Ca' Foscari University of Venice

Master's Degree programme

In International Management

Final Thesis

Ethical implications of Artificial Intelligence: cases from Italy

Supervisor
Prof. Vladi Finotto

Assistant supervisor
Prof. Andrea Stocchetti

Graduand
Erika Samorè
Matriculation Number 862137

Academic Year
2020 / 2021

TABLE OF CONTENTS

INTRODUCTION

Despite public belief, Artificial intelligence technologies are frequently embedded in software and applications. As a student of management, I have often analyzed what are the potential benefits for an organization of using new technologies, from reducing cost through automatization to creating new market opportunity and business model. Moved by my personal interest on the topic, I have read the book *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* by Cathy O'Neil, suggested in a lecture of strategy by Professor Finotto. The book highlights what are the ethical damages on the population connected with the use or misuse of software and algorithms by organizations. Intrigued by the arguments proposed by O'Neil, expert in data science, I started doing researches on the Artificial Intelligence technologies and on what could be their negative impact on the average user, from privacy breaches to being influenced through a recommendation system on how to vote in the presidential election. The aim of my research is to analyze whether Italian companies developing such technologies are aware of the negative impacts that these technologies have on the final users and what are the steps they take to prevent them. In order to gather and thus analyze data, I have interviewed Italian companies of different sizes belonging to the tech industry.

This dissertation is the outcome of my research on the ethical implications of Artificial Intelligence (AI). In the first chapter, there is a summary of key concepts and milestones connected with the development of AI from the first computer developed by Alan Turing to currently application on how to defeat Covid-19 pandemic. The second chapter reviews the today literature on ethical implications of AI starting from one of the most significant paper: *Ethics of algorithms, mapping the debate* (Mittelstadt et al., 2016). Each cluster of ethical problem is discussed and explained with examples of real damaged caused by companies to specific ethnical group or gender. Instead, in the last chapter there is a description of empirical research – how companies were selected and on the structure of the interviews. The final part of this chapter is dedicated the discussion of data gathered. Especially the focus is on the most important ethical challenges connected with the development and / or use of AI technologies that have emerged during all the interviews

that are privacy and social responsibility. Just few companies are fully aware of the side effects of AI technologies and hence they are taking proactive measures to reduce the probability of being responsible of negative consequences on the population. The fact that only few companies are doing so it is not linked with indifference but rather to a culture. In fact, Italy is not used to deal with ethics but rather with regulation. This is the reason why European Union should start legislating on these new technologies that are potentially capable of causing severe damages on individuals.

CHAPTER 1

The term Artificial intelligence (henceforth "AI") has been around for more than 60 years, but there is still no common agreed definition. It can be considered as a discipline of computer science with the aim of developing computers able to act like humans in the sense that they learn, they reason and they correct themselves (Kok et al., 2009). However, its definition changes with its evolution. In fact, once a machine is able to perform a specific task we tend to give it for granted and thus not considering it as a proof of intelligence. Nowadays it is more common to define AI as "*a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*" (Kaplan and Haenlein, 2020). Even this definition will be outdated once artificial superintelligence is reached meaning when a machine is able to be creative and social.

There are three types of AI: the first one concerns building system that thinks like human; the second one is to complete a task like robots and last to use human reasoning as a functioning model. Another necessary division regards strong and weak AI. With the former, we define machines that have an intelligence equal to human being meaning that they have a consciousness. For instance, strong AI is able to produce books or scripts that are undistinguishable from a human being. Instead, the latter, referred to as narrow AI, focuses on specific tasks and if needed it emulates the human consciousness. Virtual assistance such as Siri is an example of weak AI (IBM Cloud Education, 2020). This classification is a consequence of its evolution from strong to weak AI. Despite the improvement in the field, the major issues faced at the beginning is still present today and it involves transforming problems into mathematical equation easily understood by computers. This, together with the evolution of AI, will be explained in detailed in the following sections.

1.1 History of Artificial Intelligence

Alan Turing can be considered as the pioneer in the development of AI with his stored-program concept developed in 1935 and his idea was that machine can think. In order to test the ability of the machine to think, he developed a test known as Turing Test in 1950.

The test consists of two steps, in the first one a machine and a human have to reply to some questions and in the second step a judge has to distinguish who wrote the answers. If the judge is not able to detect the machine's answers, the test is considered successful. Unfortunately, to find a machine able to pass this test we have to wait until 2014 with Eugene Goostman, a cleverbot developed by Vladimir Veselov and Eugene Demchenko (Zaccaro, 2014).

Despite the innovative research of Alan Turing, the concept of AI is conventionally linked to John McCarthy and with the ability of the computer to store commands. In fact, Turing stored program was not real until 1949 and before that machines were just able to provide the output but they had not records of inputs. The first public discussion about AI took place at the Dartmouth Summer Research Project on Artificial Intelligence in 1956. During the convention, McCarthy defined for the first time AI as *"the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it"*. A problem of this definition is that it does not explain how to define intelligence. Therefore, a lot of invention can fall under the definition of AI. According to Newell and Simons, their program, Logic Theorist, was improving the intelligence of the machine. Indeed a computer was now able to prove mathematical theorems – nowadays the Logic Theorist is recognize as the first AI algorithm.

In the two decades following the Research project, there has been a sharp increase in the research and development of AI in different fields. As a result of the cooperation among universities and tech companies and with investment of the U.S. government, computers with new software were provided to the market at a cheaper price and with higher ability to process information. The latter was mainly due to the microprocessor – an important innovation made by Intel in 1969 and marketed two years later with the name Intel 4004 processor (Intel, ND). Furthermore, there has been an increase in complexity of the programs in the computers. This has been possible due to the higher quality of computers and at the same time of the development of high-level programming languages – the more popular was Lisp developed in 1959 at the Massachusetts Institute of Technology. It soon

became the common language for AI program especially because it allowed machine to self-modify the programs through their own learning (Hemmendiger, 2000).

Another important milestone for the AI is the introduction of the expert system by Feigenbaum in 1982. In his paper (Expert system: principle and practice, 1992), he described it as a "*program that reasons using knowledge to solve complex problems*" rather than using only arithmetic calculations. Before expert systems, knowledge was provided to a computer by human through a program. Whereas, now an expert could provide his knowledge containing both facts and heuristics on how he would behave in specific situations. With this expert system a non-expert user could describe a specific problem and the computer is able to provide a solution using the knowledge previously acquired with the expert. However, the main problem with this technology lays in its over specialization. Therefore, if the problem of the user is outside the knowledge of the machine, the system fails in providing an answer since it is not able to generalize and draw conclusions by itself (Feigenbaum, 1992).

Overall, from the end of the 70s and during the 80s there has been a decrease in the growth of innovation as a consequence of skepticism of the U.S. Congress. For them it was not worth to invest more on AI, because there were not enough progress to be made and the technology had already reached a mature stage (Haenlein, 2019). However, their thoughts were proven wrong in the following decades. In fact, the 90s ware full of milestones for the AI. For example, in 1996 for the first time a computer program (IBM's Deep Blue) won a chess match against a chess champion – Gary Kasparov. It was such a success because the machine was able to imagine combination of moves and possible scenarios way beyond human capability. Another milestone is Kismet, an artificial robot able to display emotions and behave like a human being especially when it comes to learning. Indeed, Kismet through its sensor eyes is able to learn from the environment as a child does (MIT, 2001). Both examples underline the relevance of the knowledge in the AI.

A drastic shock in the industry has happened with the development of World Wide Web and its availability to the private consumers through Netscape first then Yahoo! and

Google. With World Wide Web it is now possible to collect an huge quantity of data regarding users and therefore it was necessary to create some technologies that could facilitate the management of infinite data such as super-fast speed data processing machines (Duan et al., 2019). Moreover, new fields of research of AI have emerged in particular text mining, information retrieval and semantic web (Ramos et al., 2008). Regardless the progress made in AI from the 1950s to 2010s, a proper boom happened in the last decade. In fact, according to the World Intellectual Property Organization, over half of the inventions in the field have been published from 2013 on. According to Chethan Kumar (2018), all of these innovations can be grouped into six cluster:

- MACHINE LEARNING identifies algorithms that allows a machine to learn by itself when it faces new data. The machine is able to make predictions or association without being provided with the necessary codes. An example of this is the ability of Google's API Vision[1] to assign to a photo huge amount of tags to better classify it for future researches (Google).

- NATURAL LANGUAGE PROCESSING is used to understand, interpret and manipulate natural languages like speech and text by a software. Belonging to this category there is the spam filter that it is able to detect which are important mails.

- EXPERT SYSTEM as we described above are useful to emulate the decision making of human being. For example, DXplain[2] is a support system in the healthcare sector that suggests diseases based on previous cases.

- VISION is defined as the ability to see and capture visual information. It performs better than a human eye since it allows a machine even to see through walls. An interesting case is the Google translate app that allows a user to translate directly from a picture.

---

[1] https://cloud.google.com/vision/docs
[2] http://www.mghlcs.org/projects/dxplain

- SPEECH includes the ability to read a text in any language but more important to transform a speech into a text. The most common way to achieve this objective is through Python[3] – a high level programming language.

- PLANNING involves the decision making process and the sequence of actions that allows a machine to achieve a specific goal.

- ROBOTICS allows machines to perform tasks that are difficult for a human being because they require precisions, consistency or heavy lifting objects. From just manufacturing robots to socially-interactive one such as the waitress-robot Amy[4].

Another important point to underline is the emergence of China in the AI development. In fact, so far we have discussed about inventions and discoveries that have been possible due to huge investments from US government first and then private organizations. However, in the last decade given the improvement of AI, China has decided to step in. In 2013, China was investing and researching more than the US on deep learning technologies allowing the nation to be so competitive in the upcoming years (Meyer, 2018). The main disadvantage for China was its lack of competences on the hardware. As stated by Zheng, director of the Institute of Artificial Intelligence and Robotics at Xi'an Jiao tong University, China is still dependent on American corporations for semiconductor chips essentially for supporting AI. Whereas, when talking about software China is the world leader on computer vision, speech recognition and natural language processing (O'Meara,2019). The aim of Chinese government is to catch up to the US by 2025 and becoming the world leader by 2030 (Toscano, 2019). This will be possible due to a different attitude between US and China towards AI. Given the political configuration of China and the fact that the government controls the evolution of AI, it is possible to implement and improve more advance technologies even though they do not reach popular consensus. This is the example of facial recognition cameras used to fight against criminality in China. The same technology has been adopted but soon banned in San Francisco, because in US AI is driven by high tech companies that are profit oriented but

---

[3] https://www.python.org/
[4] https://www.servicerobots.com/amy-waitress/

the adoption of new technologies is influenced by public opinion. The future of AI depends largely on those two countries. Moreover, given the different perception on privacy even when comparing them to Europe it is possible that in the future there will be a creation of three separated networks aimed at providing different services depending on the local legislations and restrictions (Kaplan and Haenlein, 2020).

1.2 How AI is shaping the world

With the Big Data technologies, there is a new era for AI and for businesses. Companies like Google, Amazon or Netflix are using big data and their analysis to improve products and increase services and markets (Borges et al., 2020). Indeed, organizations started to adopt more AI tools that transform the way they do business. Nowadays these technologies are employed in almost every industry, from health care to transportation as shown in Figure 1, and in different business functions like manufacturing and marketing.
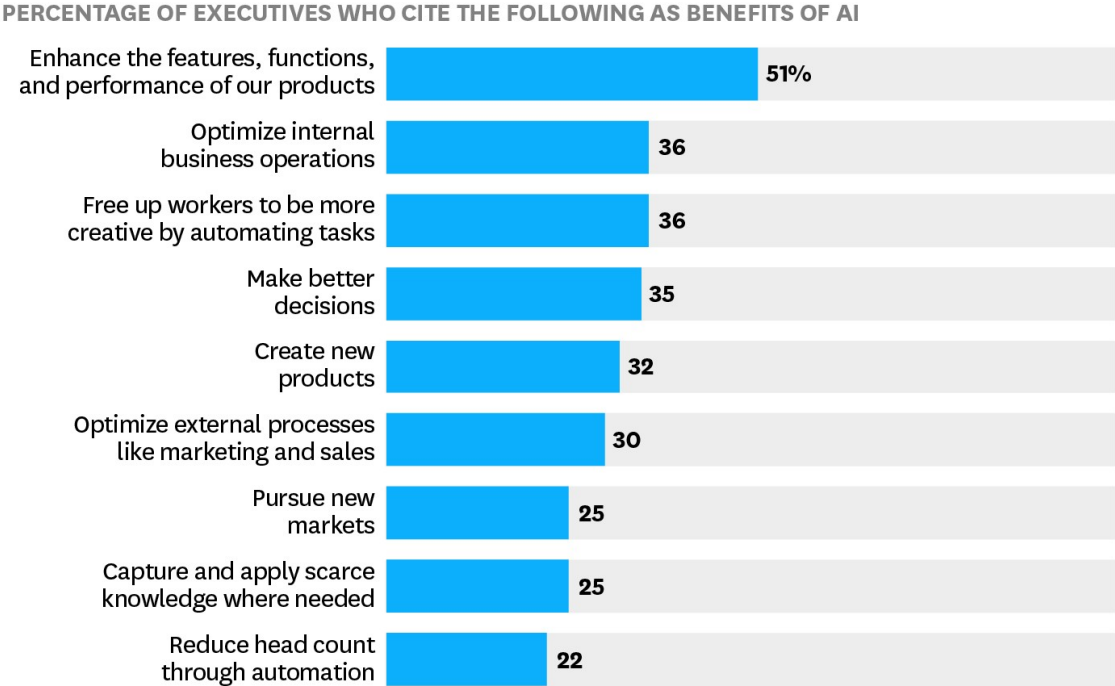
FIGURE 1
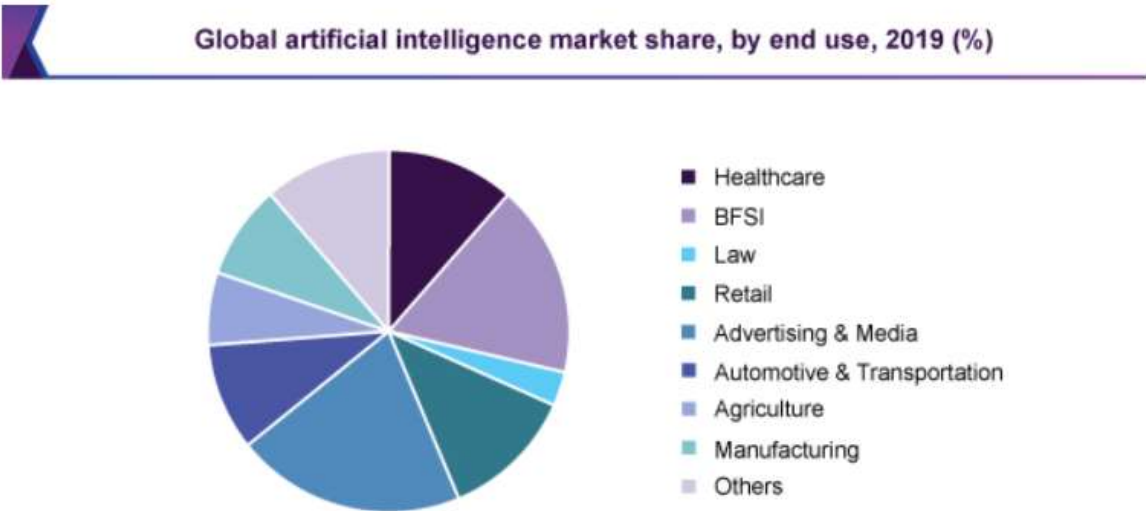
Source: Grand view research, 2020

The common belief among business managers is that AI will reshape their business model in the short-term period (Duan et al., 2019). Moreover, this is not just a business perception but rather a world vision. In fact, every nation is incentivizing businesses into

the adoption of the AI to achieve what is called Industry 4.0. The key enabling technologies of the fourth industrial revolutions are the management of big data, real time adjustment and Internet of things. The latter allows objects and machine to be connected to one another and communicate with each other. The implementation of AI allows this interconnection to be more efficient and autonomous requiring less supervision from human (Ghosh et al., 2018). The new wave of technological development influences also small businesses with a lower financial investment capacity. Indeed, firms can have access to knowledge and experts through web open platform at a cheaper price. Moreover, there is no need to develop inside a technology but it could be rented for just a period.

The implementation of AI inside organizations has several benefits. According to an analysis made by Davenport and Ronanki (2018) of 250 firms that have implemented AI, more than half of the adopters were able to enhance the feature and performance of their products. From Figure 2, we can say that overall it has increased the efficiency of the businesses in their operations that leads to a better allocation of employees toward more creative centered tasks.

FIGURE 2

PERCENTAGE OF EXECUTIVES WHO CITE THE FOLLOWING AS BENEFITS OF AI

Enhance the features, functions, and performance of our products	51%
Optimize internal business operations	36
Free up workers to be more creative by automating tasks	36
Make better decisions	35
Create new products	32
Optimize external processes like marketing and sales	30
Pursue new markets	25
Capture and apply scarce knowledge where needed	25
Reduce head count through automation	22

Source: Davenport and Ronanki, 2018

As we have discussed before and as the figure confirms, robots are substituting non-specialized and low-paid workers. Instead, as the pay increases, the risks to be substituted by a machine decreases (Muro, 2019). Even though the automation is increasing the productivity, it will not affect the population equally rather it will lead to a greater level of inequality. In fact, laid-off workers will encounter problems in finding a new occupation and unemployment is one of the major concerns of the adoption of AI (Kaplan and Haenlein, 2020). As stated by Friedman (2012), the era of the average man is over meaning that average knowledge are not enough for keeping a job or for being hired. However, automation can improve the living conditions of the citizens only in the long-run if there is a shift from a labor/wage based economy (Fagella, 2016).

