LUISS

Department of Business and Management
Major in Market Relationship & Customer Engagement

Course of Web Analytics & Marketing

Dark side of AI: ethical issues and bias in the context of recruitment.

Prof. Matteo De Angelis
SUPERVISOR

Prof. Rumen Pozharliev
CO-SUPERVISOR

Ilaria Bonsignore Matr. 747691
CANDIDATE

Academic Year 2022/2023

INDEX

INTRODUCTION

Artificial Intelligence (AI) has become an increasingly prominent and transformative force in various aspects of our lives. Its applications and potential benefits have captured the attention of managers across industries, prompting them to explore how AI can enhance efficiency, decision-making, and competitive advantage. However, alongside its remarkable advancements, AI also presents ethical challenges and potential biases that demand careful consideration. In the context of recruitment, where decisions regarding the selection and evaluation of candidates are critical, the dark side of AI and its implications become particularly significant.

The decision to address this topic for the final work of my thesis has matured because of the curiosity related to the theme of Artificial Intelligence that is still much debated today among those who think that Artificial Intelligence is the best thing that human beings could invent, and others who think it's the worst.

According to the most optimistic view, artificial intelligence will bring the human being to a higher state by making his life better in all respects and eliminating inequalities between people. On the contrary, according to the pessimistic view, artificial intelligence technology will lead the human race to extinction.

In the lively debate that has involved practitioners on the subject of artificial intelligence is not always given the right importance to the functioning of algorithms. Thus, knowing how this technology works is crucial to understanding how it could evolve and really at what point its development is. Once understood how it works, and its real applications, it is possible to have a debate on the actual developments it may have in the future.

The first chapter of the thesis sets the stage by establishing the managerial relevance of AI. It begins with a brief historical overview, tracing the evolution of artificial intelligence from its early beginnings to the present day. Following that, the chapter defines AI, providing clarity on the concept and its various dimensions. It explores the impact of AI through data and statistics, unveiling the profound influence it has had on industries, economies, and societies. Moreover, the chapter highlights the main benefits and opportunities that AI brings to the table, enabling organizations to automate processes, improve decision-making, and unlock new insights. On the other hand, it also acknowledges the threats and potential risks associated with artificial intelligence, ranging from job displacement to the ethical implications of its implementation.

Looking towards the future, the chapter shows the anticipated trends of AI, shedding light on the direction the field is headed. As AI continues to evolve, it becomes imperative to examine the ethical considerations surrounding its integration into various domains. In this regard, the chapter explores the possible combination of ethics and AI, questioning whether it is feasible to reconcile the two and mitigate the potential adverse effects. It further delves into the dark side of AI, addressing the limits and ethical implications that arise from its deployment. This section serves as a critical reminder that while artificial intelligence holds immense promise, it is not without its flaws and challenges. To deepen the understanding of the ethical implications, the chapter specifically focuses on the issues of data, privacy, and bias. AI systems heavily rely on vast amounts of data, raising concerns about the ethical collection, storage, and usage of personal information.

Additionally, biases embedded in AI algorithms and datasets can perpetuate social and demographic disparities, leading to discriminatory outcomes. To illustrate the real-world impact of these concerns, the chapter presents case studies on COMPAS and Amazon. These case studies shed light on instances where AI systems in the context of criminal justice and hiring processes have faced scrutiny due to biased decision-making and potential ethical violations.

Moving forward, the second chapter conducts an in-depth literature review that critically examines the ethical implications of AI, particularly in the recruitment context. It explores the intersection of AI and consumer patterns, providing insights into how AI algorithms influence decision-making processes and consumer behavior. Furthermore, the chapter emphasizes the importance of addressing algorithmic bias in AI applications to ensure fairness and inclusivity. It also explores the future of marketing and how AI can revolutionize marketing strategies, highlighting the potential benefits and challenges that arise from this integration. Within the realm of recruitment, the impact of AI is explored, with a specific focus on its role in human resources and the use of learning algorithms for candidate screening. To guide the research conducted in this thesis, at the end of the second chapter are presented the research questions and conceptual model that underpin the subsequent empirical studies. These research questions aim to investigate the ethical issues and biases associated with AI in the context of recruitment, providing valuable insights for both academics and practitioners.

In the third chapter, regarding the methodology, the research approach and design are detailed, with specific descriptions of Study 1 and Study 2. The results obtained from both studies are then analyzed, leading to a comprehensive discussion of the findings. The managerial implications of the research are highlighted, offering recommendations for leveraging AI in recruitment processes while addressing the associated ethical concerns. In addition, the chapter acknowledges the limitations of the research and suggests avenues for future exploration to further advance our understanding of the dark side of AI in recruitment.

