# ML-Fair Investigation

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

https://github.com/CFerrara98/TesiFairness

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

# Advisors



Prof. Fabio Palomba Assistant Professor DI Unisa Sesa Lab



Dott.ssa. Giulia Sellitto PHD Student - DI Unisa Sesa Lab

# **OUR TEAM**



CARMINE FERRARA

Master Student DI Unisa https://github.com/CFerrara98

External



FRANCESCO ABATE

Master Student DI Unisa https://github.com/frekkanzer2

## TABLE OF CONTENTS

Introduction
Fairness in a ML System

03

Research Metodology
The Survey

Research Questions
Working with fairness

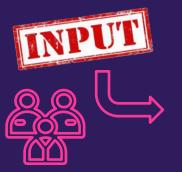
04

Data Analysis
Interpretation and consideration

# 01 Introduction

Fairness in a ML System









#### DISCRIMINATION

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

# Fairness in a ML System Some famous examples



In 2015 an **Amazon Recruting Software** made discriminations based on Gender

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





# Fairness in a ML System Some famous examples

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 output

Learning bias

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



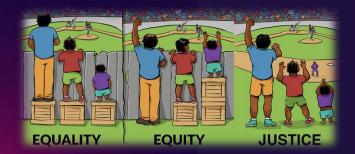
So, what is fairness?

In general..

#### **FAIRNESS**

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

#### Many levels of action to treat it



#### A very **HOT topic...**

... that has started to attract the attention of researchers in Al, 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



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

Statistical parity
Outcome test



Similarity Approaches

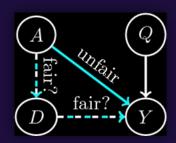
Casual discrimination

Fairness through unawareness



Casual Reasoning Approaches

Counterfactual fairness



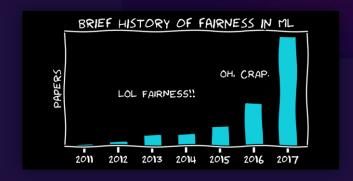


### Fairness as Research Topic



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/LIPlcs-ICDT-2021-13.pdf



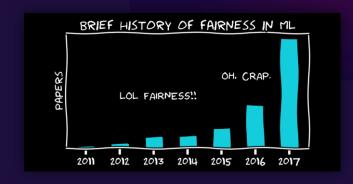
From Artificial Intelligence point of view...



### 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. <a href="https://www.researchgate.net/publication/221222533\_Fairness\_Analysis\_in\_Requirements\_Assignments">https://www.researchgate.net/publication/221222533\_Fairness\_Analysis\_in\_Requirements\_Assignments</a>

From Software Engineering point of view...

# **O2** Research Questions

Working with fairness!



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

**RQ1** 

How fairness is approached in a Al-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

**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



**RQ1.2** 

How is generally composed an a working team for the development of ML-Intensive Fair Critical System?

Data Scientists

Project Managers

Fairness Expert

Software Engineers

Analysts and Architects

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



**RQ1.3** 

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

Research community, in particular 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?



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

MIOps 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?

**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** 

# **93** 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

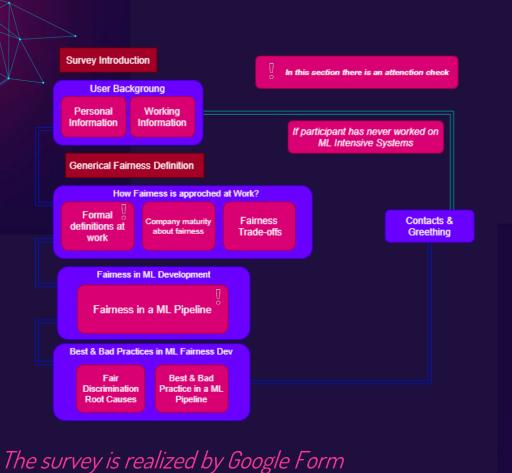
Focus Groups

#### 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



# Fairness in a ML System Survey Structure

#### Notes

Survey Structure also includes a reference to Best & Bad Practices Investigation – Francesco Abate

In some section we have included some Attenction checks in order to easily discover invalid Submissions

Survey details

# Fairness in a ML System Survey Validation

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!





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





#### It's important to remark that

All informations has been treated exlusively 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



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 Development

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?