Instead, in the health care system it is already possible to recognize the benefits derived by the introduction of AI. Not only algorithms are more precise than the human judgement in detecting diseases, but they are also capable of diagnosing them in their early stages. The University of Stanford has developed a system that allows the detection of any lung disease by taking a picture of their x ray. As the output, the program shows to the user the list of potential diseases with their probability. After few tests, the algorithm is considered more precise than doctors' judgement. Another interesting case is linked with a pre diagnosis of Parkinson through an analysis of small variation in the vocal cords, not detectible by humans. With this algorithm, it is possible to start earlier with the treatment achieving greater results with the cure. The main drawback in the development and adoption of technologies like the one for Parkinson concerns the management of personal data. Nowadays there is the need of specific regulation aimed at increasing the protection of the privacy to make sure for example that the device is just listening to the change in the voice chord and not storing relevant information about the user (DwDocumetary, 2019).

After discussing about the economic and health implication on the adoption of AI, it is relevant to analyze also its environmental implications since sustainability has become a relevant public concern. On one hand, AI tools will contribute to improve the living condition on earth. Already developed techniques concerns optimization of farming

processes to reduce water waste and increase the efficiency of the land. Moreover, there is a faster and more accurate monitoring of the pollution in the air and the identification of its sources. One interesting example of AI fighting against pollution is the Great Pacific Garbage Patch – an autonomous boat collecting trash that is floating in the ocean (Vox Creative, 2018). On the other hand, in order to produce servers and other technologies, raw materials such as nickel and lithium are needed in huge quantity and at some point in the future those materials will no longer be available. Moreover, once a device is no longer working or it is outdated it cannot be completely recycled. A further issue concerns the storage of big data and their computations in the clouds. Indeed, their servers requires a great amount of energy to avoid overheating and thus malfunctioning (Kaplan and Haenlein, 2020).

1.3 Current and future scenario on AI

The time we are living in is the era of big data. Data are at the center of the digital economy and this has been already recognized by China. In fact, in April 2020 the government stated that data are considered as a new factor of production. The availability of data is changing the way organizations operate (World Economic Forum, 2020). Relevant for this business reconfiguration is the exponentially increase in the amount of data generated and consequently collected and analyzed. Indeed, according to a research made by Internet World Stats (2020) in July almost 4.8 billion people were active online compared to 2.4 billion in 2014 and today they represent 60 % of the global population. Instead, if we take into consideration just Europe and North America, the percentage increases up to 90% of the population. The number of active users is interesting, because the higher it is the better data collected are representative of the entire population. By just being active online an individual is generating data. For example, data can be created through the email that the user sends or receives, the tweet she posts or the pictures she likes or not on Facebook or Instagram. All the information generated by users would be useless without a proper way to store them and algorithms capable of processing them. Indeed, it would be impossible for a human being to derive any conclusion from infinite numbers of data or it would require too much time and data are constantly updating. For example, in 2019 more than 347.000 people per minute were scrolling Instagram. Considering that the way an individual is moving through the app is carrying some relevant of information,

it is almost unfeasible for human being to keep up with this huge amount of information. Instead, algorithms are able to quickly process all these data. Moreover, in a short amount of time they can identify correlations among analyzed data. From a specific dataset, algorithms are capable of making reliable predictions and increasing the value of storing relevant data (Martin, 2018). Therefore, without a well-functioning algorithm that meets all the necessary requirements there is almost no value in data. However, once companies are able to develop those type of algorithms, data will acquire a deeper meaning, indeed the Economist (2017) refers to data as the new oil.

By investing in this new AI, companies are able to acquire and maintain a competitive advantage thanks to different benefits. The first one is connected to a change in the advertisements. In fact, by acquiring more information of the user there is the possibility to set up a more personalized and targeted marketing campaign. By doing so, the company can invest less in marketing or by investing the same amount of resources it is able to boost sales. Hence, algorithms can be used to increase the effectiveness of marketing campaign by investing in the right new channels or in a specific group of people who are more likely to be convinced in trying the product. Another interesting feature of personalization of the advertisement is that customers are more satisfied. The second benefit derived by implementing AI is connected with the creation of new products. Indeed, by knowing better the preferences of the customers it is easier to understand what are the future needs and desires of the market and then transform them into new features or new products (Stormon, 2019).

When evaluating the potential of AI and algorithms for organizations such as automation or better understanding of customers, everything seems perfect. However, by taking a different point of view it can be seen that digital economy has also negative aspects. First of all, there is a concentration of power to just few big high tech firms such as Amazon, Facebook, Google and Microsoft. Policy makers all around the world are worried about this oligopoly situation, because there are high barriers to entry in the market that are threatening competition. Moreover, these companies are able to collect a huge amount of data from a single consumer across products and markets (European Parliament, 2020). This leads to the second problem connected with the increase in inequality. Algorithms

used not only by these giants are creating discrimination among the population and in most cases they are reinforcing old stereotypes. The outputs of algorithms are not questioned, instead they are considered almost perfect. However, as we will see in the following chapter many unfair situations are the result of imperfect algorithm considered as perfect. This is one of the reason why in recent years the literature has been discussing the moral implications of AI, especially of self-learning algorithms. Nowadays, after the outbreak of Covid-19 the world is facing new hopefully-temporary constraints such as lockdown or curfew that have changed the way of living and of doing business. The pandemic had positive or negative impacts according to the industry in which a firm operates. However, independently on the sector, corporations have to deal with a more uncertain environment and in order to better forecast what are the needs of the consumers and the problems in the global supply chain, they have started to implement at a highest rate AI solutions.

There has also been a shift connected with the services offered to the final consumers. Given that during the pandemic it is better to reduce the number of in person meetings, customers started to engage more with brands online. According to a study made by Capgemini (2020), 54% of consumers have interacted with a firm through AI devices such as chat bots, digital assistance or biometric scanners.

Figure 3
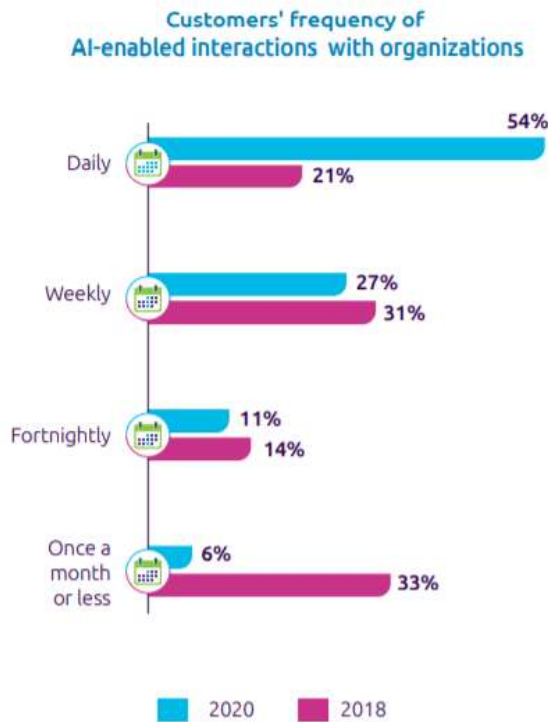
Source: Capgemini, 2020

As it can be seen from Figure 3, the interaction between an individual and an AI technology is mostly daily or weekly and there has been a radical shift compared to the situation in 2018 where almost half of the consumers had seldom interacted with them. Moreover, through the years those interactions have become more human like (from 48% in 2018 to 64% in 2020) increasing the customers' satisfactions (Capgemini, 2020).

At the same time, AI is helpful to constrain the spread of Covid-19 and in finding a cure. In fact, thanks to significant data collected by the Italian Civil Protection it is possible to develop complex model that are able to detect with accuracy the dynamic of the pandemic. Moreover, it is possible to determine the future shape of its spread curve. This has been helpful for policy makers to run simulations for different scenarios in order to avoid the collapse of national health care system. (Pham et al., 2020). A common solution to limit the spread of the virus has been the development of a tracing app. Thanks to big data combined with machine learning technologies, it is possible to send messages to a user that had been in contact with a person who tested positive to Covid-19. Even though the idea is the same in several countries it has been carried out in different ways. For example,

the Italian app *Immuni* is not compulsory and the government has faced some issues in obtaining a large amount of download that is crucial for a proper functioning of the app. Instead, China to overcome this problem had forces its citizens to download it and to update it every day. Without the app showing the negativity of an individual, he/she cannot enter into shops or public transportations. This way of tracking people allows China to shortening the recover from the pandemic and so far the country has not faced a second wave as it is happening in Italy and in the rest of Europe. Moreover, the Chinese government has used facial recognition camera not only to monitor people and checking whether they were respecting the lockdown, but also to screen people and secretly checking their temperatures (Gaglio and Vitabile, 2020). China is creating a mass surveillance system that is efficient for the pandemic but it drastically reduces the freedom and right of privacy of its citizens. For these reasons, the Chinese model is not replicable in the European and American countries where the privacy concerns are becoming more popular. This partially explains why Italian citizens avoid the use of *Immuni*, because more mainstream applications such as social media are been downloaded without taking into consideration this privacy concerned.

Almost a year after the start of the pandemic, it is now possible to start analyzing what have been the results of those tracking apps. For example, *Immuni* has not taken off – the download rate has never reached the necessary threshold to be useful. Singapore to overcome the same problem with the app TraceTogether announced that in an upcoming future it would be essential to shop or to work. In this way the government was able to increase the download up to cover 80% of the overall population. However, privacy concerns were always present and it came out that citizens' apprehension was justified. In January 2021, the government has announced that data that initially was just used for Covid-19 tracking is now provided to the police for criminal investigations. This change in the way data are used has two major consequences. The first one is connected with a diminishing trust in the government from the citizens and it is linked with inside boycott in the use of the app. The second consequence is that other countries can imitate the behavior of Singapore threatening freedom of more people (Tarabay, 2021).

Among all the different strategies adopted to avoid or slow the transmission of Covid-19, Vietnam's approach has stood out. In Vietnam, without the development and the imposition in the usage of a tracing app, the infection rate has been quite low. This has been mainly due to a well managed health care system that is the result of the lessons that the nation has learned from previous epidemic such as SARS and avian influenza that have been experienced during the first decade of XX century. What made a difference in managing the pandemic is the massive and comprehensive testing system even in the rural areas. If an individual resulted positive, he/she would have been placed in quarantine in a governmental structure, hence avoiding additional spread at home. At the same time, people that have directly or indirectly been in contact up to the third grade with the one tested positive had to be tested. In this way, Vietnam without gathering and relying on sensitive private data and sophisticated technology has been a positive example on how to manage Covid-19 outbreaks (Luong et al., 2021).

To sum up, AI is still a growing business with a lot of potential in different industries, for example there is space for improvement in efficiency through automation taking into consideration a business point of view. Moreover, AI is proved to be helpful in implementing sustainable environmental changes. Therefore, several are the positive consequences for an increase in adoption of AI in our daily life, but there has been a limited number of researches regarding the impact that those technologies have to the society (Duan et al., 2019). Nowadays, AI is so powerful that is the main tool to stop the spread of Covid-19. However, at the same time its application in different sectors is generating a new class of accidents and abuses especially at the expense of already marginalized communities. Most of the times the problem is not perceived as such due to an asymmetry of information and lack of transparency on how those AI tools operates (Galston, 2018). The aim of this thesis is to analyze the currently ethical concerns derived by the usage of AI tools, especially regarding the implication of recommender algorithms.

CHAPTER 2

In our daily life, the use of technology is almost inevitable. As it has been described in the first chapter, more and more companies are implementing AI tools for those activities directly linked with customers. In addition, in the last years, there has been an increase of interaction even in private domain, for example there has been a rise in the usage of digital assistance such as Google home and Alexa. According to Verbbek (2014) besides facilitating some aspects of our life, AI enabling tools are also playing a relevant role in our perception of morality. Given their influence on our decisions, it is important to take into consideration the way these technologies are build, moral and ethical principles reflecting human one should be embedded in the design process of those AI tools.

Ethics is a moral principle that guide decision by establishing what is right or wrong. The definition of what is ethical is not the same to what is legal. In fact, there could be cases in which a behavior is considered as legal but it is morally wrong or unethical. Ethics usually is a concept related to human behavior. Nevertheless, given the increase role played by AI tools ethical concerns are also applied to them. The ethical dilemma is not a new debate but rather a topic that has been discussed for decades. In fact, Remenyi and Williams in the paper *"Ethic and research into the silicon brain"* (1996) were already debating the ethical aspects on developing powerful processors that inevitably would have led to several social damages. During the following years the debate about the topic has enlarged and thus scholars were even focused on some specific ethical issues. The current debate on ethics can be described by a framework developed by Mittelstadt et al. (2020) and summarized in Figure 4**.**

Figure 4
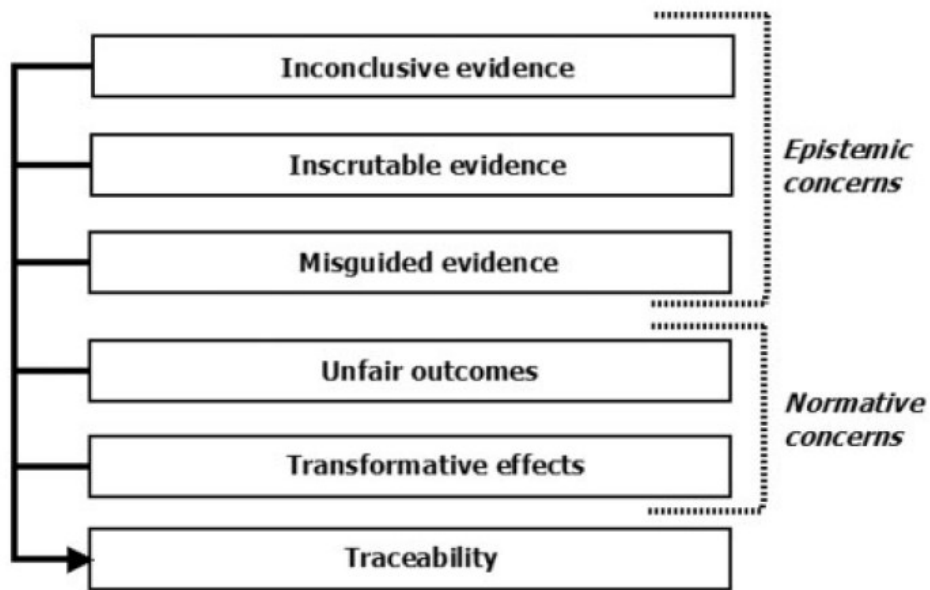
Source: Mittelstadt et al., 2020

The framework describes the ethical dilemmas concerning algorithms that are programmed to derived conclusions from a set of data and then from that output some actions will be executed. The major ethical issues can be divided into epistemic and normative concerns. In the first cluster problems are related to the reason behind the rise of discrimination. For epistemic factors, input data plays a relevant role, given that the quality of output of algorithms depends on the type of data used and their accuracy. Indeed, the normative concerns deals with the ethical consequences of implementing autonomous algorithms, such as privacy concerns and the autonomy of end users.

2.1 Inconclusive evidence

Given the ability of algorithms to process huge amount of data, associations and correlations between different variables on the dataset are easy to find. Causality is not established by designers but it is left to the machine. This is mainly due to the belief that a sufficient amount of data is enough to provide credible evidence and reliable correlations. However, high amount of data is not enough. In fact, strong correlation among variables can be true when analyzing a population but this is not significant when applied to a single individual.

When analyzing a dataset, it is important to remember that the way data are acquired poses some limitations in deriving correlations. Moreover, the output derived from a dataset is dependent on the assumptions used to collect those data. Indeed, using two different datasets may lead to different answers and when using poor quality data sets, algorithms can generate inconclusive correlations. An example of inconclusive correlation is the website "Spurious correlation". This web page is a collection of strong correlations between two random variables. As shown in Figure 5, there is a correlation of 99.79% between the US spending on science, space and technology and the number of suicides by hanging, strangulation and suffocation (Vigen, ND).