In conclusion, this thesis aims to explore the ethical issues and biases arising from the integration of AI in the context of recruitment. By examining the managerial relevance of AI, delving into its history, definition, future trends, and ethics, this thesis provides a comprehensive foundation for understanding the multifaceted nature of artificial intelligence. The literature review further enriches this understanding by exploring AI’s ethical implications, algorithmic bias, and its transformative potential in marketing and recruitment. Through empirical studies, this thesis seeks to shed light on the dark side of AI in recruitment, offering insights to inform ethical decision-making and advance responsible AI practices.

CHAPTER 1: MANAGERIAL RELEVANCE

1.1 A brief history of Artificial Intelligence

When the topic concerns the future, artificial intelligence is always mentioned, whether it is in the medical field, economic, financial, industrial or social. The truth is that AI is not the future, but the present. Surely, it is not even known how many tools that exploit algorithms are used throughout the day. Yet, when people select the Spotify suggested playlist, they are using a tool based on recommendation systems that use algorithms. When Siri is asked to call a friend, when Alexa updates on the news happened or on the weather forecast or even when Netflix recommends which movie to watch. In all in these cases there is artificial intelligence, in fact all the most modern platforms are based on algorithms of analysis, prediction and recognition; among these Google, Facebook, Amazon and the aforementioned Spotify and Netflix. To date, systems using algorithms are widespread in all industries; with the aim of reducing the human efforts, speed up time and get more accurate results.

Before continuing with the present, it is good to take a look at the past by defining when and how artificial intelligence was born. Assistant mechanics have been a part of our culture since times of Homer, when he wrote about mechanical tripods waiting for the gods at dinner. However, only in the second half of the last century is there official talk about artificial intelligence. Logically, this recognition did not emerge from the theoretical vacuum, in fact it is the result of numerous investigations into the nature of intelligence and various attempts at machine production that would reduce the fatigue of man's intellectual effort, while eliminating the errors to which it is subject. When discussing the history of AI, is often made reference to the advent of the first electronic calculators and cybernetics, thus describing the discoveries prior to 1956, the year of official birth of the discipline.

In this context, the studies of the American author Pamela McCorduck (1979) were fundamental: indeed, she focused her research on technological progress in understanding the human mind. According to this tradition of research, "artificial performance is part of human practice as is natural performance, in the direction of a continuous attempt of man to imitate and reproduce himself". The formalistic tradition of inquiry into the mind and the man's tendency to self-imitate are elements that find a place in the projects of Charles Babbage (1792-1871). Among these, are significantly mentioned the Machine of Differences and the most advanced Analytical Engine; the first invention was designed to compile tables of logarithms mainly used in nautical calculations; however, the tables featured numerous mistakes that, sometimes, were fatal. The second machine was intended to calculate any math function, not just logarithms. In fact, Babbage understood that the Analytical Engine would have been able to play some strategy games, such as chess. However, actually this car was never built.

Many developments occurred in the context of World War II, when the first real machines were created. In particular, in the United States, a group of scientists had developed the ENIAC (Electronic Numerical

Integrator and Computer), a machine used for calculation of bombing tables, which made a significant contribution to the war effort. In Great Britain, a mathematician and logician who became a pioneer in the science of computer science and artificial intelligence emerged as a prominent figure: Alan Turing. In 1936, the mathematician published the article "On computable Numbers, with an application to the Entscheidungsproblem", in which he described for the first time what would become Turing's machine. During the World War II, Turing undertook the decryption of the codes used in the German communications, working within the Bletchley Park group of cryptographers. A famous case is that of the Enigma machine, a device used by the Nazi armed forces which guaranteed the security of German communications in the initial part of the conflict. Only recently it has become public knowledge how crucial this job was to avoid defeat in the first years of the war; in fact, the British government had imposed silence on all those who had participated in the decryption of codes (Garnham A., 2017). Following the war, funds for the development of electronic computers were returned available.

In 1950, Turing published the article on which subsequent studies on artificial intelligence were based: "Computing machinery and intelligence". In this writing, the Turing machine was described: Conceptual machine capable of being in a finite number of different states and executing a limited number of actions, in order to be able to express any type of defined procedure. With this article, still considered a point of reference today, Alan Turing laid the true foundations of what, six years later, would be called "artificial intelligence". Indeed, in 1956 John McCarthy, a Stanford researcher, uses the term artificial intelligence for the first time within a lecture at Dartmouth College. The aim of this workshop was to bring together researchers from various fields to create a new area of research planned at building machines capable of simulating human intelligence. In this context, artificial intelligence is recognized as a real scientific discipline.

