning Lecterature provides some methods to conduct this kind of investigations!
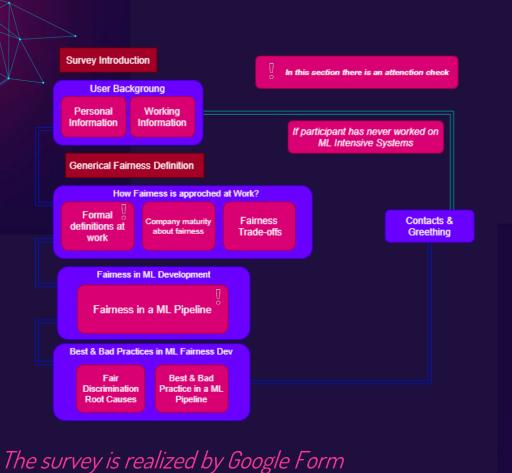
Interviews

Surveys

Focus Groups

Why have we (as project team) chosen the Survey?

It's one of the most common and used method to collect data for a mirate kind of population!

Using the correct dissemination platform, we can involve peoples with different cultures and backgrounds

A well Structured Survey permit to obtain omogeneous and mirate data that easier to analyze

Before starting with Survey Dissemination we have conducted a little pilot test with SE4AI students!

Pilot Testing Participants Requirements

Bachelor Degree
Knowdlage of SE4AI and MLOps
Introduction Overview of Fairness and
Bias in a ML System

Result & Changes

Estimation of Survey Compilation Time
Under 15 minutes

Demographic Questions
Obbligatory -> Optional

Content Remodulations
Some redundant question has been deleted!

**PILOT
TESTING**

A key choice for survey diffusion and the answare gathering is the platform to disseminate the survey!

For the dissemination we have chosen **Prolific**!

Prolific has given us the possibility to apply some restrictions on sampling recruiting

For Instance...

Fluent Knowdledge of English
Technical working sector – Computer Science, Technology, Engineering
Completation of an high study level – Diploma or Higher

DISCLAIMER – Necessity of knowdlage in Data Science or Software Engineering

Prolific

It's important to remark that

All informations has been treated exlusivelly to Investigation Goals

Participants aren't obblied to insert Sensitive Informations
(Optional Questions or Prefer Not To Say answare)

If participants were subject to work restrictions on informations provided,
We have invited them to avoid the submission

Survey Dissemination
ETA – 21 Days: from 12/05/2022 to 03/06/2022
Survey Close-Out: 16/05/2022
Total Submission Collected: 203

Prolific

04

Data Analysis

The Survey

Befor starting to analyze data

We have analyzed the integrity and consistency of the submissions recived

In particular, we have rejected:

17 Submissions because the participants provided **unlogic answare to Attenction checks**

2 Submissions because the relative partecipants not provide us **Prolific ID for Payment**

68 Submissions because the relative partecipants hasn't had **experience in ML Develompent**

After Data Cleaning we have obtained **116 answare util** to perform Data Analysis & Result Formalizations

Research Goals & Survey Questions Mapping

For Instance...

RQ1.1 - What are the best definition or approaches to deal with fairness in a working context?

Survey Questions

In your opinion, how much do the following aspects represent the generic definition of software fairness given above?

Considering your working experiences, how much do you apply the following approaches?

Do you use other approaches to deal with software fairness?

Data Preprocessing

Fairness Definitions Mapping

RQ1.1 – What are the best definition or approaches to deal with fairness in a working context?

Definitions Group	Explainations
General	Software fairness is impartial and just treatment or behaviour without favouritism or discrimination assumed by software systems. Software fairness has started to attract the attention of researchers in artificial intelligence, software engineering and law communities, with more than twenty different notions of fairness proposed in the last few years. Yet, there is no clear agreement on which definition to apply in each situation.
Probability Definitions	Treating similar individuals in a way that they are equally likely to receive a specific outcome
Similarity Definitions	Do not favor certain subjects over others on the basis of sensitive attributes, e.g., race, gender, etc.
Casual Reasoning Definitions	Taking decisions by protecting individuals and groups from mistreatments

Data Preprocessing

Qualitative Scale to Quantitative Scale Transformations

RQ1.3 – How much is fairness important compared to other Non-functional aspect ?

