

**POLITECNICO DI MILANO**  
**Master of Science in Computer Science and Engineering**  
**Dipartimento di Elettronica, Informazione e Bioingegneria**



**[Title of the Thesis]**

**Internal supervisor: Prof. Viola Schiaffonati**  
**Internal co-supervisor: Prof. Letizia Tanca**  
**External supervisor: Prof. Pierre Senellart**  
**External co-supervisor: Prof. Karine Gentelet**

**M.Sc. Thesis by:**  
**Riccardo Corona, 927975**

**Academic Year 2020-2021**

**About this template**

With this template I want to give you some input on how to structure your thesis if you develop your thesis with me in Politecnico di Milano. Next to the pure structure, which you should reuse and adapt to your own needs, the document also contains instructions on how to approach the different sections, the writing and, sometimes, even the work on your thesis project itself. Sometimes you will also find boxes like this one. These are meant to provide you with explanations and insights or hints that go beyond the mere structure of a thesis.

I hope this template will help you do the best thesis ever, if not in the World, at least in your life.

Florian Daniel  
October 12, 2017

*Disclaimer:* Sometimes I may make statements that are general, if not over-generalized, personal considerations, or give hints on how to do work or research. Be aware that these are just my own opinions and by no way represent official statements by Politecnico di Milano or its community of professors. If something goes wrong with your thesis or presentation, you cannot refer to these statements as a defense. You are the final responsible of what goes into your thesis and what not.

*Acknowledgements:* The original template for this document was not created by me. I would love to acknowledge the real creator, but I actually do not know who it is. The template has been passed on to me by a former student, who also didn't know the exact origin of it. It was circulating among students. However, to the best of my knowledge at the time of writing, it seems that Marco D. Santambrogio and Matto Matteucci may have contributed at some point with considerations on structure and funny citations. Both were helpful and enjoyable when preparing this version of the template. I will be glad to add more precise acknowledgements if properly informed about the origins of this template.

**Supervisors and co-supervisors**

If the supervisor is internal to Politecnico di Milano (a professor or researcher), then on the first page use "Supervisor" plus the titles "Prof." and "Dr." for professors and researches, respectively. If the work was co-supervised by someone else, refer to him/her as the "Co-supervisor." If the work was supervised by someone external to Politecnico di Milano, use "External supervisor" for the external supervisor plus "Internal supervisor" for the internal supervisor that mandatorily must co-supervise the work with the external supervisor.

*Optionally, here goes the dedication.*



# Abstract

The abstract is a small summary of the thesis. It tells the reader in few words (up to one/one and a half page of total text) everything he/she needs to understand:

- ☐ the *context* of the work (e.g., chatbots),
- ☐ the specific *problem* approached by the thesis (e.g., the development of personal bots by non-programmers),
- ☐ if applicable, clearly state the *research questions* you would like to answer (e.g., “is it possible to enable non-programmers to do X using A?”),
- ☐ the three/four *core aspects of the proposed solution* (e.g., use pre-defined rules, use machine learning, assisted development, etc.),
- ☐ the *concrete outputs* produced by the thesis (e.g., a state of the art analysis, a conceptual/mathematical model, an application, middleware or API, an empirical study with/without users, etc.), and
- ☐ the *findings and conclusions* that one can draw from the evaluation of the approach (e.g., that under some very specific conditions non-programmers are indeed able to implement own chatbots effectively using the proposed technique).

## Checklists

Now and there I propose checklists with items, such as the one just above this box. They are meant for you to check if you included all the content that is relevant and that should be included, in order to make your text complete. When reading your thesis, I will look for all these items.

**Writing style**

This is a M.Sc. thesis. It's neither Facebook nor Twitter nor an email. This is going to be an official document with legal value that will decide on the final mark of your yearlong university career and perhaps even on your future work perspectives. So, you surely don't want to be judged badly because of grammar errors, flawed/wrong vocabulary or superficial layout and/or text structure. It is a must that what you write is always *correct* content- and language-wise (no false statements or claims, no language mistakes), *readable* (no sentences that cannot be understood) and targeted at the *average-skilled reader* (professors, but also your own colleagues).