The years following the Dartmouth workshop were filled with significant inventions in the field of AI. One of the first achievements was the General Problem Solver (G.P.S.) program, developed by Nobel laureate Herbert Simon and RAND Corporation scientists Cliff Shaw and Allen Newell, which was able to automatically fix some kinds of general formalized problems. A further example is the famous computer program ELIZA, created between 1964 and 1966 by Joseph Weizenbaum at the Massachusetts Institute of Technology (MIT). It is one natural language processing tool that can simulate a conversation with a human, in other words the first chatbot in history. Towards the end of the 1970s, numerous criticisms regarding the expenses for research and the questioning of the optimistic outlook of researchers led to a period of disappointments and governments ended their financial support (Haenlein M., & Kaplan A., 2019).

To talk about the turning point of artificial intelligence it is necessary to wait for 1997, when the program Deep Blue, designed by IBM (International Business Machines Corporation), wins a chess game with the world champion Garry Kasparov. This calculator, with an incredible computational strength, managed to process 200 million possible moves per second. After this victory, IBM engineers set out to design a new artificial intelligence system that could compete in more complicated challenges. In 2009, Watson is born,

who defeats all his opponents in the American TV quiz show Jeopardy, where the winner is whoever correctly answers the questions as quickly as possible (New York Times, 2009).
In the first decade of the 11th century, the first vocal assistants were developed: these are programs based on artificial intelligence algorithms that are able to interact with humans through voice commands. Apple's Siri assistant was released as an independent app in 2010 and integrated into iOS in 2011. In the following years, Microsoft introduced Cortana and Amazon launched Alexa. Finally, Google's smart speaker was announced in 2016, a golden year for the US online services company. Indeed, AlphaGo, a system of its own, wins against the world champion at the Chinese board game Go. This seemingly simple game presents an infinite number of possible moves, making it much more complicated compared to chess. Therefore, it was impossible for the system to codify all the moves, but, through the machine learning, AlphaGo was able to learn, programming itself based on experience (Di Fraia G., 2020).
Furthermore, in the last twenty years the automation of processes has taken place in various areas, including marketing, trade, finance. It was the turn of augmented and virtual reality, and biometric authentication systems, such as facial recognition and digital fingerprint scan. Artificial intelligence algorithms have become an integral part of social networks, in order to analyze the huge amount of data produced and generate new business models, monitor user comments, understand their behavior and feelings. Finally, in recent years, there has been a lot of discussions about quantum computing, referring to computers which, through the laws of physics and quantum mechanics, could give life to new perspectives and discoveries in the field of AI (Pictet, 2020). The technological progress that has characterized the last twenty years has been possible thanks to multiple elements, among which emerge the development of computational power and exponential growth.

1.2 Definition of the phenomenon: what is AI?

AI is first and foremost a universal field that encompasses many sub-categories, some more generic, such as learning or reasoning, while others more specific, like playing chess or proving mathematical theorems. In literature, when it comes to artificial intelligence, stands out the definition given by Stuart J. Russell and P. Norvig in the book "Artificial Intelligence: A Modern Approach", where AI is defined on the basis of four elements: action, man, thought and rationality. Starting from this assumption and from the fact that the research of AI must be partly a science related to psychology and, on the other hand, a science related to mathematics and engineering, four different combinations can be identified (Stuart J. Russell, & Norvig P., 2020):
1. systems that act like a human;
2. systems that think like a human;
3. rationally thinking systems;
4. systems that act rationally.
The first combination refers to the study and construction of machines predisposed to perform activities that would require the use of intelligence if performed by a human. To evaluate this conformity, in 1950 by Alan

Turing (1912-1954) was proposed a test able to determine whether a machine behaves like a human being. To pass the Turing test, a machine should have the following capabilities: natural language processing to successfully communicate, memorization thanks to the representation of knowledge, automatic reasoning to answer questions and draw conclusions, and machine learning to adapt to new situations. Later, other researchers proposed a total test that, in addition, predicts a physical simulation of the human. “Artificial intelligence is the study of how to make computers do things that, right now, men do better” (Rich E., & Knight K., 1991).

The second approach, namely systems that think like humans, concerns the comparison and correspondence between the timing and reasoning sequences of the machines and those of humans. This is to say that, with regard to similar questions, the reasoning performed by systems equipped with artificial intelligence must be analogous to that of a human being. An important role in this context is that of sciences and cognitive neurosciences, disciplines that study the areas of human brain and its modalities of learning. These insights have allowed the development of AI models based on neural patterns and, therefore, much more complex. “Artificial intelligence is the study of how to make computers think in the literal sense of the term” (Haugeland J., 1985).