### 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

### Fairness Approaches Mapping

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

Definitions Group	Explainations
Approach1	We focus on guaranteeing high probability to obtain ethically correct outcomes regardless of sensitive features
Approach2	We focus on guaranteeing that machine learning predictions are not going to discriminate by sensitive features
Approach3	We model the relation between attributes and outcomes, verifying that the outcome does not depend on sensitive attributes

### Fairness Company Maturity Levels Mapping

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

Level of Maturity	Meaning
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

#### 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

### Qualitative Scale to Quantitative Scale Transformations

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

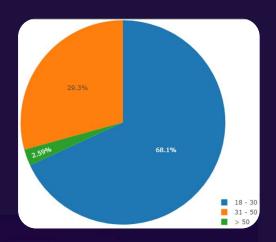
Level of Impact	Meaning
1	1 - Very low impact
2	2- Below average impact
3	3 - Average impact
4	4 - Above average impact
5	5 - Very high impact



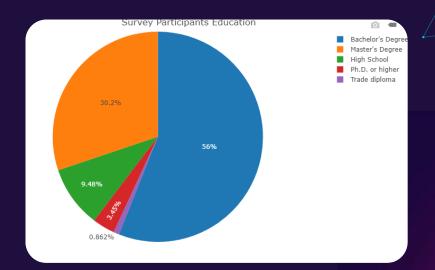
# Data Analysis

Participants Background

# Background Characteristics



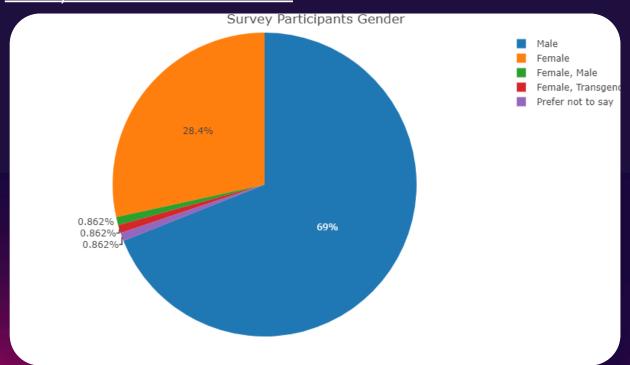
Participants Age Distribution



Participants Educational Distribution

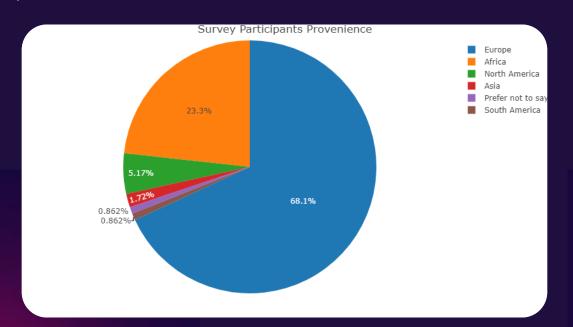
# Background Characteristics

# Participants Gender Distribution



# Background Characteristics

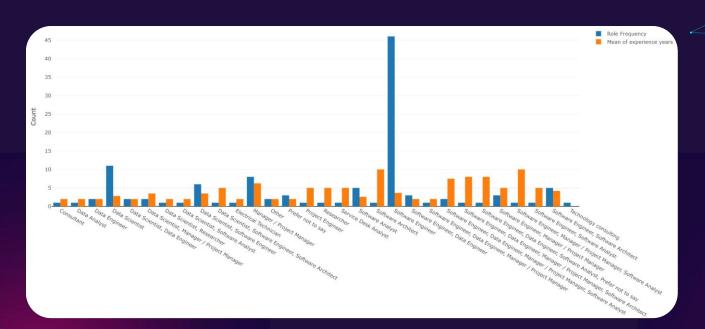
# Participants Provenience Distribution





# Research Question 1.1

# Participants Professional Roles Distribution



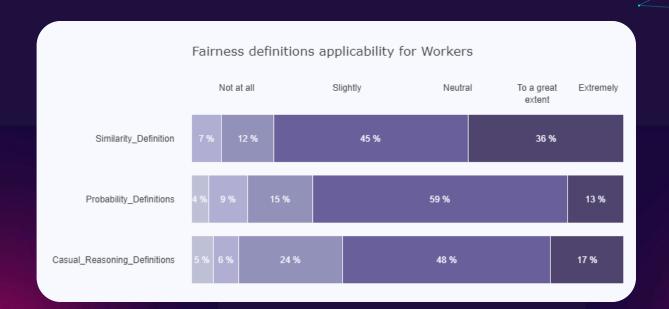


# Data Analysis

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 Approaches in a Working Environment