**Plagiarism**

This is a M.Sc. thesis. It's neither Facebook nor Twitter nor an email. This is going to be an official document with legal value that will decide on the final mark of your yearlong university career and perhaps even on your future work perspectives – yes, I plagiarized myself here a little bit. So, you surely don't want to copy/paste material from scientific articles, online resources, books, and similar without adequately acknowledging the holders of the respective intellectual property rights. If you do so, it is a must that you properly *cite* each source where you take text or inspiration from. It is fine to do so – actually, citing someone is a compliment! – but it becomes a crime if the source is not cited. Not only M.Sc. titles but also Ph.D. titles have been withdrawn for fraudulent “reuse” of others' intellectual property. Be aware that Politecnico di Milano, like most higher educational institutions that issue university degrees or scientific publishers, may use specialized software to automatically detect plagiarism.

# Sommario

Here goes the translation into Italian of the abstract. If the thesis is written in Italian, no translation into English is needed. Hence, one of the following must be checked:

- ☐ Thesis written in *English*, properly proofread translation needed
- ☐ Thesis written in *Italian*, no translation needed, chapter omitted





# Acknowledgements

If you would like to thank somebody for given support, this is the right place to do so.



# Contents

<b>Abstract</b>	<b>I</b>
<b>Sommario</b>	<b>III</b>
<b>Acknowledgements</b>	<b>V</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Context: [topic] . . . . .	1
1.2 Scenario and Problem Statement . . . . .	3
1.3 Methodology . . . . .	4
1.4 Contributions . . . . .	5
1.5 Structure of Thesis . . . . .	7
<b>2 Ethical Preliminaries</b>	<b>9</b>
2.1 Bias . . . . .	9
2.2 Discrimination . . . . .	11
2.3 Human Rights . . . . .	11
2.4 Equality & Equity . . . . .	12
2.5 Fairness . . . . .	14
<b>References</b>	<b>17</b>



# Chapter 1

## Introduction

The introduction is one of the core chapters of your thesis. It expands what has already been said in the abstract with additional details on the content and contribution and on the structure of the thesis. It is meant to introduce the reader to the work he/she will be reading in the rest of the document and, most importantly, to get the reader curious about reading on, knowing more about your work.

### 1.1 Context: [topic]

This thesis is about describing the work you are doing in your final thesis project. You have been working on it for months, and nobody knows the work better than you do. This is great and exactly how things should be: by doing your thesis project you became an expert – if not *the* expert – in this specific field of research and/or technology.

But attention: being the expert is also dangerous when it comes to explaining others what you did and why you think you did a great work that deserves attention (I give it for granted that you work does so). There are only very few people around you (your supervisor and possible co-supervisor, some friends, maybe someone else) who are as expert as you are in this topic. So, if you start in a full-impact fashion to tell that you implemented an extraordinarily cool, new algorithm to solve X, or that you discovered this extremely surprising finding Y, or that you mathematically proofed that Z, etc. (you got it), your reader will not understand anything. Therefore, before talking about what you actually did, you need to introduce the reader to the context of your work, provide the necessary core definitions that are needed to understand the terminology you will be using in the rest of the thesis (if it's not standard IT terminology).

Therefore:

- ☐ Tell the *research area(s)* your work/project focuses on. If you are doing your thesis with me, likely candidates of research areas are Web Engineering, Data Science, Crowdsourcing, Service-Oriented Computing, Business Process Management.
- ☐ Tell possible *sub-areas* that are more specifically related to what you are doing. Again, if you are doing your thesis with me, likely candidates of sub-areas are chatbots, social knowledge extraction, business process matching/modeling, quality control in crowdsourcing, etc.
- ☐ Make the *heading* of your context section self-explaining by substituting “[topic]” in heading 1.1 with the sub-area most relevant to your work. It should read like “Context: quality control in crowdsourcing” or similar.
- ☐ If needed, introduce some *key definitions* (no need to introduce everything here, but be sure that the introduction does not use terminology the reader may not be familiar with). For instance, if you are working on chatbots, this is definitely a term that needs to be introduced here; it’s not yet commonly known but it’s crucial for the understanding of the rest of the thesis and introduction.
- ☐ Use *examples* to make definitions and ideas concrete and clear.
- ☐ Throughout, make *references* to the relevant literature.