The third notion, which refers to rationally thinking systems, is based on logic. Logic, already studied at the time of great philosophers, especially Aristotle, concerns the processes of reasoning defined as irrefutable. Towards the year 1965 emerges a logistic current that translates into the functionality of programs to solve problems, describing them in logical notation. However, logic requires total and certain knowledge of the world, a condition that cannot be achieved in all domains. Nevertheless, accurate reasoning originating from uncertain information is made possible by probability theory. “Artificial intelligence is the study of mental faculties through the use of computational models” (Charniak E., & McDermott D., 1985).

Finally, the combination that merges action with rationality takes place in an agent who perceives and consequently acts in the right way, optimizing the achievement of a goal. Assuming that perfect rationality does not exist in too complex environments and that the computational demands are often too high, the question of bounded rationality arises, which is made explicit in acting in the most appropriate way possible and in designing the best possible programs based on available computational resources. “Artificial intelligence is the study of how to explain and emulate intelligent behavior through computational processes” (Schalkoff R. J., 1990).

A more intelligible and accessible definition, even to the less qualified, has been disclosed by European Parliament on the official website. Artificial intelligence is defined here as “the ability of a machine to display human capabilities such as reasoning, learning, planning and creativity”. The explanation continues by illustrating the AI functionality, which “allows systems to understand their environment, put themselves in relationship with what it perceives and solve problems, and act towards a specific goal” (European Parliament, 2021). As is well known, for a considerable part of public opinion this innovation corresponds to worrying scenarios and represents a real threat to humanity. To counter this mistrust, the term augmented rather than artificial intelligence is being used more and more frequently in different contexts. This is to highlight the centrality of man within decision-making processes and reaffirm the support function of AI, aimed at increasing human intelligence. However, the processes of choice and use of the aforementioned terminology must be accurate and specific; in fact, these two concepts are not synonymous at all.

1.2.1 Data and statistics on the impact of AI

AI has the potential to deliver great business value for organizations, and its adoption has been sped up by the data related challenges of the pandemic (Forbes, 2022). According to Forrester, it is esteemed that nearly 100% of organizations will be using AI by 2025. In 2030, the artificial intelligence market is expected to reach 1,597.1 billion dollars, growing at a CAGR of 38.1% from 2022 to 2030 (Bloomberg, 2022). In addition, the IBM report "Global AI Adoption Index 2022" reports that 34% of companies claim to use artificial intelligence in their business. Conversely, a further 42% of respondents say they are exploring AI systems.

To date, the most advanced country is China, with an overall adoption rate of 88%. Italy is fourth, preceded by China, Singapore and India (Il Sole 24 Ore, 2022). Furthermore, the data in the IBM report shows that larger companies are twice as likely to have actively implemented artificial intelligence. While smaller realities have greater chances to explore or not pursue AI.

According to a recent survey of Harvard Business Review Analytic Services, 31% of respondents who use AI in sales and marketing say it has increased their revenue and market share. Moreover, 67% of respondents agree that AI in marketing and sales will be critical to their company's ability to compete in the future (Harvard Business Review, 2020). Indeed, according to Deloitte, 46% of organizations plan to implement AI in the next three years. Among the key drivers of AI adoption, IBM identifies the growing accessibility of technology, the need to reduce costs and automate key processes, and the increased implementation of AI in standard business applications (Il Sole 24 Ore, 2022).

Although AI use cases are increasingly diverse today, customer-centric applications remain the most common. A survey conducted by Statista reports that: 57% of respondents state that customer experience is the main case of AI use; 50% say they use them to generate customer insights and intelligence; 48% for customer interaction. By spreading AI initiatives, companies are getting more value from their investments. According to IBM research, 30% of global IT professionals say their employees are already saving time with new AI automation software and tools. Additionally, a study by Deloitte reports that organizations that adopt intelligent automation are expected to reduce costs by an average of 31% over the next three years. It is also predicted that 45% of total economic gains by 2030 will be the result of AI-driven product improvement: the latter will also be able to stimulate consumer demand. Therefore, all geographic areas will benefit from artificial intelligence (PwC, 2020).