Figure 5
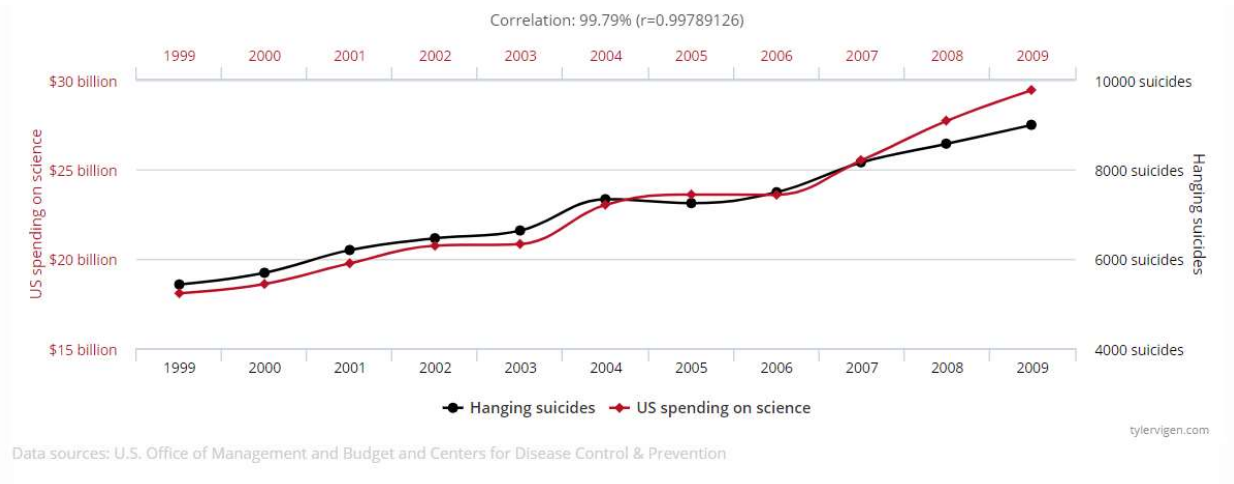
Source: Vigen T., ND

This demonstrates that a strong correlation with two variables is not sufficient to establish causality in a data set. Therefore, a machine can not operate on its own but it needs human guidelines. For this reason, correlations obtained by an algorithm cannot be directly used to support human decisions, there is the need of checking whether those correlations are representative of the reality or not. Moreover, there is the need of checking also what are the social implications of using a certain algorithms. Without taking into consideration the impact that AI has on daily life, it is possible that algorithm through its derived correlation could generate social issues (Tsamados et al., 2020). For instance, Amazon was promoting the one-day delivery in some neighborhoods in big

cities to incentivize consumers to buy online rather than in physical shops. The decision whether it was possible or not to deliver the order in the same day depended on many variables, one in particular was the closeness to one Amazon's warehouse. The company developed an algorithm that was deciding the feasibility of one-day delivery and its output was directly implemented. Thus, according to user's ZIP code, one-day delivery was available or not. According to a Bloomberg analysis (Ingold D. and Soper S., 2016), the algorithm was generating distort outcomes or it is better to say was discriminating minorities' communities. In fact, in those neighborhoods, the fast delivery was not available but the decision was not linked to the distance from the warehouse. The most striking case takes place in Boston in which the one-day delivery was feasible everywhere but in Roxbury. However, as shown in Figure 6 this community was surrounded by places where Amazon was delivering in one day.

Figure 6

Source: Ingold D. and Soper S., 2016

After the Bloomberg research went public, Amazon defended itself by stating that their algorithm was also taking into consideration the concentration of Prime account in one area and if the number was sufficiently high the one-day delivery was available. Even though this decision was made taking into consideration a cost and benefit analysis, the

outcome of the algorithm was reinforcing inequality in accessing to some services. Five days after the publication of this research, Boston's Major through a discussion with Amazon was able to extend the same day delivery to every neighborhood of the city (Ingold and Soper, 2016).

Unfortunately, discriminating situations coming from the implementation of the algorithms without proper checks do not apply only to previously discussed Amazon case. Nowadays, checking the way an algorithm is working prior its disclosure to the public is not common. This is mainly due to a complete trust on the efficiency of the operations performed by algorithms, but as we have seen there is the need to carefully evaluate what are the consequences of the adoption of a particular algorithm.

2.2 Inscrutable evidence

In order to control and evaluate the decision making process of AI tools there is the need to understand how it works and this is not easy when talking to machine learning algorithms. In fact, they are called machine learning because they have the ability of self modifying their own structure by creating or changing governing rules and this happens every time a new dataset is analyzed. Given that the older the algorithm is the more governing rules are deviating from the original and programmed structure, it is hard even for programmers to understand their decision making process. Therefore, even when they want to check on the fairness of the output, they would have to know the process that most of the time is unclear and the check would be needed every time the algorithm changes its rules. In addition, algorithms are easily malleable and they can be updated and modified by programmer whenever it is needed. On one hand, this allows to quickly correct errors but on the other it creates noise in the structure and thus decreasing the transparency. Lack of transparency increases the need of human to trust the algorithm because it is difficult to read and check the governing rules of both self-learning and modified algorithms. Nevertheless, as we have seen in the previous paragraph human cannot completely trust the outcome of algorithms, given that strong correlations may not imply causality.

Lack of transparency is reinforced by the owners of algorithms. In fact, companies do not want to disclose information about them for two reasons. First, most of the algorithms are developed by profit-oriented firms and they can provide a competitive advantage. They rather prefer to leave codes, input and output of algorithm in a black box (Hosanagar and Jair, 2018) hence they are not replicable by other organizations. The second reason is that building a transparent algorithm is complicated and time consuming. Moreover, even with a transparent algorithm common end users will still not be able to understand how it works because this requires above the average technical knowledge. Therefore, having completely transparent algorithm would not work as a unique solution regarding issues arising with the use of machine-learning algorithms. Instead, testing the behavior of the algorithm before exposing it to the end consumers could be the answer. During the test, it is possible to understand whether the algorithm shows negative tendencies such as unfair discrimination (Tsamodos et al., 2020). For example, in 2016 Microsoft has launched a chatbot on Twitter called TayTweets. The aim of this AI user was to carry out normal conversations by emulating the way millennials talk to each other. The more users were interacting with AI, the smarter it will get. At the beginning, the experiment was successful, but soon users took advantage of the situation. Indeed, they started teaching to Tay sexist and racist messages. In its first 24 hours, it has written offensive tweets such as "Bush did 9/11 and Hitler would have done a better job than the monkey we have now. Donald Trump is the only hope we've got." and some tweets against women. Therefore, Microsoft had to stop the experiment and block the online interaction (Hunt, 2016; Benfatto, 2016). One approach to avoid this type of conduct is to test the software before exposing to the public. A company can use for instance Glassbox, a test developed by Deloitte, to detect unethical behaviors in the structure of AI algorithms before its launch.

To sum up, transparency is not a term easily applicable to self-learning algorithms because machines are continuously changing the way they operates. At the same time, lacking of transparency is wanted by companies owning the algorithm to prevent the disclosure of information since they can be a source of competitive advantage. Lack of transparency is indirectly linked with unfair algorithm. In fact, it prevents internal or external analysis on its behavior, so it is more difficult to detect unfair outcomes. One way to overcome the problem is by testing them.

2.3 Misguided evidence

Another concern regarding the ethical issues generated by AI is linked with bias. In this case, algorithms are not providing explicit unethical outcome like Amazon's or Tay's case studies but still they are generating discriminating outcomes. Different are the sources of bias. First of all, an algorithm is reflecting the value of the people building it or the culture of the company. Usually when developing algorithms, programmers are mainly focused on achieving their goals and usually putting ethical concerns aside. This type of bias is classified as social value bias. Most of the time it is an unintentional bias. In fact, the machine has the same way of thinking of its programmers and if they belong to the same social group ethical issues may arise. The second one is connected with technological constraints deriving from errors or design decisions leading to discriminating outcome. In this regard even the trained data can lead to biased decisions and this represent the major driver in bias algorithm. For example, when training an algorithm to evaluate resume for a specific position it is necessary to build a data set from previously hired people. Information such as experience, qualification, type of education are important in the evaluation process. Nevertheless, if the position has been held only by white male, the algorithm will interpret it almost as a necessary condition for new candidates rather than noise. The training algorithm will end up discriminating against women and non-white men. Unfortunately, this type of algorithms is used even by big companies like Amazon. Hence, developers should carefully deciding which dataset to use in training algorithms (Praharaj, 2020). Another example is represented by ImageNet, a broad dataset containing more than 14 million labelled pictures, used to train computer vision programs. The major issue is in the composition of image that it is not representative of the reality. Indeed 45% of images belong to US when in reality the country counts for just 4% of the overall population, in contrast just 3% of photos are describing Indian and Chinese population that represent 36% of the world. The way computer vision programs are trained influences their performance and it has been show that they failed to recognize non-white individuals (Du and Xie, 2020). Despite this, they are largely used to help US police in dealing with crime and generating misguided evidence.

One solution to overcome the problem of rising bias in the training process could be to use synthetic data meaning that they do not come from reality but they are generated

through AI. Nowadays, generative adversarial networks are able to provide a diversified dataset in terms of gender and race. By feeding algorithms with such dataset, biases do not arise in the training session (Tsamodos, 2020). However, it does not prevent bias arising in the initial phase done by the programmers. Moreover, bias can also arise when algorithm developed for a specific aim or context are used in other situation. Therefore, when implementing an already existing algorithm is essential to understand whether the change in the context is compromising its ability to function without generating biases. The control on existing bias is crucial even on new algorithms.

2.4 Unfair outcomes

In order to prevent direct or indirect discrimination, fairness should be taken into consideration when building an algorithm. However, fairness is not an easy concept and there is no agreed definition on it or a standard way to measure it. For instance, algorithm can be fair if it does not discriminate against minority. Nevertheless, when defining fairness the developer need to choose their focus – it can be a single user or a group. Generally, algorithms are gathering information about a specific group and then tends to generalize user's behavior. In doing so, even when discriminating data are left out from the analysis such as gender or race, unfair outcomes can arise by profiling people according to their ZIP code, as in the already discussed one-day delivery Amazon case study. Hence, using apparently neutral features the algorithm, if it is not controlled, could still generate discrimination.

Another issue related to fairness is personalization meaning targeting just a segment of the population for a specific purpose like a job offer or a news. Fairness is defined according to a specific user at the expenses of other stakeholders. Nevertheless, algorithms should display fair behavior for all the agents using the platform hence for example, Facebook the algorithm should be fair to the company advertising on the platform and to individuals. When measuring fairness it is also important to avoid local optimum solution. When evaluating the outcome of the algorithm, fairness should not just reflex how the society is today, but rather how the society should look like. Hence, avoiding optimizing around existing economic and social discriminating dynamics.

When talking about fair algorithm there is an important trade off to take into consideration which is accuracy. In fact, it is almost impossible to develop an algorithm that minimize unfairness and at the same time it provides the lowest overall error. However, one way to deal with this problem is by undertaking the optimal solution, described by Kearens and Roth (2020) in their book, T*he ethical algorithm* and described in Figure 7.

Figure 7
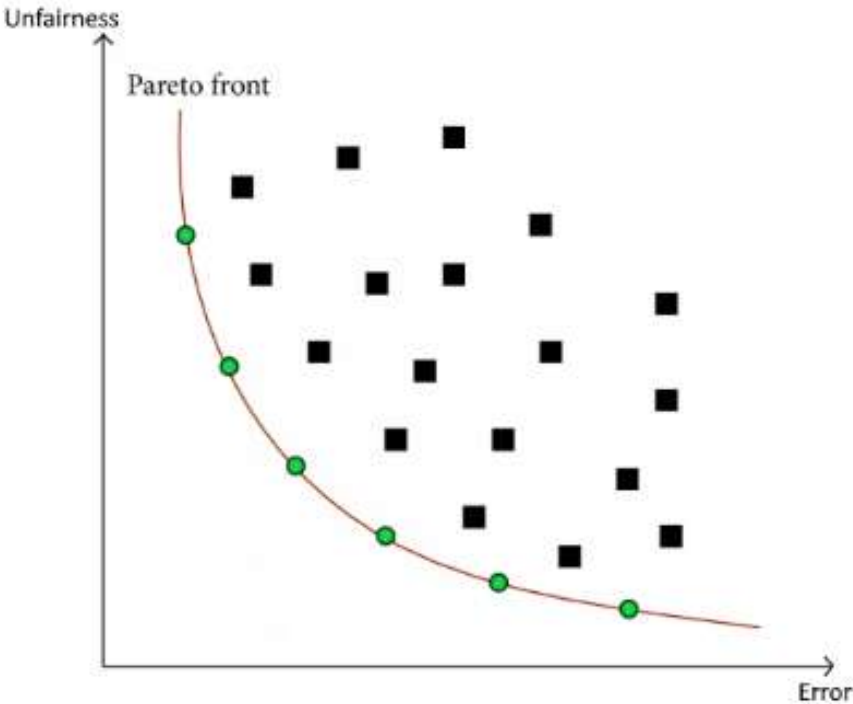
Source: Kearens and Roth (2020)

As we can see each dot and square represents a model that has x value of error and y value of unfairness. By representing all the available models build and use for the same purpose, it is possible to compare them all together. All the dots connected with a line represent the Pareto frontier, which is the set of all algorithm that better define the trade-off between accuracy and fairness. However, the optimal solution depends on the judgement of the individual. There is no right answer since fairness is a subjective perception.

From the next section on, recommender systems are used as the underline example in explaining the moral and ethical implication of AI, regarding the normative concerns. The

main reason is that this type of algorithms, as we will see, is able to nudge the way individual thinks and acts. Thus, this technology is capable of shaping our future.

2.5 Recommender system

There is no common definition of a recommender system, but rather it has a different meaning whether we analyze it according to computer science or marketing point of view. For the former it is a way to provide users with recommendations based on the behavior of the community. By looking at it by a marketing or more in general by a business point of view, recommender system is a way to increase sales and attracting more customers (Rezaeina, 2016). At the beginning, when world wide web started to be available to the private consumption, recommender system was just generating valuable suggestions for the user and it was highly appreciated (Konstant, 2004). In fact, it was focused on predicting the likelihood of the purchase of item x by having purchased a product of the same category of item x. However, this type of algorithm was suggesting items that the user would have bought anyways and it did not bring any additional sale for the company using it. Therefore, recommender system needed to shift toward providing more broad recommendations.

In the following years, machine-learning techniques previously discussed started to be applied to recommender systems in order to improve the precision of those suggestions by taking advantages of the increasing availability of data about the single users and about its targeted group. Nowadays, the use of recommender system is vital for online businesses and companies are building their competitive advantage over a precise recommender system (Rezaeina, 2016). For example, 80% of the total hours of streamed video on Netflix depend on its recommender system that it is worth US$1billion per year (Du and Xie, 2020).

Figure 8
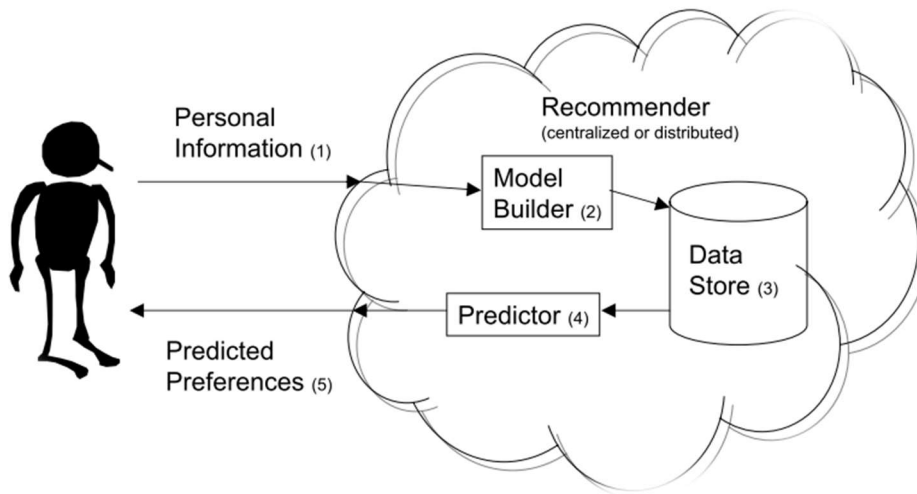
Source: Lam S. K. et al., (2006)

A brief summary on how a recommender system functions is represented in Figure 8. In reality it is more complex, recommender systems can be clustered into five categories based on the way suggestions are made. The first group is composed by the content-base and it represents the oldest type of recommender system, but it is still the more efficient. This type of system makes suggestions by analyzing user's historical interactions independently of his/her direct or indirect feedback. Nevertheless, in order to provide a more precise recommendation, it needs a huge amount of data that usually is not available for a new customers or a new products. The second one, represented by collaborative filtering, combines the rating of several users to generate a suggestion. The system collects relevant information from other individuals that are similar to the targeted user and then it provides the recommendation. The third bundle is composed of knowledge-based filtering. In this case, recommendations revolve around just on the individual, but instead of using historical data, it depends on external previously specified knowledge. The forth is represented by demographic filtering. The underlying assumption is that individual with common features such as sex, age and country share the same preferences. In the last group there is hybrid filtering in which two or more filtering are combined together (Karlsen and Andersen, 2019).

Therefore, with the incorporation of machine learning and deep learning techniques to recommender system, it is possible to carry out nonlinear transformation of the data and thus capturing more complex patterns and better predict user's preferences. Moreover, it reduces the need of a complex hand-written code and the algorithm can learn from different sources of information – texts, images, video and audio. Another advantage of deep learning is that it increases also the flexibility of the process.

Recommender systems have also a positive impact for the end user. Indeed, it has facilitated surfing the net, providing that individuals when using internet have due major constraints: bounded rationality and limited amount of time. Concerning the former, an infinite amount of information, posts, photo or videos is uploaded daily for example, every second there are on Instagram 995 new posts (Omnicore agency, 2020). An individual does not have the cognitive ability to process all of these information and filtering them according to their interests increases their satisfaction and thus utility. For example, recommendation system applied to new feeds avoids the user to read the complete newspaper. Indeed, it rearranges the order of the article according to individual's preference. Thus, in this case news provider website, but more in general companies are competing with each other in providing the best recommendation service to attract the highest number of users. More users equals to more data to feed the algorithm and thus better recommendations, it is a virtuous circle. However, companies are not only competing with firms in their industries news provider with news providers and streaming platform with streaming platform. Instead, organizations are competing against each other to conquer human attention – it is their second constraint. In fact, the main competitors of Netflix are YouTube, Facebook and sleep (Tristan Harries, 2017).