#### **Use of tenses and pronouns**

Writing a thesis is writing a scientific document like scientific articles or research publications. There are two conventions that are usually applied in this kind of publications (admittedly, they may seem somewhat odd if not used to):

First, the most used tense is the *simple present*. The thesis is meant to describe a piece of work, from problem statement, to the conception of a solution, its implementation and evaluation. Yet, it’s not a novel about your life, and it’s not meant to provide a chronological story about what you did and didn’t do. Content is presented in an order that is most effective to convey its message, not in time order. In this spirit, it’s much more effective to say “in order to get result A, first we do X, then we do Y and then Z,” instead of saying “in order to get result A, we did Y after having done X, then we went on doing Z.” The order of actions, their interconnections, inputs and outputs already tell the dependency – if properly described. Most of the times, the most effective way to describe a solution or

methodology only becomes clear after trial and error. It's enough to explain the result, not how you got there chronologically.

Second, the *pronoun* used to talk about the own work is "our" (work). That is, it is custom to say "we" instead of "I," even if you are writing your thesis alone. However, don't forget about all the people that helped you get there: your supervisor, co-supervisor, colleagues, etc. This may sound strange at the beginning, but, at the other hand, using "I" too often risks to convey the impression that you are self-focused and egoistic, which is never good.

## 1.2 Scenario and Problem Statement

Now that the reader got the general context of your work and has an intuition of the problem you will be solving in the rest of the thesis, it's time to be clear about which *specific problems* your thesis project is going to solve. One way of doing so is by describing a *scenario* (a description of a real situation, with all its actors, roles, tasks, instruments, etc.) that provides evidence that there are one or more real problems right now that, with the current technology and understanding of the domain, are hard to solve or not solvable at all. If instead the problem(s) can be solved already, it should be evident from the scenario that this is possible only at a prohibiting cost or with unsatisfying guarantees on the quality of the result or not within useful time for the target user.

It's important that the scenario is written in such a way that the reader, after reading it, agrees with you that the problem you are focusing on is a relevant one, one that deserves being studied and solved. Consider that if you convince the reader here that your thesis is needed (after all, that's what this section is about), he/she will be very open to possible solutions and happy to see how you solve it. If instead you fail to convince the reader – let me be harsh – the whole rest of your thesis is useless in the eyes of the reader. This is the worst outcome you want.

Conclude this section by explicitly stating which of the problems evident in the scenario you are approaching. Don't raise false expectations! Never ever tell the reader there are five core problems and then solve only two of them in the thesis, without telling upfront that this is what you intended to do in the first place. As soon as you list problems, the reader wants to see a solution, unless you stop him/her immediately from thinking so by telling that out of the described problems you focus on a subset only, usually because this subset is already a huge research and development problem in its own.

In summary:

- ☐ Describe a *real scenario* that provides evidence of *real problems*.
- ☐ Convince the *reader* that the problems need to be solved.
- ☐ Use an *illustration* or *figure* to help the reader understand.
- ☐ If possible, provide *references* to literature that backs your assessment of the problem.
- ☐ Provide a clear *problem statement* that summarizes what came out of the scenario and your specific focus.

### 1.3 Methodology

Fixed the problem(s) you want to approach, you can approach it/them in thousands of different ways. Your way is just one of the thousands, and the reader may have (and very likely will have) a very different intuition of how to solve the problem(s) you just pointed out. So, clarify how you intend to proceed:

- ☐ Tell if you follow an existing *methodology* or not; if yes, name it and provide a reference to literature, if available. For example, Design Science [?] is a likely methodology to cite here.
- ☐ Tell which of the following *procedures*, *techniques*, *methods* you use in your work and for which purpose (put them also into the right order, so that their application or use makes immediate sense to the reader):
  - ☐ *Systematic literature review, survey*
  - ☐ *Statistical hypothesis formulation and testing*
  - ☐ *Software prototyping*
  - ☐ *Iterative development*
  - ☐ *Participatory design*
  - ☐ *Performance evaluation*
  - ☐ *Comparative studies*
  - ☐ *User studies*
  - ☐ *Expert interviews*
  - ☐ *Simulation/emulation*