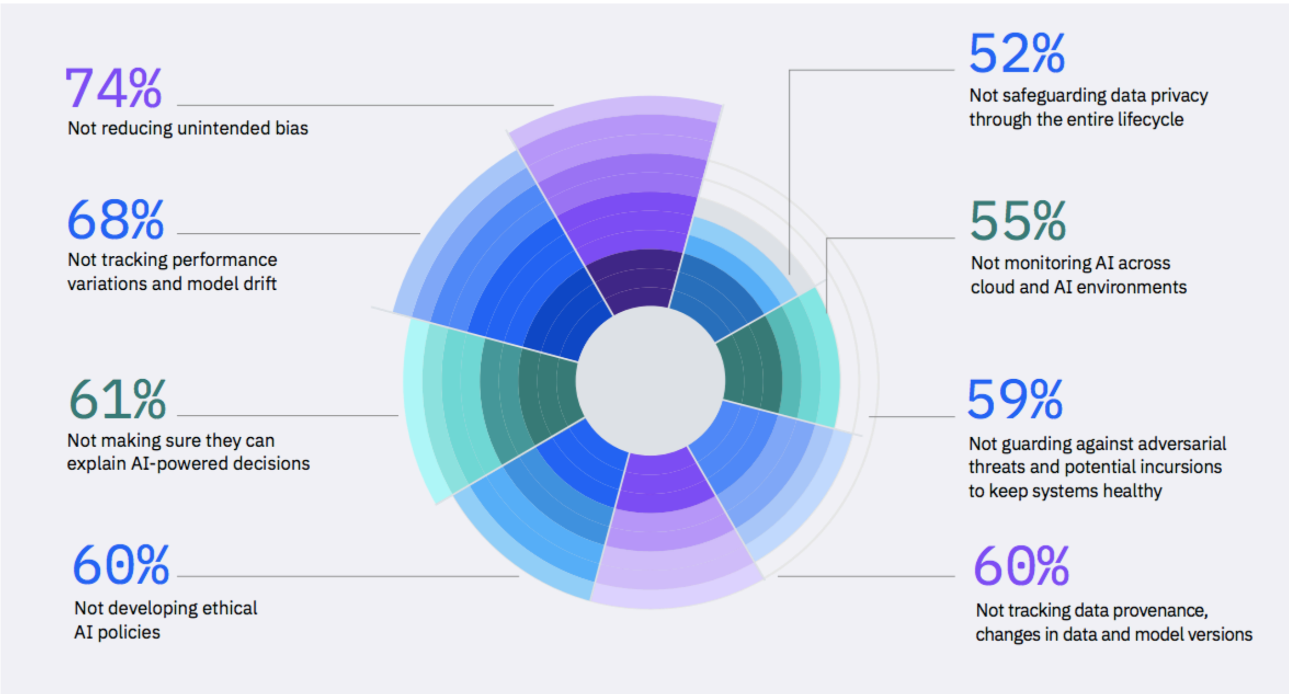
However, the exponential growth of this market is putting many companies in difficulty which, trying to benefit from the many advantages in this field, are unable to find qualified personnel. This shortage has the potential to hold back economic growth and digital innovation. According to the IBM "Global AI Adoption Index 2022" research, 34% of organizations consider insufficient AI skills, competences or knowledge as the main reason blocking their successful adoption. With the belief that a specialized team can exploit artificial intelligence to its full potential, 68% of companies are budgeting for the retraining and updating of existing employees; 58% to identify and recruit qualified talent from other companies and organizations; 49% for hiring college students (Statista, 2022).

Among the most requested specializations by companies are: coding, programming and software development (35% of companies), understanding of governance, security and ethics (34%), data visualization and analysis (33%), advanced degree in a field closely related to AI (27%). When it comes to the soft skills needed to fill these roles, 37% of respondents in IBM's survey "Addressing the AI Skills Gap in Europe" believe that problem solving is the most critical soft skill and 23% of tech recruiters struggle to find candidates with this attitude.

According to Statista and IBM, the main challenge that prevents AI from reaching their potential is the inability to develop its projects. 85% of IT professionals agree that consumers are more likely to choose a company that is transparent about how its AI models are built, managed and used. In this regard, the IBM research "Global AI Adoption Index 2022" reveals that the majority of organizations have not taken key measures to ensure that their AI is reliable and accountable.

Top issues include failing to reduce unintentional bias (74% of firms), not accounting for performance variations and model drift (68%) and not making sure the possibility to explain AI-based decisions (61%).

Figure 1. A majority of organizations haven't taken key steps towards trustworthy AI.



Source: IBM Global AI Adoption Index, 2022

As can be seen from these statistics, every company, regardless of the sector, has an infinite number of scenarios for adopting AI and high probabilities of success in case of pursuing the initiative to employ them within their processes. The progress in the use of artificial intelligence, aimed at obtaining tangible economic gains, needs to be supported by a holistic approach: a vision, that does not focus on the implementation of scattered solutions in response to specific needs, is therefore essential. It is necessary to consider these two elements as factors of business transformation, which can lead to improved decision-making, modernization of current systems; and which will be, in the not too distant future, essential for any company (Peruzzo A., 2022).