Level of Importance	Meaning
-2	Less important than fairness
-1	A bit less important than fairness
0	Neutral
1	A bit more important than fairness
2	More important than fairness

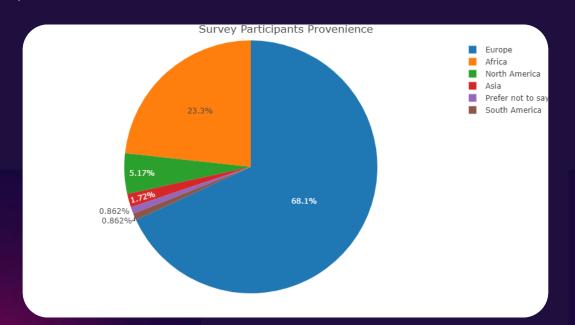RQ1.4 – In witch phases of a tipical ML Pipeline is important adopt strategy to preserve fairness in a ML Solution?

Level of Relevance	Meaning
1	Not at all
2	Slightly
3	Neutral
4	Very
5	Extremelly

Data Preprocessing
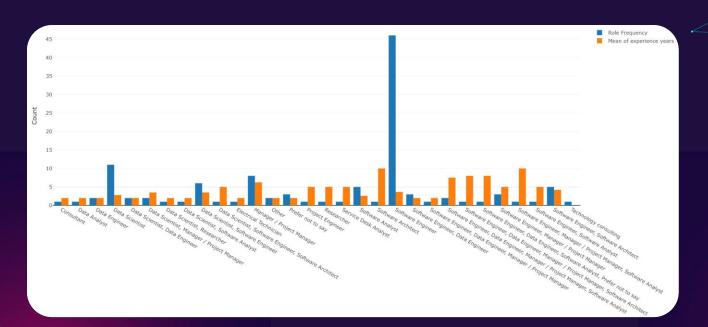
Participants Provenience Distribution

Survey Participants Provenience

- Europe
- Africa
- North America
- Asia
- Prefer not to say
- South America

23.3%

5.17%

1.72%

0.862%

0.862%
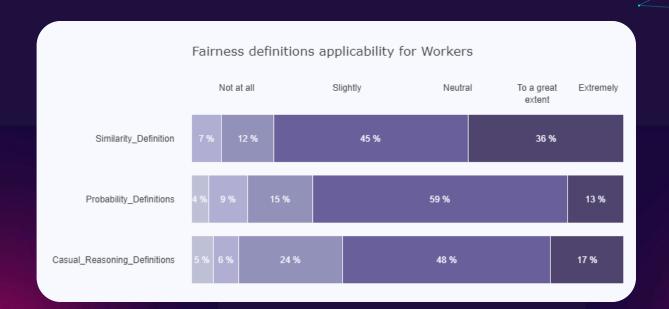
68.1%

Participants Professional Roles Distribution

Data Analysis
Research Question 1.1

Fairness Definitions in a Working Environment

RQ1.1 What are the best definition or approaches to deal with fairness in a working context?

Fairness definitions applicability for Workers

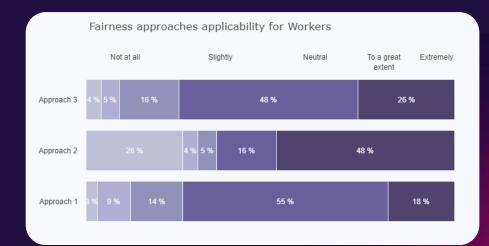|	Not at all	Slightly	Neutral	To a great extent	Extremely
Similarity_Definition	7 %	12 %	45 %		36 %
Probability_Definitions	4 %	9 %	15 %	59 %	13 %
Casual_Reasoning_Definitions	5 %	6 %	24 %	48 %	17 %

Fairness Approaches in a Working Environment

RQ1.1 What are the best definition or approaches to deal with fairness in a working context?

Approaches
1. We focus on guaranteeing high probability to obtain ethically correct outcomes regardless of sensitive features
2. We focus on guaranteeing that machine learning predictions are not going to discriminate by sensitive features
3. We model the relation between attributes and outcomes, verifying that the outcome does not depend on sensitive attributes

Fairness approaches applicability for Workers

	Not at all	Slightly	Neutral	To a great extent	Extremely
Approach 3	4 % / 5 %	16 %	48 %	26 %	
Approach 2	26 %	4 % / 5 %	16 %	48 %	
Approach 1	3 % / 9 %	14 %	55 %	18 %	

Data Analysis & Results

RQ1.1 What are the best definition or approaches to deal with fairness in a working context?

Do you use other approaches to deal with software fairness?