- ☐ *Live experiments*
  - ☐ *Case studies*
  - ☐ *Mathematical theorem proving*
  - ☐ *Mathematical modeling*
  - ☐ *Pseudocode*
  - ☐ *Graphical modeling* (e.g., UML, ER)
  - ☐ *Model-driven development*
  - ☐ *Automatic code generation*
  - ☐ ...
- ☐ Tell if you use some special *software instruments* that help you in your work. We are of course not talking about Word or Google Search. Perhaps you can tell that you used R for data analysis or some specific modeling instrument for automated code generation or simulation.

## 1.4 Contributions

Now that the reader knows what you want to solve and how you intend to proceed, you can anticipate the contributions your thesis makes to the state of the art. Attention, a thesis project may produce lots of different *outputs* (e.g., a software prototype, a set of registrations and transcripts of interviews, datasets collected during experiments) and *contributions* (e.g., a demonstration that some software solutions solves a given problem under well defined conditions, a formal proof that some property holds, empirical evidence that something works as expected). The former are all the artifacts produced throughout the work. The latter refer to *new knowledge* (if you are doing a full thesis) or the most important, *final output* (if you are doing a tesina). Sometimes, outputs and contributions overlap, but not necessarily.

Typical contributions are (multiple choices may apply to your thesis):

- ☐ A *systematic literature review* of the state of the art providing evidence for some argument
- ☐ The design of a *model* (mathematical, graphical, algebraic, etc.) describing how to solve a real world problem in a reusable fashion
- ☐ The drawing of *conclusions* (findings) from the analysis of a dataset describing some physical or virtual phenomenon

- ☐ The implementation of a *software prototype* solving a real world application problem
- ☐ The design of a *language* (textual, graphical) enabling others to solve own problems or to solve them easier
- ☐ *Formal proofs* of correctness, completeness or other properties of the proposed models or theorems
- ☐ *Objective evidence* from empirical studies (e.g., performance analyses or simulations) that demonstrate that the proposed prototype or solution works / works better than existing software or solutions that solve the same/similar problem(s)
- ☐ *Subjective evidence* from user studies or expert interviews backing the claims of viability of the proposed problem or solution/artifact
- ☐ A reasoned *argumentation*, e.g., based on a detailed case study, supporting the viability of the proposed problem or solution/artifact

#### **Thesis vs. Tesina**

Let me spend some words on the difference between these two. Before that, however, it is important to clarify the very purpose of your final project, be it a thesis or a tesina (a small thesis). The purpose of it is giving you the possibility to show that, after years of attending classes and giving exams, you are also able to *apply* the knowledge you acquired during your studies. In short, it's all about you showing that you are *mature*. Mature from a knowledge perspective, mature from an application perspective, mature from a work/teamwork perspective, mature from an ethical perspective.

It is common that a thesis project is not very well defined in its beginning and that even the supervisor does not really know how to approach a given problem or which problem to focus on in the first place. This may even be annoying to you, but attention: there is no intention behind it. Your supervisor is not withholding information from you to test you or to see if you get something. It's just the nature of real *problem solving*. If things were clear from the beginning, there wouldn't be any problem! Fledging out the problem and agreeing on a solution and methodology is a core part of you demonstrating your maturity – if not the most important one. *How* you proceed from the inception of the thesis idea to the final solution is as important as *what* you find and/or produce in the end.

This being said, a *thesis* in Politecnico di Milano usually requires you to make a contribution to the literature (the so-called state of the art). Making a contribution – from a science point of view – means creating new *knowledge*, that is, finding something that nobody knew before, demonstrating a property that nobody showed

before, improving the performance of a given system with a new algorithm, and similar. For a thesis, it is therefore not enough to produce a perfectly engineered solution. It is key that you also demonstrate, provide empirical evidence or proof that your solutions performs as claimed. Well, for a *tesina* this last demonstration is usually not required, and the focus is on the engineering of the solution. In addition, perhaps in the case of the *tesina* the solution to be engineered is also less complex then for a thesis, but this depends on the context and on how you want to measure complexity.