This has definitely been a positive year for the Artificial Intelligence sector, characterized by continuous progress in machine capabilities and by the exploits of “Dall-E2” and “ChatGPT”, which involved tens of millions of people in just a few weeks. In Italy, the AI market in 2022 reached 500 million euros, with an annual growth of 32%, of which 73% commissioned by Italian companies and 27% represented by project exports. Currently, 61% of large Italian companies have already launched at least one AI project, 10 percentage points more than five years ago. And among these, 42% have more than one operational. Instead, between SMEs, 15% have at least one AI project started (in 2021 it was 6%), almost always only one, but one in three plans to start new ones in the next two years. These are the results of the research made by the Artificial Intelligence Observatory from Polytechnic of Milan presented on the occasion of the conference “Artificial Intelligence: the age of implementation”.

Moreover, 93% of Italians have already heard of “Artificial Intelligence”, 55% affirm that AI is very present in everyday life and about 4 out of 10 (37%) in working life. But there is no shortage of perplexities: 73% are concerned, especially on the impacts on the world of work, even though only 19% of the population is firmly against the entry of AI into professional activities. The daily experience of Italians is centered on recommendation systems and virtual assistants. Specifically, chatbots, already used by 81% of people, are now almost as widely diffused as voice assistants (83%). In addition, interest in the recommendations received from AI engines for e-commerce is even increasing and one in four users has made a new online purchase after using them (Mc Kinsey, 2022).

Furthermore, according to the Accenture research report, artificial intelligence it is becoming a priority for the business; in fact, companies from the year 2017 to the year 2019 have spent $306 billion on AI. In accordance with the data, companies using AI technologies can achieve approximately three times the return on investment (ROI). In detail, the percentage of Italian managers who consider AI essential for their business is greater than the world average (it is 88% of Italian managers in contrast at 84% of the global average). In the marketing and communications sector, this percentage decreases slightly globally: 82% of managers consider AI fundamental and 71% believe that their business will be at risk without scaling AI (Deloitte, 2022).

1.2.2 Main benefits and opportunities

To get a general overview, it is important to understand what the main benefits of artificial intelligence are. It is clear that artificial intelligence is increasingly present in everyone's daily life. This can present risks but also benefits for businesses in several areas, such as security, democracy and employment (Kopalle P.K., 2022). Europe's prosperity and economic growth are tightly linked to how data and connected technologies will be used. It must be said that AI can make a big difference in terms of impact, positive or negative. For this, the European Parliament set up a committee that looked into the impact of the technology and proposed a long-term EU roadmap towards AI. Therefore, it can be useful to understand what the opportunities and risks of future applications of artificial intelligence are (Floridi L., 2018). EU countries are already strong in business to business applications and digital industry. Indeed, with a high-quality infrastructure and a regulatory framework that protects freedom of expression and privacy, the EU could become a global leader in the data economy and its applications.

Regarding the opportunities of AI for citizens, artificial intelligence could mean better healthcare, safer cars and other transportation systems, and even tailor-made, cheaper, and more resilient products and services. It can also facilitate access to information, training and education (Mc Kinsey, 2020). With the COVID-19 outbreak, distance learning has become a necessity. AI helps make the workplace safer, because the most dangerous work can be outsourced to robots, and offer new jobs thanks to the growth of artificial intelligence industries. Moving on to the main benefits of artificial intelligence for businesses, artificial intelligence can lead to the development of a new generation of services and products, even in sectors where European companies are already in a strong position, such as the circular economy, healthcare, agriculture, tourism and fashion (Financial Times, 2020). In fact, it can offer more fluid and optimized sales paths, improve machinery maintenance, increase both production and quality, improve customer service and save energy. Thanks to artificial intelligence, an increase in labor productivity is estimated from 11% to 37%, by 2035 (European Parliament, 2020).

Analyzing the benefits in public services, AI applied in this field can reduce costs and offer new options in education, public transport, energy and waste management, and could even improve product sustainability (Statista, 2022). For this reason, it would contribute to achieving the objectives of the European Green Deal. Moreover, it is expected a reduction from 1.5% to 4% in global greenhouse gas emissions by 2030, attributable to the use of AI (European Parliament, 2020). Another relevant aspect to consider is that data-driven audits, prevention of disinformation and cyber-attacks, and access to quality information can help strengthen democracy. In this way, diversity and equality of opportunity could also be supported, for example by easing hiring biases through the use of analytical data.