Domain Specific Approaches - *Tring to make the outcome available to people with disabilities so we at least can make it equal in that end*

Empirical Metodologies – *Yes, by conducting surveys and taking in consideration of other's opinions.*

Data Management Improvement – *We procure that data training is sensible enough that the results are the less biased possible.*
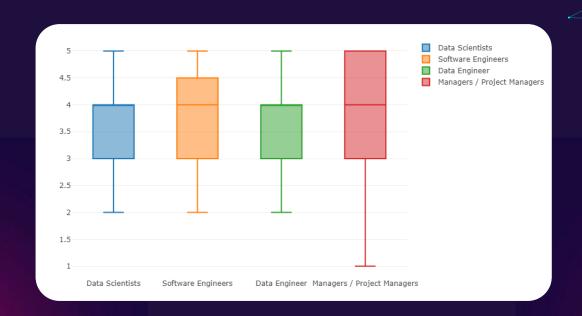
Correlation Analysis – *Analise strong-week correlactions between outcomes and features*

... And something else

Fairness Critical Roles in a development team

RQ1.2 How is generally composed an a working team for the development of ML–Intensive Fair Critical System?

Professional Role Impact box-plot 1

Data Analysis & Results

Fairness Critical Roles in a development team

RQ1.2 How is generally composed an a working team for the development of ML-Intensive Fair Critical System?
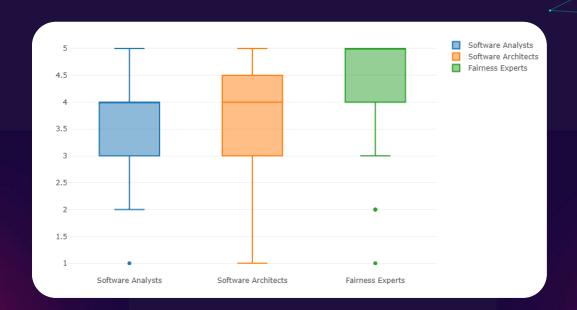
Professional Role Impact box-plot 2

Data Analysis & Results

Data Analysis
Research Question 1.3

RQ1.3 How much is fairness important compared to other Non-functional aspect ?

Performance vs. Fairness in a machine learning system

Legend
- Less important than fairness
- A bit less important than fairness
- Neutral
- A bit more important than fairness
- More important than fairness

Security and Safety VS Fairness

RQ1.3 How much is fairness important compared to other Non-functional aspect ?

Security and Safety vs. Fairness in a machine learning system

Degree of Importance

Individuals

Fairness
Security And Safety

Legend
- Less important than fairness
- A bit less important than fairness
- Neutral
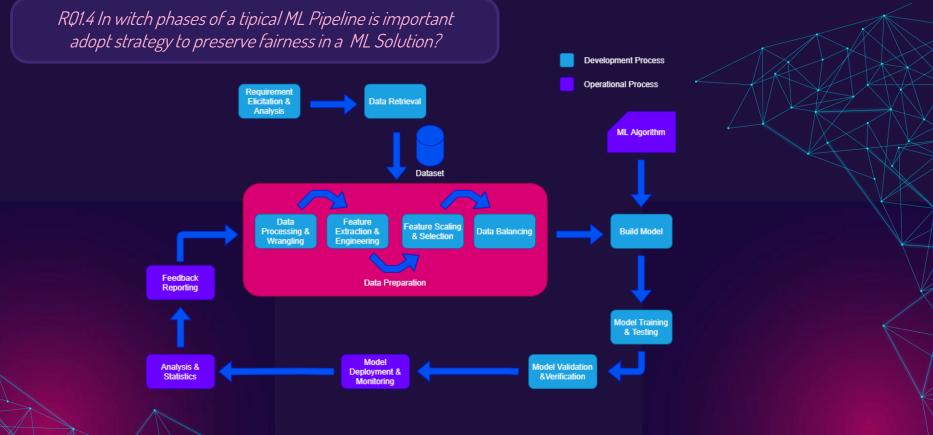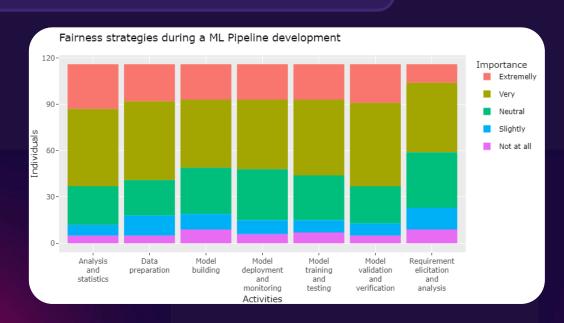- A bit more important than fairness
- More important than fairness

Data Analysis & Results

Fair Strategy Applicability in a Machine Learning Pipeline

RQ1.4 In witch phases of a tipical ML Pipeline is important adopt strategy to preserve fairness in a ML Solution?