2.6 Business model

Human attention is so important that companies are fighting with each other in order to obtain the highest share of it. In order to do so organizations are using persuading techniques such as auto play, basically, at the end of a video there is already a new one playing that is chosen by the algorithms. This feature, initially developed by Facebook, has been proven to increase the time spent in one website/app and consequentially has been

adopted even by other companies such as YouTube, Instagram and Netflix (Tristan Harries, 2017).

This interest manifested by companies to increase the attention of users towards their website is connected to their business model, in some cases the more the user is using the website/app the more the firm is generating profit. Indeed, there are two groups of company: on one hand companies such as Netflix are requiring a small monthly fee in order to have access to their services. On the other one, there are organizations providing a free service to their users and this is the case of all social networks. In both groups profitability is connected to the number of users. In the first case, the higher the members the higher the profit. Therefore, the company is willing to retain and attract as much individual as possible. Instead, in the second case, the higher the number of active user the higher is the value and the potential gain of the company. Despite being free, social networks are generating profit.

According to Zuckerberg, founder of Facebook, the platform is free to allow every individual, regardless of his/her income to be a part of the community. However, the main reason why users are not paying a monthly fee is that price that they are willing to pay for Facebook is way lower than the price paid by advertising companies to invade users' privacy. Indeed, less than 10 % of today users are willing to pay for a service with no advertisement, so charging a fee on users will decrease dramatically the online community and thus the value of the social network (Sherman, 2018). As seen so far, for social network the main source of revenues is represented by advertisement. Among all social networks, despite being the pioneer, Facebook has still the highest growth and revenues; currently its leading position is threatened mainly by Tiktok – a Chinese social network. Facebook is generating money mainly through two different forms of advertisement that counts for 85% of total revenues allowing the company to become the leading marketing and advertising platform (Iqbal, 2020). The first one is self-serve advertising in which the ads are seen by users in the right side of the feed and it can be created in an autonomous way. The second one is targeted advertising, more effective than the first one because it creates higher engagement. Companies are able to target their potential customers in a precise way. In fact, Facebook is holding huge amount of data for

every user such as age, gender, location, her passions, political opinions etc. Therefore, firms can pick the relevant characteristics of users seeing the ads generating a higher return compared to a marketing campaign run in television (O'Connel, 2018).
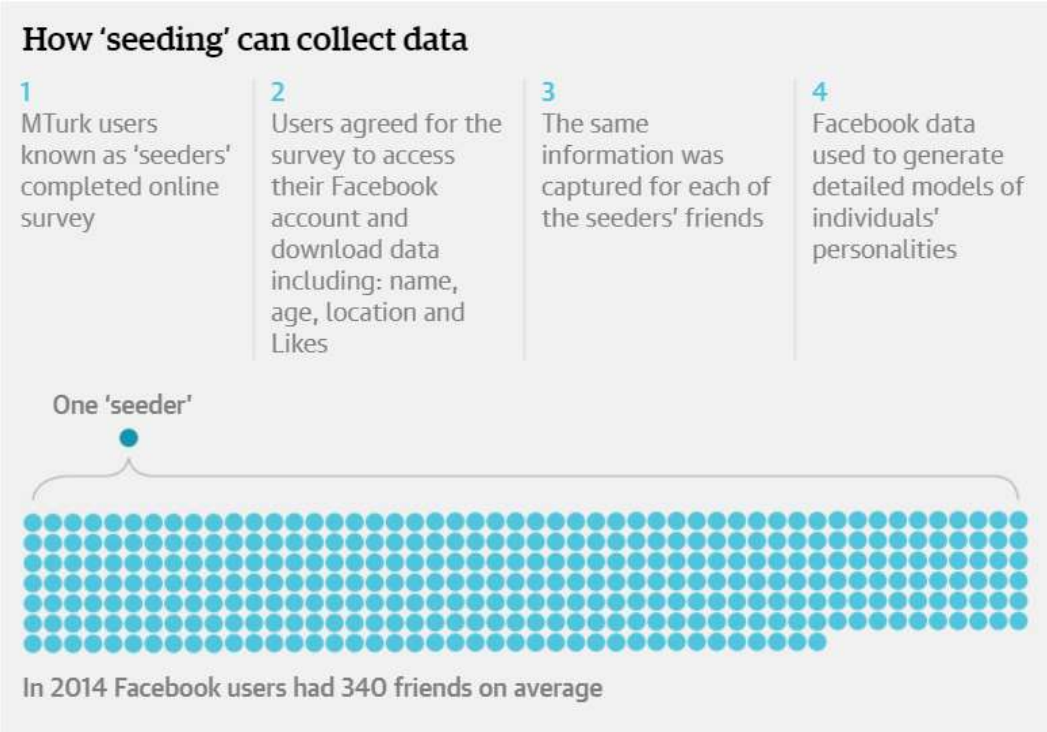
Besides advertising, the other source of revenues of Facebook is linked with data. According to Zuboff (2015), this new business model is connected with surveillance capitalism and the humanity is facing "*A new economic order that claims human experience as a free raw material for hidden commercial practices of extraction, prediction, and sales*". It is known that Facebook has been sharing users' sensitive data such as name and private chats with other firms such as Netflix and Amazon in exchange of their data. With the transfer of knowledge, Facebook was able to be connected with every other relevant applications and it was promoting this feature to its users as a way to increase their satisfaction. However, the main reason of the exchange of data was that Facebook was increasing its online presence and thus building its empire at the expenses of users and their privacy. This was its *modus operandi* until Cambridge analytical scandal (Madrigal, 2018).

2.7 Cambridge Analytica

Cambridge Analytica is a subsidiary of Strategic communications laboratories, the organization that introduced and implemented the idea of using data coming from social media to determine the personal trait of individuals in 2014. Thus, through its algorithm Cambridge Analytica was capable of clustering people according to their habits and behaviors. In this way, it is possible to develop and target every individual with the most efficient advertisement. Compared to other companies that are able to leverage the user according to their tastes, Cambridge Analytica can do it even with their emotions. Their algorithm is so precised and fed up with data that the company needs just to know 70 likes an individual has put on Facebook to understand his/her personality. In addition, with only 150 likes the algorithm has a broader understanding of the individual compared to his/her family (Manietti, 2018). However, the main problem for the company at the beginning was to find a sufficient number of data to build and train the algorithm. To overcome this issue, Cambridge Analytica has developed an app called "this is your digital life" in which users were paid a small amount of money between $ 2-4 to fill a

questionnaire allowing both of the parties to understand what was their psychological trait. Users were connecting to the app through their Facebook account to save time, since they could avoid the creation of a new account. At the same time by connecting with Facebook, they were authorizing the app to acquire relevant information about them and all of their friends – nowadays Facebook banned the possibility of acquiring information about the users' connection. Considering that on average a user has 340 connection, just few people taking this test were allowing Cambridge Analytica to store a huge amount of data. A sum up of on how data are collected is described in Figure 9.

Figure 9

How 'seeding' can collect data

1 MTurk users known as 'seeders' completed online survey

2 Users agreed for the survey to access their Facebook account and download data including: name, age, location and Likes

3 The same information was captured for each of the seeders' friends

4 Facebook data used to generate detailed models of individuals' personalities

One 'seeder'

In 2014 Facebook users had 340 friends on average

Source: Davies H., 2015

With data obtained through the online questionnaire, Cambridge Analytica has generated 253 types of personality following the OCEAN scale, meaning that it takes into consideration Openness, Consciousness, Extraversion, Agreeableness and Neuroticism (Davies, 2015). Given their precise targeting service, the company has been chosen by politicians in order to improve the effectiveness of their political campaign for the 2016 presidential election. For instance, Trump's committee has delegated the firm to acquire and analyze data about consumers. The aim of the marketing campaign was to target

persuadable individuals and convince them to vote for the conservative candidate rather than the liberal one. Among all the persuadable, it was essential to catch the attention of people of certain district in the *swing states*, meaning the one that can change the result of the election, allowing Trump more chances to win the election.

The strategy adopted to change the mind of voters was the following. Voters were boomed with personalized contents on all the platforms that they were using, until it was sure they would end up voting for Trump. The selected content depended on the targeted group each individual belong to, the common feature was not to promote Trump's believes but rather to discredit Clinton (Hern, 2018). For example, it was common for black individuals to come across a quote of Clinton dating from 1996 in which she defined African-American as "superpredators". Instead, white female were often seeing articles stating that Clinton's husband has no respect for women. In the end, the online promotion was efficient, individuals were persuaded in favor of the conservative thus Trump was elected the new president of the US. From this presidential campaign, two are the important aspects that have emerged: the power of social media to nudge people and the way private data are traded online. Both implications represent two of the most discussed ethical issues in the literature but also in public debate.

2.8 Nudge Theory

To win the election Trump's online strategy was not focused just on showing real facts that were discrediting his opponent rather he was carrying on an unfair campaign. In fact, to support him, huge quantity of fake social media accounts were created and automatically managed. Those accounts using popular names and friendly pictures had the aim of spreading fake news, pictures or comments in order to enlarge the debate regarding a specific topic. The activity of those fake accounts was carried out for the entire length of the campaign, it was most relevant during public event of Trump such as debates among the candidates – the most awaited events of the year (Manietti, 2018). During those debates, it is common that viewers read other opinions online and with no doubt they would end up reading some post of those accounts containing even fake news. The aim was to convince and manipulate people to vote for Trump no matter what, even if this means playing unethically.

What has been done by Cambridge Analytica to support Trump it can be described by the Nudge Theory developed by Richard Thaler that granted him the Nobel Prize in 2017. In the book *Nudge improving decisions about Health, Wealth and Happiness* (2008), Thaler describes nudge as "*any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives*". Moreover, as explained by Hansen (2016), it is considered as nudge any attempt aimed at manipulating the way people think or behave and this can happen because of cognitive biases, habits and routines embedded in any individual and in the society. Hence, according to Thaler's studies, individuals can be influenced because they do not behave as the Homo economicus – individual used as a point of reference to develop all major economic theories – but rather as what he defined as Homo sapiens. The latter has bounded rationality, meaning that it is not able to process all the information available in the world to make evaluation of all the possible alternatives before making his/her decisions. For example, it is common that users stick to default options no matter what they are. Therefore, it is possible to obtain the consent of individuals by imposing a certain condition as default option. Another relevant difference between Homo sapiens and Homo economicus is the distorted perception of reality – people tend to be way more optimistic. It has been proved that it is possible to influence the way individuals think in specific situations by recommending a starting point for the thought process. As a matter of fact when guessing a number such as the value of an object or its weight, the hypothesis of an individual is going to adjust to the hypothesis stated out loud of other people – in this case the default option.

However, in order to be capable of influencing a thought or a behavior it is necessary to understand how the human brain works. Human can use two different types of thinking, the first one is called Automatic System and it functions when the situation is intuitive and automatic. It has been shown that voters on average rely just on the automatic system and first impressions of a candidates combined with their way of arguing are the most important determinant in choosing to support them or not. Instead, the second one, Reflective System, is triggered when there is the need for reflection and rationality and it is considered as more self-conscious. Therefore, when trying to build a nudge system is

important to understand what are the best strategies to obtain the desired outcome taking into consideration the system that needs to be triggered (Thaler, 2008).

Another factor relevant in building an efficient nudge is determined by the social context in which the user is making a decision, especially regarding the type of available information available in that moment and the peer pressure. Through experiments, it has been shown that an individual when he/she is with other people is more likely to reply to questions in the same manner as the other, independently whether they provide the correct answers or not. The situation does not apply only to tough question. Hence, it is even possible to nudge people to state something crazy as long as other people do so. This can explain the reason why Trump administrations has been using fake accounts. For an individual it is hard to distinguish whether behind an account there is a real person or not. Nevertheless, seeing hundreds of tweets or posts supporting Trumps providing similar evidence can nudge the individual. For example, in one experiment described in Thaler's book, individuals were asked whether they agree or not to this statement: "Free speech being a privilege rather than a right, it is proper for a society to suspend free speech when it feels threatened." In the control group 19% of them agreed to the assirtation, in the other group with 4 people were showing their support for the quote, the overall percentage of people agreeing was much higher, 59%. However, this does not reflect the initial purpose of the studies conducted by Thaler. Nudge theory has been developed as a tool to promote a healthier and more sustainable lifestyle both by public and private institutions. For instance, concerning tobacco the aim of the government is to reduce its consumption to decrease the probability individuals will suffer from smoking related issues thus increasing their lifespan. The most common strategy so far adopted is to inform the consumers about all the negative consequences related to smoking. Instead, by following the Nudge theory, the strategy should be focused on altering the condition in which smocking decisions are made. In any case, a nudge does not imply a restriction on the available alternatives (Alemanno, 2012).

With the development of internet and especially nowadays, in the economy of big data, it is easier to construct an environment that promotes nudges. The process required to develop a digital nudge is composed of two steps. In the first part, AI algorithm collects

for each individual all the relevant information available online such as the list of friends, which type of content she/he prefers to read or watch, what are the items that have been recently purchase etc. The collection of data does not come from a single domain but rather through cross domains. For example, Facebook is collecting information even about individuals that are not members of the platform (Reisach, 2020). By processing these information, AI algorithm are able to provide a complete description of the individual's life path. In this way, it is possible to understand whether the individual is a potential candidate for the nudge and if it can easily influenced. If the machine is considering the individual a right target then the analysis proceeds to the second step. Now, there is the need to identify what are the best ways to influence the targeted user, because there is no one size fits all approach. Indeed, what is proving successful for an individual does not mean it is efficient for everybody. That is why AI in essential, because it is able to personalize and target in different manner individuals in a short amount of time. Moreover, through feedback machine-learning algorithms can determine whether the applied nudging techniques have been efficient or not by monitoring and detecting some change in the behavior of the user toward the goal. In the latter situation, they are able to self-modify their persuading tools to improve the efficiency of the nudge. Every time the algorithm finds a better way to nudge a specific individual it modifies the initial model. Moreover, AI algorithms allow storing information about the type of techniques that have been effectively working to nudge the user in the past (Karlsen and Andersen, 2019). The whole process behind building a successful nudge is described in Figure 10. Hence, the more the algorithm is used the more precise its persuading techniques are matched with individuals in the right moment (Risdon, 2017). Usually on social networks, nudge is implemented through personalized recommendations that push the individual toward the desired behavior.

Figure 10
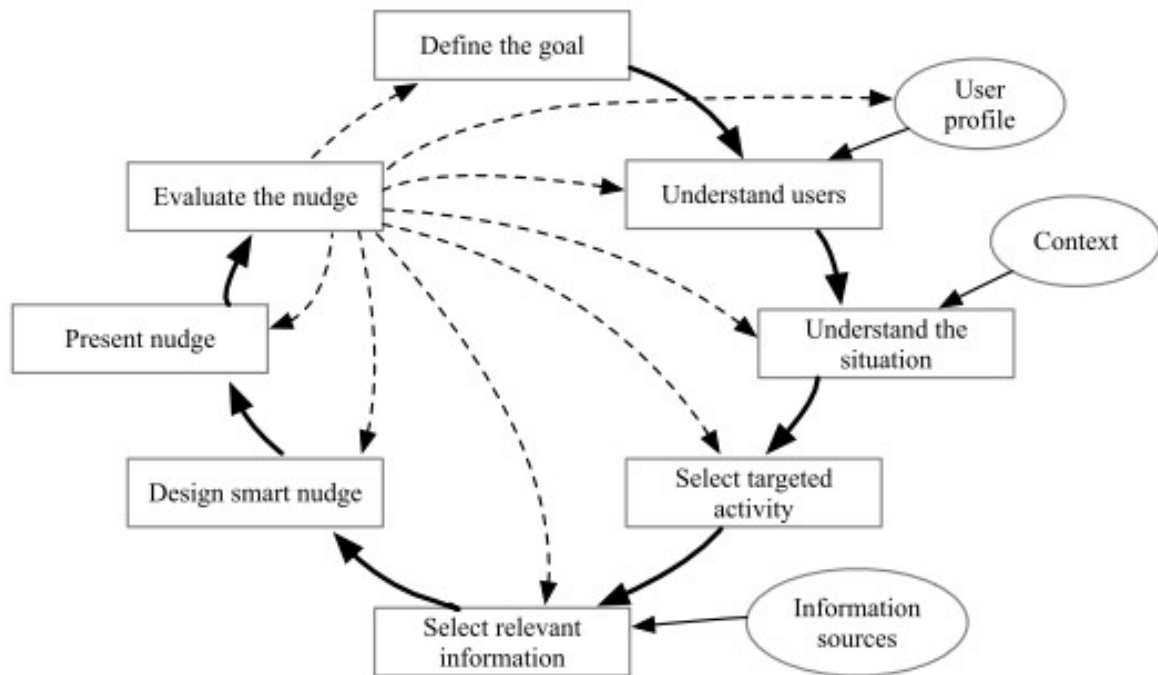
Source: Karlsen R. and Andersen A., (2019)

Different types of techniques are needed even depending on the level of education and on the age of the targeted individual. Concerning the former, most educated people requires more time and thus resources to be influenced. Therefore, low educated people have a higher risk of being easily nudge. The matter is of increasing importance especially in those countries with high rate of functional illiteracy such as Italy. In fact, there citizens suffering from functional illiteracy or relapse into it counts for 45-50 % of the population. For them, it is difficult to distinguish whether the information they are reading is real or not. Hence, given they are easily persuasible they are the perfect target for any type of nudge. According to an investigative report made by Gianni Riotta in 2010, "*Il lato oscuro della rete*" (the dark side of the net), there is a link between the literacy level and the way the individual is casting his/her vote. The process begin when the individual is nudge into believing something is real. In addition, the individual through the bubble effect will only see confirmation of the nudged though or behavior that will reinforce his/her credo. This is reinforced even more by the confirmation bias – the tendency to search information that support the individual's hypothesis. In the era of big data, it is easier to find other individuals that believe the same thing, even if it is controversial. In this stage it does not

matter whether confirmation comes from real or false evidence – that is when fake news or conspiracy theories emerge. Moreover, the targeted individual will spread those messages to his/her online and offline net (Moderato, 2017). This report based on the Italian population underlies how 2016 US election is not an isolated case but rather something that can have some relevant implication in our life no matter where we live.