In addition, the use of artificial intelligence for security is very important, as it could be used in crime prevention and as an aid in criminal justice, because it would make it possible to process large volumes of data faster, assess the risks of escape of prisoners more accurately, even predict and prevent crimes and terrorist

attacks (Forbes, 2022). Besides, AI is already being used by online platforms to detect and respond to illegal or inappropriate online practices. Last but not least, in the military field, AI could be used for defense and attack strategies in case of cyber-crimes or for targeting key targets in cyberwarfare.
The broad range of AIs positively impacts the benefits businesses can achieve. Among the companies that have already invested in artificial intelligence projects, according to data from the Artificial Intelligence Observatory, 50% have the objective of reducing costs through process improvement. On the other hand, 37% aim to increase revenues, while the remaining 13% wanted to develop solutions to support the decision-making process (Il Sole 24 Ore, 2022). Among the companies that have launched AI projects, only 4% have not achieved the set objectives, while as many as 68% declare that they have projects underway that are proving to be very successful. The remaining 28%, however, is not yet able to express an opinion. In conclusion, these data highlight how companies are starting to see AI solutions as a real opportunity which, despite the complexities of start-up and implementation, can offer great benefits at a technological, economic and organizational level.

1.2.3 Threats and potential risks

Artificial intelligence has changed digital marketing through numerous applications, which optimize the conversion rate and make the overall funnel more performing. It is clear that these technologies represent an opportunity and more and more companies are adopting these technologies. However, together with considerable advantages, AI is a vector of limits and fears about, above all, people's lives. The most widespread concerns are related to the violation of privacy, fairness and transparency of algorithms, impacts on society and on man. Another issue that has been highly debated is the impact on jobs, which is also a legitimate concern according to statistics; indeed, according to a Forrester report, AI technologies and automation will result in a net loss of 7% of jobs by 2025 in the United States. However, this does not apply to the marketing sector, where the required skills will be progressively limited to the field of data science or creativity. (Conick H., 2017)
As has just been pointed out, the increasing use of AI systems also carries risks, in particular that of overuse and underuse of artificial intelligence (Floridi L., 2018). More in detail, not using artificial intelligence to its full potential is a risk: poor implementation of important programs, loss of competitive advantage over other regions of the world, economic stagnation and fewer opportunities for companies. Underutilization has several causes, starting from lack of public and business trust, up to poor infrastructure, lack of entrepreneurial initiative, low investments, fragmented digital markets. Regarding the latter cause, since AI machine learning depends on data, a fragmentation makes it less efficient. But abuse is also a problem. For example, it shouldn't be used for problems it isn't suited for, such as explaining or solving complex social issues (De Bruyn A., 2020).

Moving on to the topic of civil liability and artificial intelligence, it is often asked who is to blame. A major challenge is determining who is responsible for the damage caused by an AI-powered device or service: indeed, in an accident involving a self-driving car, the question is whether the damages should be paid by the owner, the builder or the programmer (Hage J., 2017). If the manufacturer is free from liability, there may not be sufficient incentive to provide a safe and efficient product. The public may have less faith in the technology. But at the same time, too strict regulations could discourage attempts at innovation.

Other possible threats of artificial intelligence are related to fundamental rights and democracy. The results AI produces depend on how it is designed and what data is put into it. This process can be influenced intentionally or unintentionally. For example, some important aspects may not be programmed into the algorithm or may be programmed to reflect and perpetuate structural biases (Akter S., et al., 2022). Furthermore, the use of data and numbers to represent a complex reality makes the AI seem factual, accurate and independent even when it is not, the so called "math-washing". For example, if not programmed correctly, AI could lead a decision regarding a job offer, loan offer and even in criminal proceedings, influenced by ethnicity, gender, age (Mc Kinsey, 2021). AI can also threaten data protection and the right to privacy. It can be used, for example, in devices for facial recognition or online profiling. Furthermore, it is capable of putting together the information it acquires about a person without their knowledge. The threat to democracy represented by artificial intelligence passes through information. It has already been accused of creating "bubbles" online, where content is presented based on content the user has interacted with in the past, rather than creating an open environment for multi-voiced, inclusive and accessible debate. It can also be used to create fake but extremely realistic images, videos and audios, known as deepfakes, which can be used to scam, ruin reputation and question trust in decision making processes (MIT Sloan Management Review, 2019). All of this, risks leading to the polarization of public debate and the manipulation of elections. Artificial intelligence could also threaten freedom of assembly and protest, as it could allow for the tracking and profiling of individuals associated with certain groups or opinions.