## 1.5 Structure of Thesis

Here you explain the structure of the thesis, so that the reader knows how to read it. Consider that not every reader wants to read through the whole thesis to find some specific information. Actually, only few will do so (your supervisor and co-supervisor, and the possible reviewer for sure). Many more will just leaf through it and look for specific types of information (e.g., the context of your work, your findings, how you implemented something, which technologies you used). It is your duty to accommodate them all. How? By telling them how your thesis is structured.

Therefore, in this section you provide a brief description (2-3 sentences) for *each* chapter that follows this introduction. Use an itemized or numbered list to structure the text, like this:

- ☐ Chapter 2 introduces the state of the art and...
- ☐ Chapter 3 provides...
- ☐ ...

### Structuring text

Besides telling the reader how the content of your thesis is organized into chapters, it is important that you master some basic text structuring techniques. To organize your text there are lots of instruments you can use: chapters, sections, sub-sections, paragraphs, itemized lists, numbered lists, code examples, figures, images, screen shots, captions below figures, tables, and so on. Use them all! Don't write text without structure. Never.

Be aware that the structure of your text, that is, how you present your work, conveys a lot of information about how well you actually understand what you are writing about, how much you care about being clear and helping your reader understand, and how much value you give yourself to your own thesis. A well

structured presentation of content that the reader can understand and agree with is a huge plus in this respect. Text that lacks proper paragraphs, does not use lists where needed, etc. is a minus and also much harder to read (think about how much a well structured text can help you go back ten pages and find concepts you know you read about compared to a text that comes without an easy to memorize formatting and structure). When writing, think about some of your textbooks. Since you are doing an engineering degree, I'm sure these are textbooks that make exemplary use of the different formatting instruments available.

## Chapter 2

# Ethical Preliminaries

The aim of this chapter is to provide to the reader preliminary notions of ethical and sociological rather than technical nature.

Starting from the concept of *bias*, passing through *discrimination* and *human rights*, we will discuss about *equality*, *equity* and finally *fairness*, by providing definitions and relevant examples from the literature on these topics. Because of the “dual nature” of this research, this chapter has to be seen as complementary to the next one, in which some technical bases will be provided, together with an overview on the tools adopted.

### 2.1 Bias

Although the word “**bias**” doesn’t have an intrinsically negative meaning (it is informally used to indicate a deviation from neutrality), it is mostly adopted in contexts where it entails a moral dimension. As reported in [7, p.332], *we use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate.* Therefore, it is important to underline that unfair discrimination due to bias is strictly related to systematic and unfair outcome.

By following the classification provided in [7], we can distinguish three overarching categories of bias:

- **Preexisting bias:** it has its roots in social institutions, practices and attitudes. Preexisting bias may originate in the society at large, in subcultures and organizations, and can enter a computer system

either voluntarily or implicitly and unconsciously, even in spite of the best intentions of the system designer.

- **Technical bias:** it arises from the resolution of issues in the technical design. Technical bias may originate from design choices, constraints and technological tools, or exacerbate preexisting bias.
- **Emerging bias:** it emerges some time after a design is completed, as a result of changing societal knowledge, population, or cultural values. Emerging bias is strictly related to a context of use, and it is the most difficult to detect.

For the purpose of this research, we will focus on preexisting bias (in particular, *societal bias*) and technical bias, but it is important to point out that emerging bias shouldn't be underestimated in the long run, especially when it arises from a *mismatch between users and system design* due to *different values*, that is bias originated when a computer system is used by a population with different values than those assumed in the design, because society is in constant change and systems should be readjusted or reinvented in order to keep up with the present.

A significant example of bias is provided in [1]: a commercial tool called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) was used in courts in the U.S. to automatically predict some categories of future crime to assist in bail and sentencing decisions. On average, the tool correctly predicted recidivism 61% of the time, but blacks were almost twice as likely as whites to be labeled a higher risk but not actually re-offend. The tool made the opposite mistake among whites: they were much more likely than blacks to be labeled lower risk but go on to commit other crimes.