Instead, regarding the age we can say that digital nudge is more efficient on the young part of the population, since their information processing behavior is the most rapid – they have the lowest attention span. Indeed, it takes less time for them to process any type of information they come across on social networks. Moreover, due to their continuously multitasking activity it is more likely that they make errors when handling information. Nevertheless, their opinion is so easy to influence that just rearranging the options or setting a specific layout is sufficient to obtain the desired outcome (Lembcke et al., 2019). The fact that it is so easy to influence people is positive only when the nudge is used as designed so to help citizens to make decision aimed at improving their life. Unfortunately, most of nudges do not belong in this category. As we will seen in the following chapter, legal systems all around the world are protecting the most vulnerable people – the children. Regardless the legal protection of minors of 13 years old, everyone regardless his/her age is constantly receiving digital nudge through their smartphone. Notifications can be used to remind to the user the desired behavior, for instance smartwatches notify the user to do some physical activity like a walk after hours of being sat. Besides promoting a good desired outcome as smartwatches encouraging to move, notifications are also used by platforms as well as social networks to promote themselves and to invite the user to login after a period of non-being active. In this way, by pushing the user to enter to their app there are more chances of increasing their profit coming from advertisement as described in the previous paragraph and they also decrease the probability of the user switching to a new platform (Lembcke et al., 2019). Notifications are one of the many tools a company can use to retain attention. In fact, corporation uses millions of calculation to find a way to tweak online experience and they are building the experience so the user wanted to have more of it. This mainly explains why people are so addicted to their smartphone and they are checking it on average every 15 minutes even though no notification has alerted them (Anderson, 2017).

In order to manipulate the mind of the online users of a platform, the AI can spread fake news initially through bot or fake accounts and then if the news is interesting or shocking it has the potential to go viral, meaning that everyone will read it somehow. Providing that social networks are becoming the primary source of news discovery, it is easier for fake news to spread and for people to believe them. This is mainly due to multitasking effect as stated before and the fact that an increasing number of individual read just the title of the article. In this way disinformation is generated. Moreover, it is easy to spread fake news given that bot accounts represents 9-15 % of all the users enrolled in social networks (Reisach, 2020). Therefore, the problem of bot and fake news used to nudge individual is not just related to Facebook and Trump's election. By 2016, social networks instead of increasing the freedom of individual and being part of revolution as it happened during Spring revolution in north Africa had become a threat to the democracy (Harris, 2018).

2.9 Privacy

The 2016 election has been remembered for the Cambridge Analytica scandal. Basically, months after the election and following an inside leakage of information there was a public disclosure about the way the company has collected data from Facebook and the way it used it to run the Trump's campaign. However, the main problem is not that hackers have been breaking the system and then trade personal data. It is rather a problem on how the social media is functioning. Indeed, Facebook has been designed to generate data about single users and to extract and exploit those data (Carrie Wong, 2019). One issues is connected with the disinformation of individuals regarding the type of data collected and the way they are used by Facebook. As described in Figure 11, users when talking about privacy they are taking into consideration posts and messages that they directly generate. For this type of record, the user is free to change in the settings who it is able to see them. Nevertheless, when talking about privacy we are also referring to the unseen part of the iceberg.

Figure 11
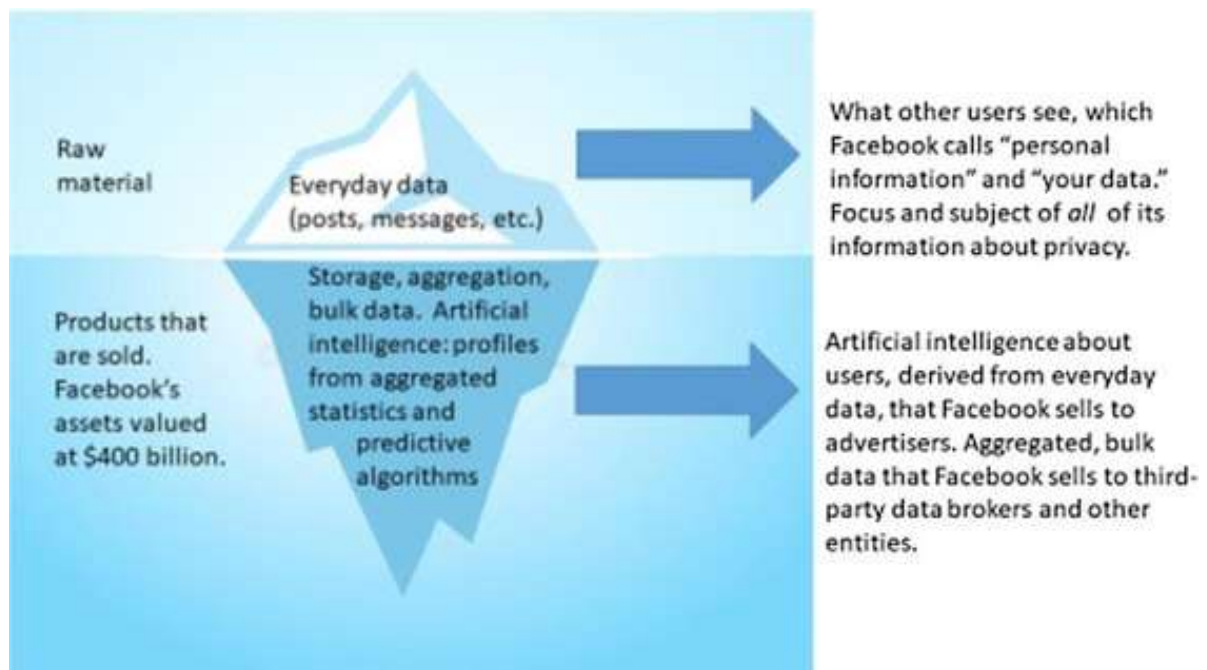
Source: White C. L. and Boatwright B., 2020

More in general, to build a personalized algorithm a huge quantity of data of the entire population and of the single user is needed. However, this can increase the risk of disclosure of personal and private information. The concept of personal data is pretty unclear and vague. According to the legal system, several are the way in which privacy is described and subsequently different are the methods used to grant protection of personal data. For example, EU and US have two different approaches when it comes to privacy protection. In the US there is not a unique federal law, but it is rather an aggregation of federal and state laws. There has been many attempts in creating a single federal law to provide to everyone the same level of protection, but none of them has been successful. Despite the legislative difference among states, a common feature is that safeguarding citizens is as important as preserving the needs of industries. It is common that lobbies, such as Privacy for America, work to ensure that any new law concerning the way private data are collected and then managed does not harm the organizations. Instead, in the EU, the rights of individuals are more important than business concerns. The way personal data is treated is subject to a single regulation (n. 2016/679), called General Data Protection Regulation (henceforth "GDPR"). The regulation defines the way personal data can be collected and the way they can be processed: wholly or partially by

automated means. More important is how the regulation defines the concept of personal data that is

> "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*"

From this is clear that a data is considered as personal when through some analysis is possible to link it to a specific individual. Therefore, personal data is not simply composed of name, surname, social security number, bank details or health records, but rather by some small details about individual preferences that are sufficient to discover his/her identity. Given this definition of personal data, even disclosing information about users' reviews of films watched on Netflix can be considered as a privacy violation. In fact, from the title of the film is possible to find the individual who has been watching it and from the way the individual has reviewed it is possible to derive sensitive information such as political or sexual orientation that an individual may does not want to disclose (Kearns and Roth, 2020). This is considered as privacy violation when the company collecting information is disclosing it to the public or is selling to another company without the consent of the user. With the ability of AI technologies to re-identify an individual from small and apparently insignificant data, individuals have more probability of encountering privacy violations through data breach. The latter is defined as the exposure of sensible data to third parties.

Another relevant aspect of GDPR lays in the article 5. It can be considered as a framework to guide organizations and individuals on how to handle personal data. The basic principles are: lawfulness, fairness, transparency, purpose and storage limitation, data minimization, accuracy, integrity/security and accountability. The last principle is the only one new compared to the principles laid down in the previous 1995 data protection directive. With the accountability principle, organizations must show they are complying

with the GDPR. In fact, companies now have to keep track of how data is collected and managed. Moreover, they have to train employees on how to protect data and on how to behave to comply with the GDPR. Therefore, with the introduction of the accountability principle, companies have to demonstrate they are doing everything in their power to safeguard the privacy of European citizens. A second implication for companies is the necessity to have an internal or external DPO. The data protection officer is an individual with juridical and computer knowledge. Her/his task is to evaluate the data process management and improve it whenever it is necessary to make sure the activities carried out in and by the company comply with the GDPR.

Nowadays, as we have seen with the EU regulation, disclosing personal information is considered illegal as well as unethical. Despite this improvement related to online privacy, there are still other activities connected to the way companies are managing privacy that are permitted by the law but wrong taking into consideration the ethical point of view. One of them is the way companies are able to collect information about every user. The most common is same-site tracking. Basically, the organization stores all the activities performed by every user in its own website or app. It is possible to achieve this through session cookies that allow the user to stay logged in and keep items in chart even if we close the window. As we can see from Figure 12, every website stores different type of data. Among all the available information, the most stored are email, name, home address and bank account details. When looking at the amount of data stored, we can see that Facebook is at the first place and indirectly in the second one since Instagram is managed by Facebook. After them, there are Tinder, Grind and Uber collecting half of available data. Despite the huge storage of data through in-site cookies, ethical issues are more related to the other methods. Among them there is cross-site tracking that allows the company to store information about the user's activities in other sites. This can be done through: traditional cookies in which the users are linked with a code and they are depositing it in every website they visit; super cookies that are similar to the traditional one but they are more difficult to erase; fingerprints that collects information about the location of users and the type of device they are using (Matsakis, 2019).

Figure 12

clario.
The companies that know most about you

#	Company	% of personal data collected
1	Facebook	70.59%
2	Instagram	58.82%
3	Tinder	55.88%
4	Grindr	52.94%
5	Uber	52.94%
6	Strava	41.18%
7	Tesco	38.24%
8	Spotify	35.29%
9	MyFitnessPal	35.29%
10	Jet2	35.29%
11	Credit Karma	32.35%
12	Lidl Plus	32.35%
13	Netflix	26.47%
14	Nike	26.47%
15	Asos	26.47%
16	Depop	26.47%
17	Ryanair	26.47%
18	Ocado	26.47%
19	Airbnb	26.47%
20	American Airlines	26.47%
21	Ikea	23.53%
22	Trainline	23.53%
23	Amazon	23.53%
24	PayPal	23.53%
25	eBay	23.53%
26	Walmart	23.53%
27	Deliveroo	20.59%
28	Twitter	20.59%

Data categories (column headers): Email, Name, Age, Gender/Sex, Sexual Orientation, Marital Status, Race, Religious Belief, Live Location, Home Address, Employment Status, Job Title, Pet/Animal Ownership, Mobile Number, Landline Number, Type of Phone/Device, Hobbies, Interests, Height, Weight, Next of Kin, Mother's Maiden Name, Current Employers, Past Employers, Bank Account Details, Salary, Social Profile (Friends), Social Profile (Hobbies), Social Profile (Interests), Country of Birth, Allergies/Intolerances, Health & Lifestyle Info

Source: Atamaniuk M., 2020

Cookies do not work in the same manner everywhere as it used to be. In fact, since the beginning of the internet, when the user was surfing the net cookies where attached to the unique code related to his/her. The user did not need to provide a consensus on this practice – it was given for granted by already checking the consent box. In 2019 things started to change for users living in Europe. Indeed, for the EU legislation the pre-filled box giving consensus has no legal meaning. Moreover, as stated in the General Data Protection Regulation, each individual needs to be able to decide which type of cookies to accept. In general the choice is among functional, statistics and cookies for marketing

purposes. In order to not violate any EU legislations, web site need also to communicate the way user's data are going to be store and the length of the storage. In this way users are more aware of the type of data they are giving away in order to use a website. Instead, in the US it is still valid the old method, the automatic consensus.

However, cookies are not the only method used to gather information. For example, companies are interested in the way individuals are touching their phone. More issues regarding violation of privacy come up when analyzing emerging technologies such as virtual assistances. In order to provide virtual assistance, this type of devices is continuously listening to conversations and noises waiting for a command like "Alexa!" or "Ok Google" to turn on. Nevertheless, it has been shown that all the information that they collect while being in alert for a command are used to learn more about the user. Indeed, they are trained to be triggered and registering information when people are stating their preferences or when they are planning to go somewhere. Then the algorithm is processing all the information and it transform them into ads that the targeted individual will see everywhere online. For example, Oscar Schwartz (2018), in an article published by theoutline.com, described his experience. Briefly, he got as a present a Masomoto kitchen knife and without searching online more information about the gift or taking picture of it, he soon saw advertisement posts on Instagram about the same knife. The situation experienced by the author is not an isolated case. In the website theneworgan.com there is a wide collection of similar stories showing how smartphone and more in general smart devices are collecting information about us even when we think they are not. Given the increase of this type of event, concerns about privacy and possible leakages of private information has emerged. For companies and hackers it is easier to spy on individual and profiting on it (Henderson et al., 2018).

Some solutions regarding privacy issues have been developed in the tech industry. One example is differential privacy. Essentially, it is a way to collect and public data increasing the anonymity. Before exposing personal data, they are randomized by an algorithm. In this way it is possible to analyze and run researches with the data set but there is low probability of re-identification. However, the drawback is machine learning tools are becoming more powerful and in a near future they will still be able to re-identify

individuals. Moreover, it increases noise in the data set and if that data set was used for good purposes such as new drugs implementation, it reduces the accuracy of the prediction. Therefore, it is bad outcome for the society (Kearns and Roth, 2020). As we have seen, privacy is a hot topic in the ethical debate of AI tools. Nowadays, possible solutions are coming up but we are far from solving the problem.

2.10 Limited autonomy

Connected with the problem of privacy there is the one of limited autonomy of the end user. We define autonomy as the ability of the single individual to freely enjoy a website or an application according with her/his believes and desires without being forced by third parties to behave in a predetermined manner. However, when using recommendation systems or in general AI enabled products, autonomy is always limited since they work as a filter providing to users just items that are capturing their attention. By scrolling the feed of Facebook two are the posts most likely to be seen: the one matching the individual preferences or the one creating outrage. The second type of post is creating more engagement through the app, because the user is more likely to share the post with her friends. However, the algorithm ordering and filtering the post in the feed unlikely will show to the user a post out of his sphere of interests (Tristan Harries, 2017). In this way, the AI is limiting the autonomy of individuals in choosing what is best for them. Algorithms are undermine human dignity and self determination.

Lack of autonomy of the end user is connected with three factors. The first one, it is linked with the problem of filter bubbles – the more the algorithm is learning about the user interests the less the user is able to encounter diversified contents. Connected with this, there is the second factor which is the lack of understanding by the users on how algorithm works. Therefore, it is difficult for them to take precautions. Third, it is not possible for users to change the way algorithms operate. It is almost impossible to have news on the feed providing opposite side of views. For example, Twitter has changed the way it shows post on the feed of the user. At the beginning, tweets were seeing in chronological order, from the most to the least recent. In this way the user has the same probability in seeing each tweet independently on who were the writer and what was it about. However, this algorithm has changed in 2018 and tweets were rearranged

according to their popularity and to the interest of the single user. However, there has not been a strong transition, but for a while the user was deciding whether to see the timeline with the old or the new system. Nowadays, the only available option is the second one. Therefore, even though a user has decided to follow users that express different opinions it will be difficult for her/him to see them.

There are even social issues connected with hyper personalization. For instance, when using hyper personalized advertisement the drawback is that individuals are losing the opportunity of self determining their preferences for an item over the others. The issue of limiting autonomy and self determination has been widely discussed in the literature but also in governing bodies such as European Commission and UK's house of Lords. Nevertheless, hyper personalization is one of the main characteristic of social networks and when people create their accounts they accept the rules of the game. In fact, people online have the tendency to aggregate with individual who show similar interests or think in the same way. In addition, suggestions provided by the algorithm match to user's interests. This process has been defined as eco chamber or filter bubble. If on one side this mechanism increases user's satisfaction, on the other it creates an online reality that differs from the offline life. Online, individuals are living in a bubble since they are just exposed to a limited point of view. Hence, this is preventing the user to develop new interests and to being more informed. By forcing the interaction of the user with only information supporting one side of the story, it is easier to influence the thoughts and thus the behavior of users, especially of those more fragile (Nikolov et al, 2015). The major concern is linked with polarization of political discussion since people are stuck in those filter bubbles they are not exposed to information against their beliefs. Hence, people are less informed and consequently less tolerant of opposite perspective (Kearns and Roth, 2020). Individuals create strong bonds with users of the same group and by sharing same thoughts there is the tendency to reinforce values but also prejudices. Therefore, Web is considered as an aggregator of people through the creation of online community but at the same time it fortifies old prejudices (Amrollahi and McBride, 2019). One way to deal with the issue of polarization is by forcing the algorithm to show to users opposite view news. This vision is even supported by Thaler (2008), in the already discussed book *"Nudge improving decisions about Health, Wealth and Happiness"* in which he states that

surprise and serendipity are good source of information and it can happen that an individual would appreciate also what people different from her/him like.