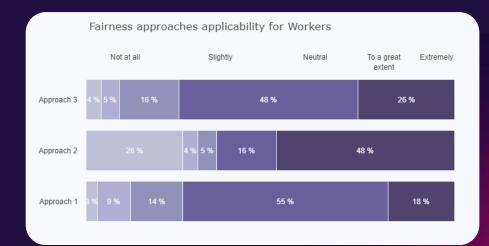
RQ1.1 What are the best definition or approaches to deal with fairness in a working context?

#### 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



### Other relevant thips

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

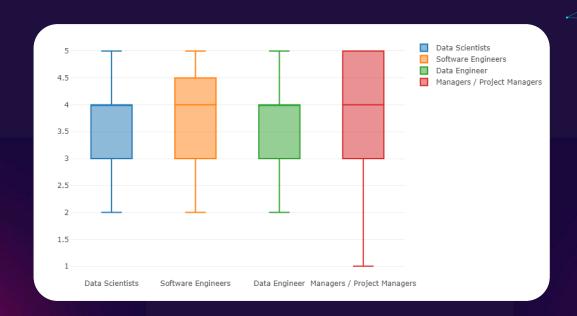
... And something else

# Data Analysis

Research Question 1.2

## 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?

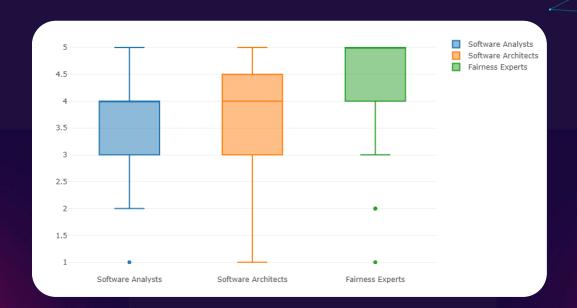


Professional Role Impact box-plot 1

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

# Usability VS Fairness

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



#### Legend

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

# Reliability VS Fairness

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



#### Legend

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

# Performance VS Fairness

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



#### Legend

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

# Supportability VS Fairness

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



#### Legend

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

# Accuracy VS Fairness

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



#### Legend

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

# Security and Safety VS Fairness

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



#### Legend

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

## Maintenability and Retraining VS Fairness

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

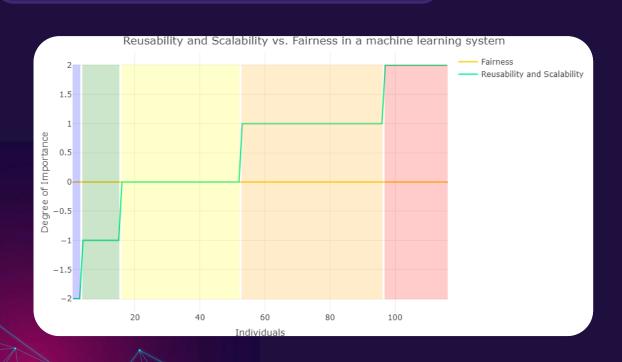


#### Legend

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