Other examples, related to gender bias, are given by [8] and [5]. The former is about a research conducted in 2015, in which a tool was used to simulate job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender. One experiment showed that Google displayed adverts for a career coaching service for “\$200k+” executive jobs 1,852 times to the male group and only 318 times to the female group. The latter concerns another Big Tech company, Amazon, whose machine-learning's specialists, back in 2015, discovered that their new recruiting engine was not rating candidates in a gender-neutral way, because the system taught itself that male candidates were preferable by penalizing resumes that included the word “women's”.

## 2.2 Discrimination

Bias can lead to **discrimination**, but what discrimination is and how it occurs is a controversial issue. As reported in [12], often the law, rather than providing a definition of discrimination, defines a list of attributes, called **protected attributes**, that cannot be used to take decisions in various settings. The list is non-exhaustive and includes characteristics such as race, sex, religion or sexual orientation. Groups of people characterized by these attributes are therefore classified as *protected groups*. Trying to formalize a bit more, we can define discrimination as the result of:

- **Disparate treatment**: the illegal practice of treating an entity, such as a job applicant or an employee, differently based on a protected characteristic such as race, gender, age, religion, sexual orientation, or national origin.
- **Disparate impact**: the result of systematic disparate treatment, where disproportionate adverse impact is observed on members of a protected class.

On a more sociological perspective, discrimination is strictly related to *human rights* and *equality*.

## 2.3 Human Rights

For what concerns **human rights**, there is a lot of debate on how to incorporate them in computer systems by following a “human-rights-by-design” approach [11], in order to contrast the negative effects of the so called “dual-use technologies”: products which may serve legitimate societal objectives but are also used to undermine human rights like freedom of expression or privacy. Of course, to reach the goal it would require a culture shift and huge efforts from both national governments and businesses, in order to let them design tools, technologies and services to respect human rights by default, rather than permit abuse or exploitation of them. A similar concept is purposed in [16], where the authors sketch the contours of a comprehensive governance framework for ensuring AI systems to be ethical in their design, development and deployment, and not violate human rights. This framework should be effective in contrasting *ethics washing*: the practice of fabricating or exaggerating a company’s interest in equitable AI systems that work for everyone, a sort of side door that companies use to substitute regulation with ethics. For the purpose of this research, we can define human rights as *inalienable fundamental rights to which a person is inherently*

*entitled simply because she or he is a human being* [13, p.3]. These norms are concerning every human being, regardless of sex, age, language, religion, ethnicity or any other status.

## 2.4 Equality & Equity

Equality is generally intended as *an ideal of uniformity in treatment or status by those in a position to affect either* [3]. The concept of equality is often associated with discrimination mostly because of the article 7 of the Universal Declaration of Human Rights (UDHR), which states: “*All are equal before the law and are entitled without any discrimination to equal protection of the law*” [2, p.2]. This principle is known as “equality before the law”, and establishes that everyone must be treated equally under the law regardless of race, gender, color, ethnicity, religion, disability, or other characteristics, without privilege, discrimination or bias. However, it is important to distinguish between two different political and social theories:

- **Equality of opportunity:** *the idea that people ought to be able to compete on equal terms, or on a “level playing field”, for advantaged offices and positions* [10]. This principle is based on the notion of *sameness*, where fairness is achieved through equal treatment regardless of people’s needs. Equality of opportunity is usually simply referred as **equality**, and from now on we will adopt the same terminology for this research.
- **Equality of outcome:** the idea that people should have access to resources (possibly of a different nature and to a different extent) in order to be able to reach the same condition. This principle is based on the notion of *need*, where fairness is achieved by treating people differently depending on their endowments and necessities. Equality of outcome is also known as **equity**, and from now on we will adopt the same terminology for this research.

Figure 2.1 provides a simple example of the difference between equality and equity. Treating people equally, in this scenario, means to give everyone the same one box to reach the fruit, while treating people equitably means to give them as many boxes as they need to achieve the goal. It is important to notice that equity could require (and often requires) unequal treatment.