To sum up, regarding online privacy, but more in general the ethical aspect of AI enabling tools, as stated before, there are some regulations at national and international level. However, when talking about artificial intelligence tools and their consequences it is difficult to provide upfront regulations protecting citizens from future losses. Regulations take time and it usually follows innovation. Moreover, another important aspect is that legislators need specific knowledge on the topic and it is not always like this. For instance, following the Cambridge Analytica's scandal there has been a lawsuit for Facebook. During the trial, congressmen have been asking to Mark Zuckerberg, Facebook's CEO, basic questions on the way the social media work such as how Facebook was generating money if users do not pay for the service. Therefore, how could congress change the game if it does not know the basic rules?

Companies now more than ever have the responsibility in shaping the future of AI. Corporate social responsibility strategies should start taking into consideration ethical principles in relation to AI tools (Du and Xie, 2020). In the following chapter, we are going to analyze the current scenario of Italian companies that develop AI technologies with a focus on understanding their perception of the ethical issues described so far.

CHAPTER 3

Once analyzed what is the current scenario on ethical consequences on the usage of AI technologies described in the literature, I wanted to understand how the situation looks like now and what is the perception of companies of all sized on the already discussed ethical problems. As regarding the geographical scope of my research, I have decided to focus just on Italian companies for two reasons. Besides wanting to explore the situation in my home country, I think it would be interesting to have the direct opinion of companies since the main finding so far presented were developed in USA and UK.

At the beginning of the chapter, there is an explanation of the research method and of the company who took part to the project, then the findings are presented.

3.1 Research method

In order to have a clear picture of the perception of Italian firms on ethical issues described in the previous chapter, I decided that the best method to gather data was through interview rather than close questions. Indeed, the former method allows for a broader exchange of opinion on topics that are complicated. The initial idea was to interview two groups of companies: the one developing AI and the one using software already developed. In this way, I could have a bigger picture of the situation.

The first thing that I have done was to collect all the name of companies connected with the development of AI. Among all companies that I could find, two were the main requirements. The company should develop directly in Italy and its technologies should directly link with ethical issues. For example, if AI products offered by an organization were focused just on better organizing internal workflow of activity, the company would not be chosen for this research.

For what concerns the second group of potentially interviewed companies, I have focused on firms using algorithms or software with strong ethical implications. The research of interesting case studies have been done mainly on the internet and in the website of companies belonging to the first category. By interviewing those type of firms I could have understood better what are the determinants of purchase of an AI technologies and whether ethical concerns play an important role.

Despite being able to collect and thus contact more than 20 organizations belonging to both categories, just few of them agreed to take part to this research. Moreover, all the companies belong to the first category and those are:

- MM ONE GROUP
- U HOPPER
- BLUETENSOR
- DATA VALLEY

Then I was able to get into contact with AEG[5] and I have decided to include as a multinational, to have a broader and also an European point of view. In the following section, there is a brief explanation about each company and the reason why it has been chosen.

3.1.1 MM One Group

It has been established more than 20 years ago in Noventa di Piave, in Veneto and nowadays it counts more than 50 employees. The company through its growth has acquired a relevant position in the CRM segment allowing MM ONE to reach 6 million in revenues every year. The group is focus on the B2B market in the development personalized digital strategy targeting the final consumers – us – to increase the costumers' awareness and performances. The group is working not only at national level but also at global, costumers are coming from 21 different states. Among their clients, there are significant organizations such as Caleffi and Dr. Schär.

The group has been chosen for this research for its solution on profilization using AI tools, especially regarding marketing automation and web analytics. Moreover, it was selected also for the way it promotes and reinsures privacy protection to its customers. For the interview, I was directly dealing with the founder and CEO – Mauro Cunial.

3.1.2 U-Hopper

It is a small company located in Trento, Trentino alto Adige, specialized in data analytics and it was established back in 2010 as a spin-off of an international research group called FBK. Despite the fact that the company is composed by just 16 people, U-Hopper is able

---

[5] It is a consulting company not to be confused with the international organization working in the domestic appliance industry

to work with clients coming from all over the world, for example US, Canada, Swiss etc. The aim of the firm is to help companies, especially local one, in creating value from data they own. In order to achieve their goal, they work a lot with machine learning technologies. In this way, a company can automatize the process of data analytics. The use of these technologies has been one of the reasons why I have chosen to interview U-Hopper.

Another relevant factor for which the company has been selected is its software TAPOI that is an AI technology sold to the B2B market for the targetization of any individual. Briefly, the software collect all the information that the individual consciously or unconsciously spreads on the social networks. By processing those data and combine them with important milestones (such as obtain a degree, getting married or having a child) achieved or assumed it would be achieve in a near future by the person, the software is able to predetermine her/his purchasing patter. Since the company is aware and thus sensitive to the ethical issues derived by the use of AI technologies as their blog contains articles and opinions on the matter, I thought it would be interesting to listen to their point of view and understanding how they manage technologies such TAPOI. During the interview, I had the opportunity to talk with Daniele Miorandi, the CEO of the company and the one in charge of ethical issues.

3.1.3 BlueTensor
The company is a 10 employees start-up located in Trento, Trentino Alto Adige, specialized in the development of AI software for the B2B market. The vision of the company is to promote a smart work method and a better life through the usage of AI tools. Their software are embedded with deep learning and machine learning technologies. BlueTensor keys technologies are: computer vision, predictive analysis, reasoning system that allows a faster decision making process and recommendation engine, already discussed in the second chapter.

Besides developing directly software with potential ethical concerns, it has been selected to be part of the research for its size. It would be interesting to see how companies of different sizes would face and deal with ethical issues. Moreover, BlueTensor like U-

Hopper has a blog/podcast managed by Jonni Malacarne, the co-founder and the person that I have interviewed, in which it currently explains its vision on the matter.

3.1.4 Data Valley

The organization has been established by the partnership with CRC Lex and Blum Comunicazione respectively a law firm and communication company. Therefore, Data Valley has the knowledge and expertise of both firms. Data Valley presents itself as a company that offers service of data management and data strategy through an organizational and legal point of view. Compared to the other companies so far described, Data Valley belong to another business. In fact, it does not develop AI software or other type of technologies. Instead, it brings together small Italian companies with big techs to facilitate the development of new business opportunity. Despite being an interesting activity, this is not the reason why the company has been selected, but rather for their work connected to algorithms and AI. Indeed, they offer consulting services to help firm evaluating what are the legal and ethical consequences of the implementation of any algorithm. Moreover, they provide to their clients ways on how to protect themselves on a legal point of view from the consequences of decisions made autonomously by algorithms. In the interview, I talked to Silvia Martinelli – strategic research manager at Data Valley as well as lawyer expert in law and new technologies.

3.1.5 AEG – Law Data Lab

AEG is an international consulting company settled in London and specialized in providing strategies and guidelines to companies and organizations worldwide. It is the biggest among the company selected for the research and it was established more than 30 years ago. Besides providing consultant service, it offers legal services through Leg Desk – a branch of the company. Leg desk is composed of three separated divisions: GDPR that helps organizations to be fully compliant with the European legislation, Sport Law Consulting focuses on arbitrage and sports litigations and Law Data lab. The latter is specialized in computational law and legal Artificial Intelligence. Briefly, they use AI algorithms to speed up legal consultancy. In my interviews, I have talked with Lorenzo Baldanello manager of Law Data Lab and member of the ethics committee. AEG has been chosen for the project for their ethical committee to better understand how an ethical

committee deals with AI technologies given that have managed delicate project such as the development of smart city and robot that assist elderly people.

3.1.6 Interviews

In my opinion, the best way to understand how each firm was dealing with ethical consequences in developing AI technologies was through individual interviews. The research period took place between January and June 2021 when national restrictions due to Covid-19 pandemic were promoting less contact as possible. Therefore, the only way to speak with representative was through video call. For each company, I set up an individual online meeting that lasted on average one hour.

The first part of the interview was structured thus allowing an easier classification of companies and comparison among them. Besides asking to describe the company and its history, I have asked whether they develop software inside their company. If the answer of the question is affirmative, the second question was on the description of the steps undertaken by the company before the sale of any developed software and on what are the main factors they usually check. Moving on, I ask what the ethical consequences are connected with the development of AI and how can be avoided by the company. Then, what is the company's perception of the ethical problem nowadays.

Depending on the answers of those questions, the second part of the interview was constructed, so there was no a fix set of question to ask. The conversation depended on relevant matters that have emerged in the first part of the conversation or a way to have practical examples on how they behave in specific situation. In two circumstances, with U-Hopper and Law Data lab, I have asked a second online meeting because there was more to discuss. In both cases, the second meeting was more about examples of ethical issues emerged with the development AI technologies and how they have dealt with them. All the results coming from all the interviews are discussed in the following section.

3.2 Data and results

Once all the interviews were done, I have proceed with the analysis of all the information that are presented in this section. For a purpose of clarity, I have decided to divide results into categories similar to the one described in the second chapter. In this way, it is easier

to compare the today situation of Italian companies with the situation described by the literature. The first section will focus on privacy and accountability, and then the following part is dedicated to social responsibility. In the last part, there will be a discussion on the cultural background of Italy that partly explains the reason why all the other ethical implications have not emerged while interviewing the companies.

3.2.1 Privacy

Privacy clearly is a sensitive topic. Indeed, in each interview it was nominated as the most important ethical consequence connected to the use of AI technology. Moreover, privacy is the only factor for which all companies are implementing strong protecting mechanisms. The main explanation for this common agreement on privacy protection lays on the European and thus Italian legal system. In fact, as confirmed by Silvia Martinelli, spokesperson of Data Valley, with GDPR organizations are held responsible of privacy violation linked to the usage of their product if they did not put in place a system that could have prevented the privacy violation.

Regardless of the shared vision on privacy protection, the way companies protect the privacy of the final consumer is completely different. On one hand, there are organizations such as BlueTensor and MM One Group that just comply with the law. For instance, MM One Group, besides having a DPO (Data Protection Officer), they have developed a software that automatically checks whether a developed software is compliant with the GDPR or not. They do not have ethical guidelines on the use of gathered data, they surely operate on the legality. However, as we have already seen in the previous chapter, if a thing is legal it does not mean that is ethical.

On the other hand, AEG tries to work with less data. They have understood the potential damage connected with overloading a software with sensitive information. Therefore, when analyzing a software/algorithm they try to remove data instead of adding them, if they are not essential for the output of the AI tool. Whenever it is possible, they try to use anonymized data or cryptography.

Another company that is more than compliant to the GDPR is U-Hopper. In fact, they impose stricter terms of usage to potential clients for their technologies. For example, in order to use TAPOI – a software used for the analysis of data – a company has to declare that all the data used as input are collected in a transparent way. For U-Hopper, data collected system can be considered as transparent only when individual explicitly give to a company the consent to keep and thus use their information for that specific purpose. By adding this requirement, they decrease the potential market of this technology by 80%, but for U-Hopper this is not a problem because they want to be coherent with their ethical principles.

3.2.2 Social responsibility

Social responsibility is not a concept previously described as a potential damage to the single individual or more in general to the society, that is because the concept of social responsibility is quite the opposite. It is a way to protect the society for the harm directly or indirectly linked to the usage of AI technologies. In this section, just two companies have shown proactive initiative in the protection of the citizens and not surprisingly, they are the one taking extra care in the management of sensitive data – U-Hopper and Law Data Lab. Having the organizations different size and hence resources they are dealing the problem in their own way and with the available resources.

U-Hopper's top management is well aware of the bad impact that AI could have on the society. The company has strong ethical unwritten guidelines spread top-down. As soon as the first interview started, Daniele Miorandi made it clear that he knows that software are not neutral and they can unconsciously carry unfair outcomes and discriminations. He believes that a company should not just be profit oriented but it has to provide value to its community. In this sense, a company have a social responsibility. Since they take this role very seriously, they control every software they develop along all the supply chain. Basically, it means that they try to imagine what are the consequences if the software is used as it has been developed. Moreover, they think what are the way the customer would handle it and what are the consequences linked with this alternative usage. By imagining all the possible usage, they try to prevent any harm to the final consumer.

This strong attitude to safeguard the interest of the final consumer is not reflected in an internal structural flow of activity aimed at detecting ethical damage but rather it is a check done by Daniele Miorandi – member of the management board. For every new project, he is the one making researches on potential use of the algorithm and on potential damages and his findings are then shown to the rest of the board and together they take the decision whether to reject the project if ethical issues can emerge. They prefer to prevent being associated with a company that it is likely to be involved with unethical behavior due to the bad image they would get. The reason why the company has not developed a structural flow of activity to detect unethical consequences in the usage of AI can be explained by two factors. Firstly, the company is not keen on the development of software that are linked with final consumers, the only one is TAPOI and it is not their main product in term of revenues. Indeed, U-Hopper's software are used to improve a third company workflow or to predict maintenance. Secondly, the company is too small to have complex structural analysis just for ethical implications.

AEG shares the same concerns about the use of AI and the consequences on the final consumer as U-Hopper. The main difference is that its size allows them to have a more structured process to analyze and thus detect problems. In fact, the company has an ethical committee composed of different types of expert in the field. For instance, Bruno Carenini gives with his experience as activist, his opinion on matters regarding potential infringement of human rights. The committee examines every ethical aspect connected to the project. Two are the main sources of ethical concerns: ethics connected to infringement of human rights and ethics and legality. In the former, they are evaluating even the consequences of implementing automation on the labor market of the community of the company. Instead, in the latter category belongs all the activity that are in a borderline between legality and illegality.

Every time there is a new project or consultancy, Lorenzo Baldanello is notified. If he thinks the project could have ethical implication he brings it to the attention of the committee. Another way in which the committee is in charge of ethical investigation is when the client specifically asks for it. In any case, the process is the same and it is as follows. In the first general meeting, all the experts gather together and the project is

explained in every detail, then each of them analyses the case for the following two weeks and in the following group meeting concerns are raised. The output of the committee is a report containing all the concerns and what are the action that can be taken in order to avoid them. For instance, AEG has been involved to provide their consulting in the early stage of a development of a smart city build from zero with an initial capacity of 100.000 citizens. In the city, there are biocompatible buildings strongly integrated with AI technologies and the company has to determine what would be the implication of having such high level of technology. For example, robots are used to provide faster and more precise services to the citizens especially those connected with situations of emergency, such as fire. To determine whether it was ethical the employment of a robot they have done a pro and cons analysis. What emerges is that robots are substituting human beings just for activities that would put the life of humans in danger such as entering a burning building. In this case, it would be unethical to choose an individual over a robot to enter the building. Therefore, AEG has suggested that in case like this it is ethical to use robot to provide assistance.

However, ethical concerns have been risen by the committee in another aspect of the development of the smart city. Briefly, the idea was to increase the security of the city through mass surveillance with facial recognition devices. Moreover, in the city it would be possible to access to some building only with individual biometrics parameters. This, for AEG, could constitutes privacy issues. Since the city is not established in Europe the company does not have to be complaint with GDPR and there was not such strict rules in the country. Nevertheless, AEG in the report advises its client to pay more attention to privacy issues. In addition, they suggest that people that want to live in the smart city have to accept explicitly this mass surveillance control and they proposed to make it optional the use of biometric parameters as keys to have access to part of the city. Unfortunately, at the moment of the last interview with AEG, the smart city was still in the development so we do not know if the suggestions have been followed.

3.2.3 Culture, ethics and new role of Europe
As we have seen in the previous paragraphs, a company is just compliant to the legislation or it is taking strong measures to protect the life of its community, it can be said that if an

organization is aware of what could be the ethical implications on the final consumers then it is more likely to implement a protection mechanism. In fact, MM One Group and BlueTensor were not conscious of the negative consequences on the population such as unfair outcome or discrimination, for them ethics was just a matter of privacy.

This lack of interest in the ethical consequences of AI devices displayed by Italian companies is a consequence of the lack of ethical debate on the usage of technologies – instead this is a daily topic in the US. In fact, almost all the papers cited in this dissertation has been published in the US and UK. In Anglo-Saxon countries, it is normal to deal with ethics. Indeed, their legal system does not depend on law that regulate cases but rather from a case a law is established. For this reason, as confirmed during the interview with Silvia Martinelli, in those countries the principles have more power than in non-Anglo-Saxon one. Therefore, in Italy but more in general in Europe the lack of activism on ethics of AI is due to a lack of norms and regulation in the legal system. However, EU is currently dealing with this issue and in the upcoming months a new regulation should be out. The aim of this regulation is to build a safer place for its citizens by imposing a trustworthy AI. The second objective of the regulation is to increase the competitiveness of EU in the AI market. Indeed, nowadays, when talking about Artificial Intelligence and more in general about technologies three are upfront nations: United State, China and Israel. However, Israel has less power than the other two and for this reason it is not in the battle for supremacy. At the same time, European Union wants to step in and boost its global competition on Artificial intelligence. EU has understood that it is too late to compete on innovation or against the giant of the big tech, because it would require more time and investment. The strategy adopted by the EU is to create a safer environment for its citizens by imposing the use of trustworthy AI. According to the European Commission, in order to achieve trustworthy AI three are the necessary conditions. First of all, it has to comply with the legislation, then it should be in line with ethical principal and finally yet importantly, it has to be robust. As it can be understood from these principles, ethical concerns regarding AI are at the core of the legislation. This explains also the reason behind the publication in 2019 of *Assessment list for trustworthy AI*. The document is been the result of a high-level group of expert representing a wide range of stakeholders hired to draft European ethics guideline. The output contains suggestions on how to prevent

ethical issues described in the second chapter. According to the European Commission, this guide should be applied to every individual who develops, provides and uses AI tools. The main drawback is that this is just a recommendation, it is not legally binding – this will not be a problem after the publication of the new legislation concerning AI.