Development Process

Operational Process

Requirement Elicitation & Analysis

Data Retrieval

ML Algorithm

Dataset

Data Processing & Wrangling

Feature Extraction & Engineering

Feature Scaling & Selection

Data Balancing

Data Preparation

Build Model

Feedback Reporting

Model Training & Testing

Analysis & Statistics

Model Deployment & Monitoring

Model Validation & Verification

Data Analysis & Results

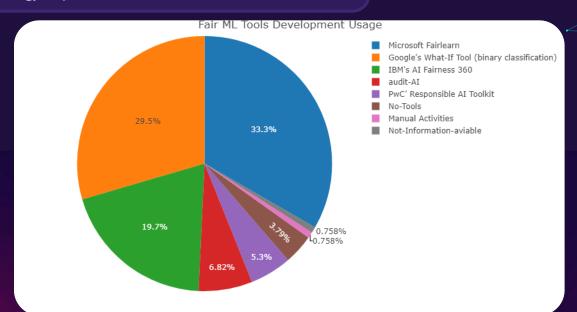# Fair Strategy Applicability in a Machine Learning Pipeline

RQ1.4 In witch phases of a tipical ML Pipeline is important adopt strategy to preserve fairness in a ML Solution?

Fairness strategies during a ML Pipeline development

Importance
- Extremelly
- Very
- Neutral
- Slightly
- Not at all

Individuals

Activities: Analysis and statistics, Data preparation, Model building, Model deployment and monitoring, Model training and testing, Model validation and verification, Requirement elicitation and analysis

Fairness Strategies Applicability in a ML-Pipeline Histogram

Data Analysis & Results

Fair Tools used during a Machine Learning Pipeline Development

RQ1.4 In witch phases of a tipical ML Pipeline is important adopt strategy to preserve fairness in a ML Solution?

Fair ML Tools Development Usage

- Microsoft Fairlearn
- Google's What-If Tool (binary classification)
- IBM's AI Fairness 360
- audit-AI
- PwC' Responsible AI Toolkit
- No-Tools
- Manual Activities
- Not-Information-aviable

33.3%
29.5%
19.7%
6.82%
5.3%
3.79%
0.758%
0.758%

Fairness Improoving Tools Applicabilty Pie Chart

Data Analysis & Results
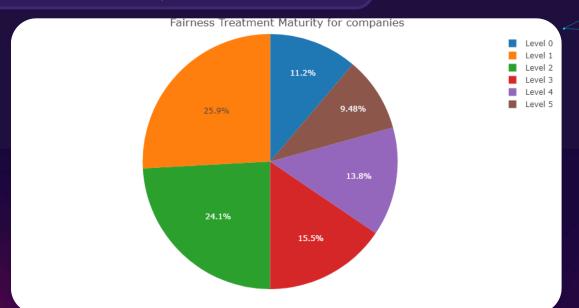
Data Analysis
Research Question 1.5

Fair Maturity Model Levels

RQ1.5 How mutch are companies "mature" to treat fairness as a non functional Requirement?

Level 0 We do not treat software fairness

Level 1 We occasionally treat software fairness, but related processes are disorganized and even chaotic

Level 2 We regularly treat fairness and related processes are established, defined and documented

Level 3 We regularly treat fairness and it develops its own standard fairness management processes

Level 4 We regularly treat fairness and it monitors and controls its own fairness related processes through data collection and analysis

Level 5 We regularly treat fairness and fairness related processes are constantly improved through monitoring feedback
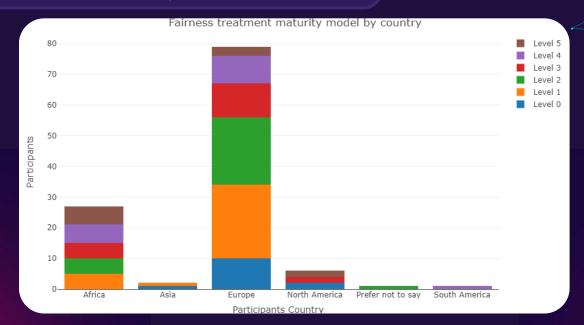
Data Analysis & Results

Country Fairness Maturity Distribution

RQ1.5 How mutch are companies "mature" to treat fairness as a non functional Requirement?

Fairness Maturity Model Bar Diagram

Data Analysis & Results

Parallel Studies

Unfair Discriminations Root Causes & Fair Machine Learning Development Bad and Best Practices
Francesco Abate

Possible Future Studies

Fair Team Management Protocol Definition

Explore and develop strategies for Fair Data Preparation

Define Strategies and Method for Fair Model Validation

Go Deeper in Fair Trade-off Analysis

Reproduce the study with other Empirical Methods, like Interviews

THANKS!

Do you have any questions?
c.ferrara49@studenti.unisa.com

Carmineferrara.altervista.com