The authors of [9] distinguish between four different facets of data equity (which, as the name itself suggest, is an extension of the equity concept in the technical context of data systems):



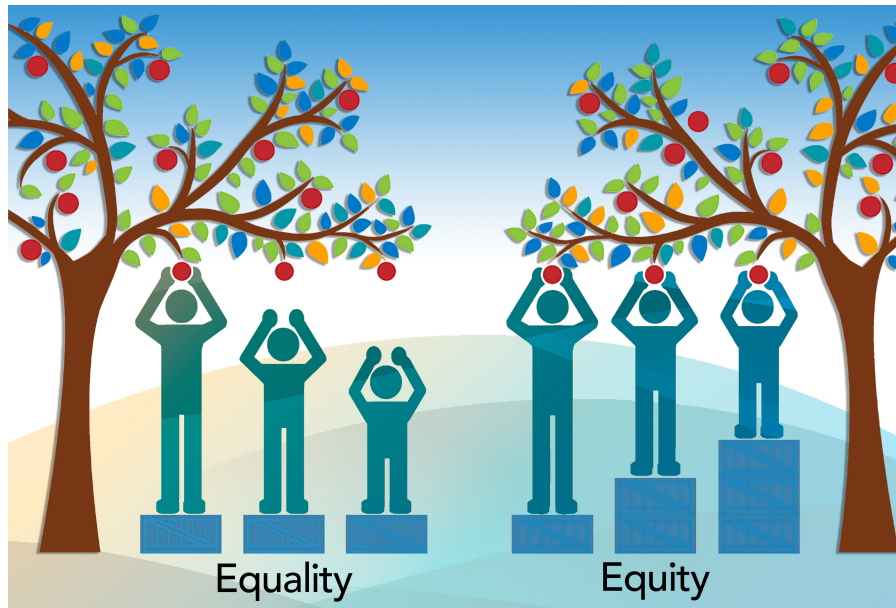


Figure 2.1: Visual example of the difference between equality and equity.

©2014, Saskatoon Health Region. Source: <https://www.nwhu.on.ca/course/rvices/Pages/Equity-vs-Equality.aspx>

- **Representation equity:** bias may arise because of material deviations between the data and the world represented by the data, often with respect to historically disadvantaged and underrepresented groups. Even when dealing with contemporary data, disparities rooted in historical discrimination can lead to representation inequities and therefore to the introduction or the exacerbation of problems. For example, in the U.S. there has been a lot of discussion about racial disparities concerning COVID-19, regarding both availability of testing (fewer test sites in minorities neighborhoods, historically poorer) and desire of individuals to be tested (black people more suspicious about the medical system, because of their history of unfair treatments). Another example is related to Amazon and a software developed by the company for screening candidates for employment: the software was trained on the already hired employees and since they were mostly males, females became underrepresented in the data and the software was much more likely to mark women as unsuitable for hiring.
- **Feature equity:** bias may arise because not all the features needed to represent a marginalized group of people and required for a particular analysis are available in the data. As an example, for a specific study

involving transgender people may be important to distinguish between their birth name and their self-assigned name.

- **Access equity:** bias may arise because of a not equitable and participatory access to data and data products across domains and levels of expertise, or even because of the opacity of data systems.
- **Outcome equity:** bias may arise because of a lack of monitoring and mitigation of unintended consequences for any group affected by the system after deployment, directly or indirectly (for example, contact tracing apps may facilitate stigma or harassment).

## 2.5 Fairness

As discussed in the previous section, both equality and equity aim to achieve **fairness**, despite the different approaches of the two philosophies. We can therefore distinguish between two different concepts of fairness [6]:

- **Individual fairness:** any two individuals who are similar *with respect to a task* should receive similar outcomes. The similarity between individuals should be captured by an appropriate metric function, usually difficult to determine. Individual fairness is strictly related to the idea of equality.
- **Group fairness** (also known as *statistical parity*): demographics of the individuals receiving any outcome - positive or negative - should be the same as demographics of the underlying population. Group fairness equalizes outcomes across protected and non-protected groups, and is therefore strictly related to the idea of equity.