The idea is that in the future, a foreign company has to be sure its product is complying with a stricter regulation if it want to provide its technology in the European market. With higher barriers to entry, it is easier to sustain local companies (Martinelli, 2021). In order to facilitate the growth of the tech industry, the European Commission has increase the funds aimed at AI by 70% in the program Horizon 2020. Besides boosting the technological development, EU aims at ensuring the right economic and social protection from the change derived by the implementation of AI. Moreover, it wants to provide adequate ethical and legal framework (European Commission, 2021).

The central pillar of the EU strategy is human- centricity, this is an unconventional choice, but it is in line with the aim of transforming the continent into a niche market. With this point of view, AI is an enabling tool that allows not only citizens but also the community to live better and to increase their freedom. The union understood that more policies are needed to create the right environment for the development of human center AI. The goals described by a high-level group of expert on AI (European Commission, 2021) of new policies are the following, it is not a ranking base on importance. First, the aim is to provide to individuals sufficient knowledge of how AI works. Second, to safeguard the integrity of humans, society and environment by preventing any harm cause by discrimination, unfairness, biases etc.. Another relevant aspect connected with the protection of society is that AI should not be used to construct mass surveillance mechanism, but instead as a way to protect the environment and help to achieve Agenda 2030's targets. Third, the goal is to increase the protection of human at work – automation derived from AI tools should substitute dangerous tasks, doing so there is an increase in the safety in the work place. Forth, the increase in adoption of AI technologies should not be linked with an increase in discrimination, but rather should be a means of inclusion of individuals in the society.

Overall, it can be said that European Union wants nations and organization to be aware of the negative ethical consequences described so far connected with the use of AI technologies not only in the short run but also in the long run. As emerged from the research within Italian organizations the main problem was that they were not conscious of the potential impacts that their technologies could have on individuals. Therefore, according to the EU the right way to prevent any harm to the population such as discrimination or unfair outcome is to provide more legislations and build a system to educate the individuals about the potentiality and threat of the new technologies.

However, the main concern on relying only to the legislation for preventing any harm to the population is timing, meaning that it is difficult for the legislation to keep up with the innovation. Moreover, the process behind the creation of a new law/directive is full of bureaucracy causing a delay on its promulgation. According to Bryson (2018), this latency exists to avoid frequent changes of the status quo that would damage business activities.

3.3 An opportunity for the Italian AI industry

The result of the research, presented in the previous paragraphs, shows companies in two different positions: compliant to the national and European regulations and the one over compliant - they have shown extra care in managing data. Today, organizations belonging in both categories are able to compete against each other in the market especially among non-tech savvy. The main reason is that average users are not taking into considerations ethics and its consequences when purchasing a technological product or using a service. However, things are changing. A growing number of individuals is choosing AI embedded products/services more consciously when their sensitive data are involved. This change has been proved also during the interview with Lorenzo Baldanello, AEG, in which he stated that more people every year are asking companies to delate the data they have gathered about him/her.

Companies in order to gain a long-term competitive advantage should start moving from just compliant to over complaint. In this way, they are able to attract this growing segment of conscious consumers. Apple for example is already making this transition as a part of its differentiation strategy. When looking at the market share of the smartphone both
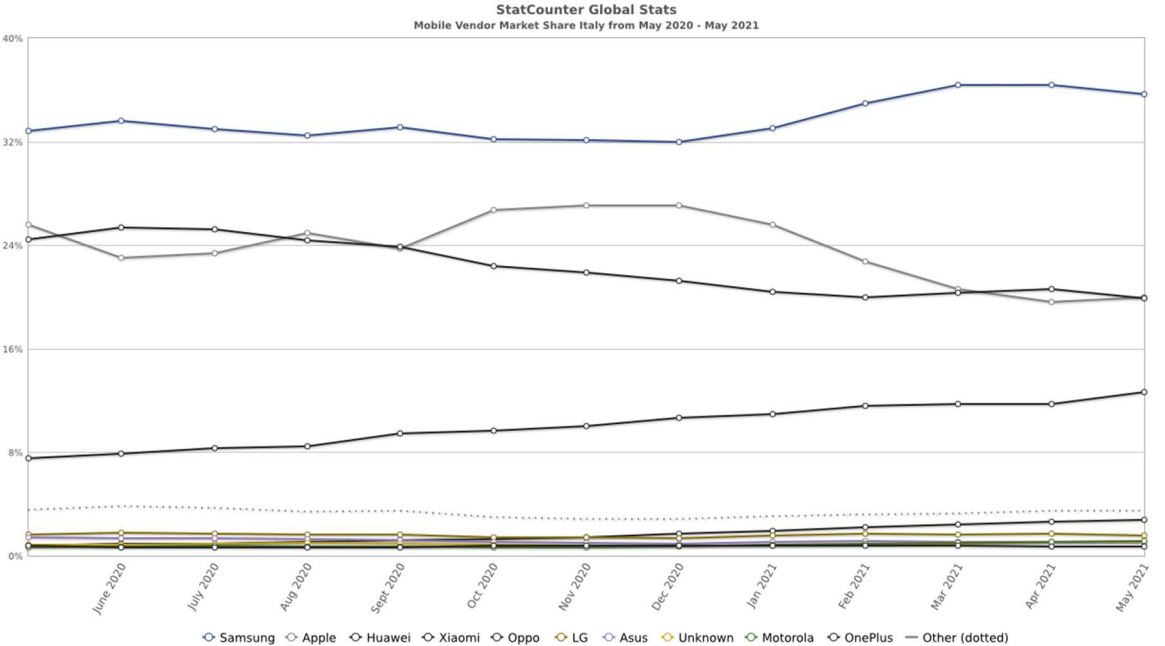
worldwide (Figure 13) and in Italy (Figure 14), the company is in the second place beyond Samsung. In Italy, Apple's market share are almost the same as Huawei, an emerging Chinese company.

Figure 13
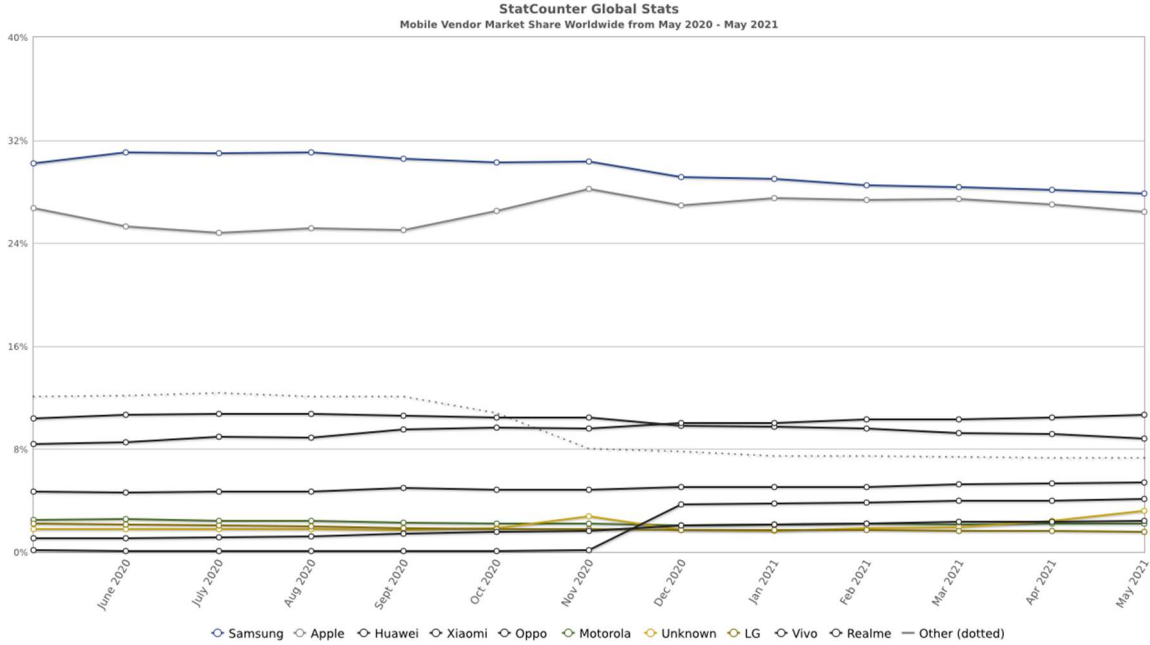
Source: (Statcounter, 2021)

Figure 14

Source: (Statcounter, 2021)

Apple, compared to all the other companies, is charging a higher price for its product and the difference in price with Huawei is quite significant. Instead of competing on prices, Apple's strategy in 2021 consists in emphasizing how the company is different from the other in treating the privacy of the users. Indeed, the new advertisement campaign called Privacy on iPhone | Tracked is not focused on the features of the phone but rather on the new management of privacy that is more transparent. With the latest system update, for each application the user has to explicitly allow or deny tracking of activity inside and outside the application. Apple is the first company in the smartphone market that is leveraging on its privacy transparency system – this move would attract this emerging niche market. Organizations more likely to lose this type of users are the Chinese one such as Huawei and Xiaomi, since their management of personal data is ambiguous and not transparent. In fact, Huawei has been accused by government of Netherland of acquiring personal data of ministers from phone calls (Henley, 2021). Therefore, Apple, by being the first mover in being over compliant, can gain a long-term competitive advantage and maybe increase its presence in the market.

During interviews with most of the organizations, the one belonging in the compliant category, my feeling was that companies were not even aware of the potential damages of AI technologies so it is difficult to imagine for them a transition to the over compliant category in the short period. It seems like that the public discussion taking place nowadays about the benefits but especially about the drawbacks of AI tools, is not happening in the same context as the one in which firms operates. I hope that the discussion would lead to concrete actions by the government and/or organizations. However, for now it seems like it is just a rhetorical discussion. For example, during the *Festival dell'Economia di Trento* (Economics Festival in Trento), Vittorio Colao, the Minister of technological innovation and digital transition, expressed his perplexity about the fragility of the system used for store data in Italy. Nevertheless, there is no planned investment for improving this system on the new PNRR (National Plan for Resistance and Resilience).

Interviewed companies have ensured me that all the technologies developed by them cannot be linked with potential harm on the population. Indeed, all the organizations are focusing on AI technologies aimed at improving the efficiency of the workflow inside the

customers' firm or aimed at predicting the maintenance of machines. Software that focus on the profilation or that are aimed at nudging people are still underdeveloped. One possible explanation stated by the companies is that at EU and Italian level there are already too many regulations that limits the explorations of new technologies and they do not have as much freedom as companies located overseas. Another explanation, more plausible, is linked with the stage of development of the industry in the northeast of Italy where most of the interviewed organizations are established. Companies located in this territory have not invested yet in data analytics and for this reason, AI technologies developed by them are more basic and the possibility of harming the population is very low. Therefore, companies are not equipping themselves on how to prevent any possible harm, none of the interviewed organizations has a proper flow of activity aimed at checking the potential damages on consumers. Nowadays, Italian companies are waiting for a problem to happen and they are not developing proactive protection mechanisms. In Italy, a proper boom of the industry is likely to happen in the following years and for this growth, a maturation is necessary both in the technical competences – to develop more sophisticated technologies – and in the prevention of ethical damages. There is the need to transform the today discussion in concrete actions as soon as possible. In this way, Italian firms would be able to switch from just compliant to over compliant acquiring competitive advantage of the early movers previously discussed.

The today situation holds huge potential for the future of Italian AI industry. There is the possibility of moving in advance and be prepared for these new technologies. I believe that to exploit better the opportunity, all the involved stakeholders – companies, citizens and institutions – should cooperate with each other. By working together, there is an alignment of power among the three actors, since companies detain more power than citizens do. Moreover, it avoids the problems connected with the organization regulating themselves – big techs overseas are starting to develop their own rules for managing ethics. This practice can generate an agency problem – there could be a conflict of interest between profiting from technologies and safeguarding final users from ethical damages. Google for example has published self-regulating principles on the AI, but on reality they do not impose regulations on how to deal with ethics and AI – it is more an ethicswashing (Canca, 2020).

61

The government should build a framework to define the context in which firms should operate. The legislations need to be general because it is impossible to regulate every single case. Another reason why it should be general is to allow organizations to experiment. A framework too narrow in which everything is regulated can even prevent the discovery of new technologies from which the society could benefit from. It is known that innovation comes from exploration and many innovative products we could not live without come from mistakes made in the past, this is the example of pacemaker and x-ray. Companies not dealing actively with ethics are gambling with the trust of consumers and with the government. As soon as something does not go as planned, a scandal happens and it is followed by punitive regulations as it has occurred with Cambridge Analytica. To avoid any scandal, which is usually connected to harm on the population and a consequent narrow framework, institutions and companies should invest more into the education of individuals towards ethics and AI. As confirmed during the interview with Daniele Miorandi, ethics course should be included in every bachelor's and master's degree. This can be a good starting point; however, it is not sufficient to effectively change the attitude of people. As stated by Cristina Pazzanese (2020), writer for Harvard Gazette, introducing a single course on ethics and philosophy is not enough in changing the mentality of programmers from "can I build it?" to "should I build it?". What needs to be changed is the teaching methodology. Nowadays, at Harvard they combine technical and ethical concepts along all the duration of the curses teaching students how to recognize ethical issues when dealing with technical problems. This way of teaching should be soon applied also in Italian institutions. The more a developer is aware of the potential damage of AI the more likely he/she would ask for tools aimed at identifying and solving ethical issues before they emerge. At the same time, companies have to implement a proactive ethics culture also in their operations. This could be done by cooperating with ethics experts through the establishment of experts committee that are independent from the company and they have the right competences to judge the feasibility of a project and its correlated potential damages (Canca, 2020).

EU with the new regulation on AI is trying to create a Niche market for the AI industry to promote the development of local organizations focused on developing AI tools. Since in

Italy this industry is still underdeveloped, the country should prepare itself by investing in acquiring new technical and ethical knowledge about these new technologies. It would be too late for Italy to invest in the transition of companies from compliant to over compliant once the EU regulation on AI is out. Indeed, it takes time for implementing changes in educational system and even more time is required for the new knowledge to be integrated in the workflow of the companies. I believe that Italy should exploit the opportunity provided by the EU to increase its competitiveness in the development of ethical AI technologies.

CONCLUSION

The tech industry has been able to provide a huge number of innovative products to the market, from chat bot to profilization. Despite the huge spread of technologies, the average users has only few knowledge about how these AI tools are functioning. The lack of knowledge represents a problem especially if it regards employers of companies developing and using the AI tools. The issue is not linked to a missed opportunity to exploit the advantages of the technology but rather it is connected to a potential harm at the expenses of an individual or more in general on a community. When talking about the side effects of AI, it is usually referred as ethical consequences such as unfair outcome, inconclusive evidence, inscrutable evidence, nudging, privacy breaches and limited autonomy.

The aim of the thesis was to understand how much Italian companies are aware of the negative consequences that a technology embedded with AI has on every individual. Companies that took part to the research are mainly focused on directly developing AI software. From the interviews, it has emerged that all companies have established mechanisms to protect the privacy of consumers, especially when sensible data is managed, mainly because of the GDPR, the European regulation for privacy. It can be said that companies are just compliant with a legislation. Indeed, when they have been asked if there were other mechanisms to protect the final users from the other ethical consequences most of them have denied.

The reason why companies are not safeguarding individuals lays in the culture, in fact, Italy, as all the other pan European countries, is not used to deal with ethics as Anglo-Saxon countries do. Therefore, it is difficult for organizations with this cultural background to question themselves on ethical problems. They only make sure to be compliant with the law. Consequently, European Union has stepped in in the discussion of ethics and AI. Indeed, it has started to interrogate itself on what would be the shape of AI in the community and what should be the guiding principles for a sustainable development in the long run.

At the moment, it seems like that the goal of EU is to safeguard the individuals by imposing strong regulations at companies by undertaking the so called human centric AI approach. At the same time, the regulation aims at building a niche market in order to promote the development of the European high tech industry. In Italy, as demonstrated by the research, the measures taken by the organization to prevent any harm on the consumers are quite low – this is mainly due to the fact that the industry is in its early development stage, technologies are still to basic compared to the ones made by competitors overseas. To close the gap, the institutions should invest in the development of the industry by promoting the acquisition of technical knowledge. Moreover, to exploit the niche market that is going to be created by the new regulation of EU on AI, Italy should also start investing in educating people on how to develop these new technologies while paying attention to ethical issues. At the same time, companies should also invest in implementing more mechanisms that help to detect and prevent any harm to the population. I believe that these are the pillars needed for a sustainable development of AI technologies.