# Reusability and Scalability VS Fairness

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



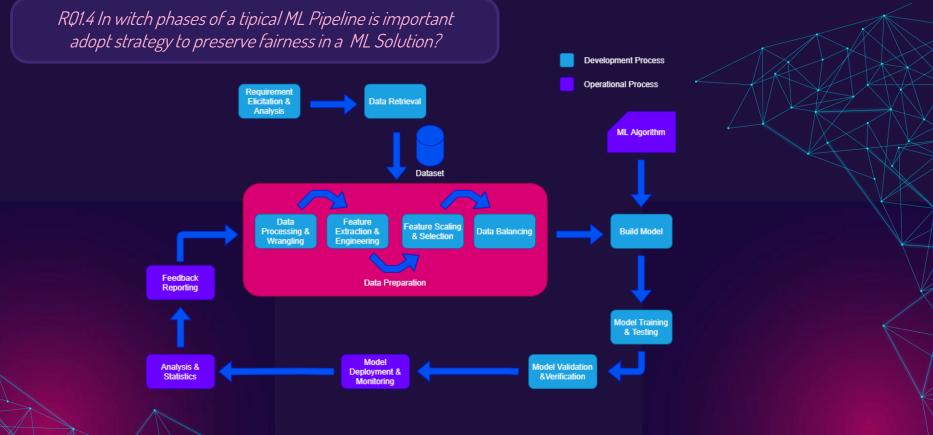
#### Legend

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

# Data Analysis

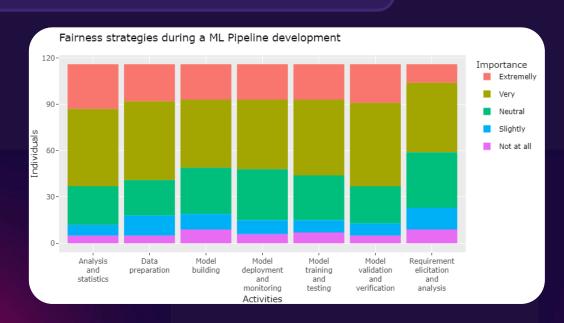
Research Question 1.4

# Fair Strategy Applicability in a Machine Learning Pipeline



# Fair Strategy Applicability in a Machine Learning Pipeline

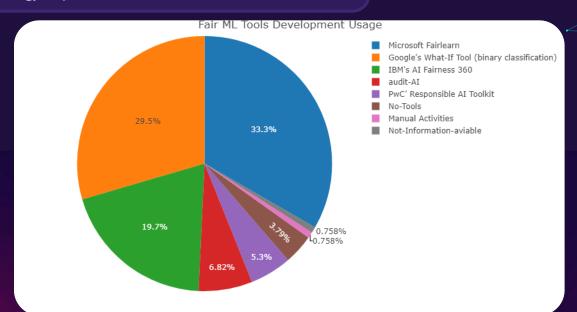
RQ1.4 In witch phases of a tipical ML Pipeline is important adopt strategy to preserve fairness in a ML Solution?





# 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?



Fairness Improoving Tools Applicabilty Pie Chart

# Data Analysis

Research Question 1.5

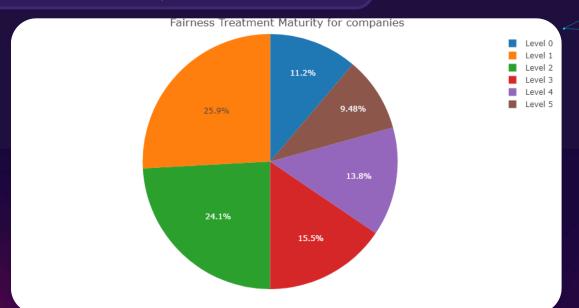
# Fair Maturity Model Levels

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
	We regularly treat fairness and related processes are established,
Level 2	defined and documented
	We regularly treat fairness and it develops its own standard fairness
Level 3	management processes
	We regularly treat fairness and it monitors and controls its own
Level 4	fairness related processes through data collection and analysis
	We regularly treat fairness and fairness related processes are
Level 5	constantly improved through monitoring feedback

# General Fairness Maturity Distribution

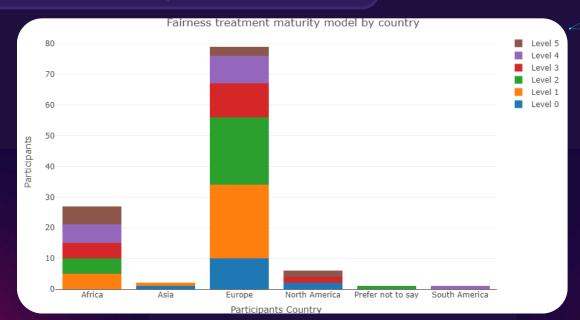
RQ1.5 How mutch are companies "mature" to treat fairness as a non functional Requirement?



Fairness Maturity Model Pie Chart

# Country Fairness Maturity Distribution

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



Fairness Maturity Model Bar Diagram



#### 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

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**