Despite individual and group fairness are not mutually exclusive in theory, in real life it is often hard to conciliate the two approaches. Furthermore, this categorization is not the only possible one: the authors of [14] collected and provided about twenty among the most prominent definitions of fairness, and applied each of them to a case study based on gender-related discrimination, in which the aim was to assign a credit score to people requesting a loan by using “Personal status and gender” as a protected attribute for the decision-making process.

Among the others, a couple of peculiar definitions, often listed together with individual and group fairness, are the following:

- **Fairness through unawareness:** protected attributes are not used in the decision-making process, and therefore the subsequent decisions

cannot rely on them. This “blind” approach relies on *impartiality* and is consistent with the disparate treatment principle, but removing features means losing information, and furthermore there could be features correlated to protected attributes that would not be removed, potentially introducing bias.

- **Counterfactual fairness:** a precise and non-technical definition is provided in [15, p.1]: *a model is fair if for a particular individual or group its prediction in the real world is the same as that in the counterfactual world where the individual(s) had belonged to a different demographic group. However, an inherent limitation of counterfactual fairness is that it cannot be uniquely quantified from the observational data in certain situations, due to the unidentifiability of the counterfactual quantity.*

The classifier resulted to be fair depending on the notion of fairness adopted, showing the impossibility of addressing fairness as a unique, broad and inseparable concept. This result is coherent with *Chouldechova’s impossibility theorem* [4], which demonstrates, taking three definitions of fairness, the impossibility of satisfying all of them.



# Bibliography

- [1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias. *ProPublica*, 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- [2] UN General Assembly et al. Universal Declaration of Human Rights. *UN General Assembly*, 302(2):14–25, 1948.
- [3] The Editors of Encyclopaedia Britannica. “Equality”. *Encyclopedia Britannica*, 2009. <https://www.britannica.com/topic/equality-human-rights>. Accessed 7 June 2021.
- [4] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163, 2017.
- [5] Jeffrey Dastin. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 2018. Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-thatshowed-bias-against-women-idUSKCN1MK08G>.
- [6] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness Through Awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226, 2012.
- [7] Batya Friedman and Helen Nissenbaum. *Bias in Computer Systems*. Routledge, 2017.
- [8] Samuel Gibbs. Women less likely to be shown ads for high-paid jobs on Google, study shows. *The Guardian*, 8(7), 2015. Available at: <https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>.

- 
- [9] HV Jagadish, Julia Stoyanovich, and Bill Howe. The Many Facets of Data Equity. In *24th International Conference on Extending Database Technology (EDBT)*, 2021.
  - [10] Andy Mason. “Equal opportunity”. *Encyclopedia Britannica*, 2019. <https://www.britannica.com/topic/equal-opportunity>. Accessed 7 June 2021.
  - [11] Jonathon Penney, Sarah McKune, Lex Gill, and Ronald J Deibert. Advancing Human-Rights-by-Design in the Dual-Use Technology Industry. *Journal of International Affairs*, 71(2):103–110, 2018.
  - [12] Teresa Scantamburlo, Andrew Charlesworth, and Nello Cristianini. Machine decisions and human consequences. *arXiv preprint arXiv:1811.06747*, 2018.
  - [13] M Sepuldeva, Th Van Banning, Gudrún Gudmundsdóttir, Christine Chamoun, and Willem JM Van Genugten. *Human Rights Reference Handbook*. University for Peace, 2010.
  - [14] Sahil Verma and Julia Rubin. Fairness Definitions Explained. In *2018 IEEE/ACM international workshop on software fairness (fairware)*, pages 1–7. IEEE, 2018.
  - [15] Yongkai Wu, Lu Zhang, and Xintao Wu. Counterfactual Fairness: Unidentification, Bound and Algorithm. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019.
  - [16] Karen Yeung, Andrew Howes, and Ganna Pogrebna. AI Governance by Human Rights-Centered Design, Deliberation and Oversight: An End to Ethics Washing. *The Oxford Handbook of Ethics of AI*, page 77, 2020.