REFERENCES

Alemanno, A. (2012). Nudging Smokers The Behavioural Turn of Tobacco Risk Regulation. *European Journal of Risk Regulation, 3*(1), 32-42

Amrollahi A., Mcbride N., (2019), HOW TO BURST THE BUBBLE IN SOCIAL NETWORKS?, Conference: 24th UK Academy for Information Systems International Conference, [Online], available at:
https://www.researchgate.net/publication/331333116_HOW_TO_BURST_THE_BUBBLE_IN_SOCIAL_NETWORKS (accessed: 07.01.2021)

Anderson C., (2009), "What is "brain hacking"? Tech insiders on why you should care", *60 minutes*, , [Online], available at: https://www.cbsnews.com/news/brain-hacking-tech-insiders-60-minutes/ (accessed: 03.01.2021)

Atamaniuk M., (2020), Which Company Uses the Most of Your Data?, Clario, [Online], available at: https://clario.co/blog/which-company-uses-most-data/ (accessed: 20.01.2021)

Benfatto L., (2016), "Microsoft blocca il software Tay: era diventato razzista e xenofobo", *Sole24Ore*, [Online], available at: https://st.ilsole24ore.com/art/tecnologie/2016-03-25/microsoft-blocca-software-tay-era-diventato-razzista-e-xenofobo--095134.shtml (accessed: 08.11.2020)

Borges A. F.S., Laurindo F. J.B., Spínola M. M., Gonçalves R. F., Mattos C. A., (2020), The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions, *International Journal of Information Management*

Bryson, Joanna J.. (2018), "The Past Decade and Future of AI's Impact on Society." In Towards a New Enlightenment? A Transcendent Decade. Madrid: BBVA, [Online], available at: https://www.bbvaopenmind.com/en/articles/the-past-decade-and-future-of-ais-impact-on-society/ (accessed: 12/06/2021)

Canca C., (2020), *How To Solve AI's Ethical Puzzles | Cansu Canca | TEDxCambridgeSalon*, [Online], available at: https://www.youtube.com/watch?v=cplucNW70II&ab_channel=TEDxTalks (accessed: 23.06.2021)

Capgemini, (2020), Research: AI in Customer Experience, [Online], available at: https://www.capgemini.com/news/ai-in-customer-experience-2/ (accessed: 08.11.2020)

Carrie Wong J., (2019), "The Cambridge Analytica scandal changed the world – but it didn't change Facebook", *The Guardian*, [Online], available at: https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook (accessed: 04.01.2021)

Chethan K. GN., (2018), "Artificial Intelligence: Definition, Types, Examples, Technologies", *Medium,* [Online], available at: https://medium.com/@chethankumargn/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b (accessed: 25.10.2020)

Cruz T., (2015), "Ted Cruz using firm that harvested data on millions of unwitting Facebook users", *The Guardian*, [Online], available at: https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data (accessed: 27.12.2020)

Davenport T. H., Ronanki R., (2018), "Artificial Intelligence for the Real World", *Harvard Business Review*, [Online], available at: https://hbr.org/2018/01/artificial-intelligence-for-the-real-world (accessed: 02.11.2020)

Davies H., (2015), "Ted Cruz using firm that harvested data on millions of unwitting Facebook users", *the Guardian*, [Online], available at: https://hbr.org/2018/01/artificial-intelligence-for-the-real-world (accessed: 22.01.2021)

Du S., Xie C., (2020), Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities, *Journal of Business Research*

Duan Y., Edwards J. S., Dwivedi Y. K., (2019), Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda, *International Journal of Information Management*, Volume 48, Pages 63-71

DW Documentary, (2019), Artificial intelligence and algorithms: pros and cons | DW Documentary (AI documentary), [Online], available at: https://www.youtube.com/watch?app=desktop&v=s0dMTAQM4cw&ab_channel=DWDocumentary (accessed: 05.11.2020)

European Commission, 2021. *Digital strategy.* [Online]
Available at: https://digital-strategy.ec.europa.eu/en/policies/strategy-artificial-intelligence (accessed: 17.03.2021)

European Commission, 2021. *Policy and investment recommendations for trustworthy Artificial Intelligence.* [Online] Available at: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence (accessed 17 03 2021)

European Parliament, (2020), Is data the new oil?, [Online], available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/646117/EPRS_BRI(2020)646117_EN.pdf (accessed: 25.11.2020)

Fagella D., (2016), "Exploring the risks of artificial intelligence", *Techcrunch*, [Online], available at: https://techcrunch.com/2016/03/21/exploring-the-risks-of-artificial-intelligence/?guccounter=2 (accessed: 03.11.2020)

Feigenbaum E. A., (1992), Expert Systems: Principles and Practice, The Encyclopedia of Computer Science and Engineering

Friedman T. L., (2012) "Average is over", *The New York Times*, [Online], available at: https://www.nytimes.com/2012/01/25/opinion/friedman-average-is-over.html (accessed: 03.11.2020)

Grand view Research, (2020), Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution (Hardware, Software, Services), By Technology (Deep Learning, Machine Learning), By End Use, By Region, And Segment Forecasts, 2020 – 2027, [Online], available at:

https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market (accessed: 03.10.2020)

Ghosh A., Chakraborty D., Law A., (2018), Artificial Intelligence in Internet of Things, *CAAI Transactions on Intelligence Technology*. 3. 10.1049.

Haenlein M, Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*. 2019;61(4):5-14

Harri J., (2018), "Why the government must help shape the future of AI", *The Brookings institute*, [Online], available at: https://www.brookings.edu/research/why-the-government-must-help-shape-the-future-of-ai/ (accessed: 05.11.2020)

Harris T., (2017), Hw a handful of tech companies control billions of mind every day [Online], available at: https://www.ted.com/talks/tristan_harris_how_a_handful_of_tech_companies_control_billions_of_minds_every_day/details?language=it (accessed: 13.12.2020)

Hansen, P. G. (2016). "The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?" European Journal of Risk Regulation 7 (01), 155–174

Hemmendinger D., (2016), "LISP", *Encyclopedia Britannica*, [Online], available at: https://www.britannica.com/technology/LISP-computer-language (accessed: 21.10.2020)

Henderson P., Sinha K., Angelard-Gontier N., Ke N. R., Fried G., Lowe R., Pineau J., (2017), Ethical Challenges in Data-Driven Dialogue Systems, *Cornell University*, [Online], available at: https://arxiv.org/abs/1711.09050 (accessed: 05.01.2021)

Henly J., (2021), "Huawei 'may have eavesdropped on Dutch mobile network's calls'", *The Guardian, [Online]*, available at: https://www.theguardian.com/technology/2021/apr/19/huawei-may-have-eavesdropped-on-dutch-mobile-networks-calls (accessed: 13.06.2021)

Hern A., (2018), "Cambridge Analytica: how did it turn clicks into votes?", *The Guardian*, [Online], available at: https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie (accessed: 27.12.2020)

Hosanagar K., Jair V., (2018), We Need Transparency in Algorithms, But Too Much Can Backfire, *Harvard Business Review,* [Online], available at: https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire (accessed: 30.11.2020)

Hunt E., (2016), "Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter This article is more than 4 years old", *The Guardian,* [Online], available at: https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter (accessed: 21.10.2020)

IBM Cloud Education, (2020), Strong AI, [Online], available at: https://www.ibm.com/cloud/learn/strong-ai, (accessed: 20.10.2020)

Ingold D., Soper S., (2016), "Amazon Doesn't Consider the Race of Its Customers. Should It?", *Bloomberg*, [Online], available at: https://www.bloomberg.com/graphics/2016-amazon-same-day/ (accessed: 29.11.2020)

Intel, (ND), The Story of the Intel® 4004, [Online], available at:
https://www.intel.it/content/www/it/it/history/museum-story-of-intel-4004.html (accessed:
21.10.2020)

Internet World Stats, (ND), INTERNET USAGE STATISTICS, [Online], available at:
https://www.internetworldstats.com/stats.htm (accessed: 24.11.2020)

Iqbal M., (2020), "Facebook Revenue and Usage Statistics (2020)", *Business of Apps*, [Online],
available at: https://www.businessofapps.com/data/facebook-statistics/ (accessed:
12.12.2020)

Kamark E., (2018), "Malevolent soft power, AI, and the threat to democracy", *The Brookings
Institutions*, [Online], available at: https://www.brookings.edu/research/malevolent-soft-
power-ai-and-the-threat-to-democracy/ (accessed: 02.01.2021)

Kaplan A., Haenlein M., (2020), Rulers of the world, unite! The challenges and opportunities of
artificial intelligence, *Business Horizons*, Volume 63, Issue 1,Pages 37-50

Karlsen R., Andersen A., (2019), Reccomendations with a Nudge, *Technologies*, 2019, 7, 45

Kok J. N., Boers E. J. W., Kosters W. A., and van der Putten P., (2009), Artificial Intelligence:
Definition, Trends, Techniques and Cases, Encyclopedia of Life Support Systems (EOLSS)

Konstan J. A., (2004), Introduction to recommender systems: Algorithms and Evaluation, ACM
Transactions on Information Systems, Vol. 22, No. 1, January 2004, Pages 1–4

Lam S. K., Frankowski D., Riedl J., (2006) Do You Trust Your Recommendations? An Exploration
Of Security and Privacy Issues in Recommender Systems, Lecture Notes in Computer
Science book series (LNCS, volume 3995)

Lembcke, Tim-Benjamin; Engelbrecht, Nils; Brendel, Alfred Benedikt; and Kolbe, Lutz, (2019).
"TO NUDGE OR NOT TO NUDGE: ETHICAL CONSIDERATIONS OF DIGITAL NUDGING BASED ON
ITS BEHAVIORAL ECONOMICS ROOTS". In Proceedings of the 27th European Conference on
Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019

Loung T., Edit O. M., Duc P., Huy V.N., Raisa C., Minh P. and Dinh S. B. (2021) The COVID-19 global
pandemic: a review of the Vietnamese Government response, *Journal of Global Health Reports*
[Online], available at: https://www.joghr.org/article/21951-the-covid-19-global-pandemic-a-
review-of-the-vietnamese-government-response (accessed: 29.05.2021)

Madrigal A. C., (2018), "Facebook Didn't Sell Your Data; It Gave It Away", *The AtlanticI*, [Online],
available at: https://www.theatlantic.com/technology/archive/2018/12/facebooks-failures-
and-also-its-problems-leaking-data/578599/ (accessed: 12.12.2020)

Manietti E., (2018), "Il caso Cambridge Analytica, spiegato bene", *Il Post*, [Online], available at:
https://www.ilpost.it/2018/03/19/facebook-cambridge-analytica/ (accessed: 12.12.2020)
1
Martin K., (2018), Ethical Implications and Accountability of Algorithms, *Journal of Business
Ethics*, volume 160, pages 835–850

Martinelli, S., 2021. Discussion on ethics of AI in Europe [Interview] (04.02.2021)

Matsakis L., (2019), "The WIRED Guide to Your Personal Data (and Who Is Using It)", *Wired*, [Online], available at: https://www.wired.com/story/wired-guide-personal-data-collection/ (accessed: 04.01.2021)

McCarthy, J., Minksy, M., Rochester, N., and Shannon, C. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial intelligence. Technical report

Meyer Z., (2018), "As China Marches Forward on A.I., the White House Is Silent", *The New York Times,* [Online], available at: https://www.nytimes.com/2018/02/12/technology/china-trump-artificial-intelligence.html?action=click&module=RelatedLinks&pgtype=Article (accessed: 26.10.2020)

MIT, (2001), MIT team building social robot, [Online], available at: https://news.mit.edu/2001/kismet (accessed: 23.10.2020)

Mittelstadt B. D., Allo P., Taddeo M., Wachter S., Floridi L., (2016) The ethics of algorithms: Mapping the debate, *Sage jourals*, [Online], available at: https://journals.sagepub.com/doi/full/10.1177/2053951716679679 (accessed: 04.10.2020)

Muro M., (2019), "Countering the geographical impacts of automation: Computers, AI, and place disparities", *the Brookings institutions, [Online], available at: https://www.brookings.edu/research/countering-the-geographical-impacts-of-automation-computers-ai-and-place-disparities/ (accessed: 02.11.2020)*

Nikolov D., Oliveira D. F., Flammini A., Menczer F., (2015), Measuring Online Social Bubbles, *Center for Complex Networks and Systems Research, Indiana University*, [Online], available at: https://www.researchgate.net/publication/272845437_Measuring_Online_Social_Bubbles (accessed: 06.01.2020)

O'Connel B., (2018), "How Does Facebook Make Money? Six Primary Revenue Streams", *TheStreet*, [Online], available at: https://www.thestreet.com/technology/how-does-facebook-make-money-14754098 (accessed: 12.12.2020)

O'Meara S., (2019), Will China Overtake the U.S. in Artificial Intelligence Research?, *Nature magazine,* [Online] available at: https://www.scientificamerican.com/article/will-china-overtake-the-u-s-in-artificial-intelligence-research/#:~:text=According%20to%20the%202018%20China,States%2C%20which%20had%20roughly%2029%2C000 (accessed: 26.10.2020)

Omnicore, (2021), Instagram by the Numbers: Stats, Demographics & Fun Facts, [Online], available at: https://www.omnicoreagency.com/instagram-statistics/ (accessed: 10.12.2020)

Panazzese C., (2020), Computer science, philosophy faculty ask students to consider how systems affect society, *The Harvard Gazette,* [Online], available at: https://news.harvard.edu/gazette/story/2020/10/experts-consider-the-ethical-implications-of-new-technology/ (accessed: 23/06/2021)

Pham Q., Nguyen D. C., Huynh-The T., Hwang W., Pathirana P. N., (2020), Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts, [Online], available at: https://www.researchgate.net/publication/342964155_Artificial_Intelligence_AI_and_Big_Data

for_Coronavirus_COVID-19_Pandemic_A_Survey_on_the_State-of-the-Arts (accessed: 10.11.2020)

Praharaj K., (2020), "Bias In AI Algorithms", *towards data science*, Online], available at: https://towardsdatascience.com/how-are-algorithms-biased-8449406aaa83 (accessed: 02.12.2020)

Ramos, C., Augusto, J. C., & Shapiro, D. (2008). *Ambient Intelligence—the Next Step for Artificial Intelligence. IEEE Intelligent Systems, 23(2), 15–18.*

Regulation 2016/679, European Parliament

Reisacj U., (2020), The responsibility of social media in times of societal and political manipulation, *European Journal of Operational Research*, [Online], available at: https://www.sciencedirect.com/science/article/pii/S0377221720308249 (accessed: 02.01.2021)

Ridson C., (2017), Scaling Nudges with Machine Learning, *Behavioural scientist,* [Online], available at: https://behavioralscientist.org/scaling-nudges-machine-learning/(accessed: 02.01.2021)

Statcounter, (2021), Mobile Vendor Market Share Italy, [Online], available at: https://gs.statcounter.com/vendor-market-share/mobile/italy/#monthly-202005-202105 (accessed: 13.06.2021)

Statcounter, (2021), Mobile Vendor Market Share Worldwide, https://gs.statcounter.com/vendor-market-share/mobile/worldwide/#monthly-202005-202105 [Online], available at: (accessed: 13.06.2021)

Stormon A., (2019) "The 5 forgotten benefits of consumer data you need to know", *Oracle*, [Online], available at: https://blogs.oracle.com/oracledatacloud/the-5-forgotten-benefits-of-consumer-data-you-need-to-know (accessed: 15.11.2020)

Tarabay J., (2021) "Government Tap Covid Data for Other Uses, Risking Backlash", *Bloomberg*, [Online], available at: urly.it/3b8zj (accessed: 14.02.2021)

Thaler R., (2008), Nudge improving decisions about Health, Wealth and Happiness, Yale University Press, New Haven & London

Toscano J., (2019), "China Will Outpace US Artificial Intelligence Capabilities, But Will It Win The Race? Not If We Care About Freedom", *Forbes,* [Online], available at: https://www.forbes.com/sites/joetoscano1/2019/12/03/china-will-outpace-us-artificial-intelligence-capabilities-but-will-it-win-the-race-not-if-we-care-about-freedom/?sh=6743d0c96ed3 (accessed: 26.10.2020)

Tsamodos A., Aggarwal Ni., Cowls J., Morley J., Roberts H., Taddeo M., Floridi L., (2020), The Ethics of Algorithms: Key Problems and Solutions, [Online] available at: https://ssrn.com/abstract=3662302 (accessed: 02.10.2020)

Vigen T., (ND) Spurious correlation, Tylervigen, [Online], available at: http://www.tylervigen.com/spurious-correlations (accessed: 26.11.2020)

Vox Creative, (2018), "How AI can help us clean up our land, air, and water", *Recode*, [Online], available at: https://www.recode.net/ad/18027288/ai-sustainability-environment (accessed: 07.11.2020)

White C. L. and Boatwright B., (2020), Social media ethics in the data economy: Issues of social responsibility for using Facebook for public relations, *Public relation review,* volume 45, issue 5, [Online], available at: https://www.sciencedirect.com/science/article/pii/S0363811120301077 (accessed: 17.01.2021)

World Economic Forum, (2020), A new Paradigm for Business of Data, [Online], available at: http://www3.weforum.org/docs/WEF_New_Paradigm_for_Business_of_Data_Report_2020.pdf (accessed: 24.11.2020)

Yanqing D., John S. E., Yogesh K. D., (2019), Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda, *International Journal of Information Management*, Volume 48, 2019, Pages 63-71

Zaccaro C., (2014),"La storia di Eugene Goostman, il tredicenne che ha superato il test di Turing", *Linkiesta*, [Online], available at: https://www.linkiesta.it/blog/2014/06/la-storia-di-eugene-goostman-il-tredicenne-che-ha-superato-il-test-di/ (accessed: 21.10.2020)

Zuboff S., (2015), Big other: Surveillance capitalism and the prospects of an information civilization, Journal of Information Technology, 30, pp. 75-89