Moreover, it is also important to analyze the effect of artificial intelligence on the world of work. Indeed, the use of artificial intelligence could lead to the disappearance of many jobs (Wall Street Journal, 2021). Even if more and better ones are created, it is crucial that there is adequate training so that the unemployed can access them and that there is a skilled workforce in the long term. Another concrete threat of artificial intelligence concerns competition: since parties with more information could gain an advantage and seek to eliminate competitors, accumulation of information could also lead to a distortion of competition. There may even be security risk: more specifically, AI applications that are in contact with or even integrated with the human body can be dangerous if poorly designed, misused or hacked (Mc Kinsey, 2021). For example, unregulated use of artificial intelligence in weaponry could lead to a loss of control over destructive weapons.

Lastly, as regards threats to transparency, inequalities in access to information could be exploited to the detriment of users (The Economist, 2022). In fact, based on a person's online behavior or other data used without their knowledge, a service provider can predict how much this person is willing to pay for a service or a political campaign can know what message to send them. Another transparency issue is that it may not be clear to the user whether they are interacting with a person or an AI system.

1.3 Future trends of AI

The Artificial Intelligence market has been growing rapidly for several years, with the industry projected to reach $37 billion by 2025 (Forbes, 2022). This momentum will continue, and it is starting to realize this as new powerful AI-powered tools and services across all industries. There has been a shift from AI's well-known role in analytics and forecasting, helping data scientists and businesses make sense of the world and chart their paths accordingly, to new and innovative systems, that are producing completely new artifacts never seen before. But what is driving this exponential growth and how will it affect the space in the future? As AI tools make their way into mainstream business applications, there are five key trends to keep an eye out in the next years:

- Ongoing democratization of AI
- Role of augmented working
- Commercialization of generative AI
- Development of sustainable AI
- Priority of ethics in AI

The first is that, due to the fact that AI is becoming a key business differentiator, the democratization of AI will continue (Sudmann A., 2019). If a company can't find deeper insights into the data, quickly and at scale, its competitors will. Supply is far below demand, and the best engineering and data science talent will remain extremely expensive. As a result, more AI consultants and higher availability of low and no-code features will become differentiators. This democratization of AI will help make it easier for those with varying levels of experience to adopt these technologies across vertical markets. Additionally, cloud service providers will increasingly merge the building blocks of their services to include artificial intelligence, leading to powerful and widely available features and solutions (Deloitte, 2019). This is important for several reasons: more and more people will use AI and it is starting to realize the bottom-line business drivers of AI, which will shift from the aforementioned major cloud providers to smaller technology players, leading to even greater AI adoption.

Moving on to the next important topic, which concerns increased work, many people will find themselves working alongside intelligent robots and machines specifically designed to help them do their jobs better and more efficiently (Wall Street Journal, 2021). This could take the form of smart phones giving us instant access to data and analytics capabilities, as it was seen increasingly used in retail and industrial workplaces. It could mean augmented reality enabled headsets that overlay digital information. In a maintenance or manufacturing use case, this could provide real time information that can help identify hazards and safety risks, such as when a wire is likely to be live or a component might be warm. Management and leadership teams will increasingly have access to real time dashboards and reports, providing an immediate and up to date overview of operational effectiveness (McKinsey, 2018). AI-powered virtual assistants will also become increasingly popular in the workplace, able to respond quickly to questions and automatically suggest alternative, more efficient ways to achieve goals. Overall, developing the ability to work with and alongside intelligent and intelligent machines will become an increasingly indispensable work skill. It can even be said that it will go a long way towards mitigating the dangers of finding some roles redundant.

Another important trend is that generative AI will be commercialized (Statista, 2022). Indeed, it will be possible to see many more products and services on the market in the next years. This area is exciting because there are many largely untapped but valuable use cases. One particularly positive point is AI-based generative language applications. In games, for example, a user can choose to sound like their on-screen character. In a virtual meeting, a person with a cold can make their voice easier to understand, allowing people to focus on their work contributions rather than potential misunderstandings. Unlike AI-generated images, which have gained a lot of attention recently, there is a lack of business use cases. Speech-to-speech (S2S) technology, on the other hand, has the potential to change the way we work. For customer service, this can be a game changer. For example, contact center agents can use generative AI to clearly understand callers from anywhere in the world, helping them solve issues faster and feel more empowered.

Regarding the development of sustainable AI, in the short term all businesses will be under pressure to reduce their carbon footprint and minimize their impact on the environment. In this respect, the race to adopt and profit from AI can be both a blessing and an obstacle. Artificial intelligence algorithms, as well as all the infrastructure needed to support and deliver them, such as cloud networks and edge devices, require increasing amounts of energy and resources. A 2019 study found that training a single deep learning model can result in the emission of 284,000 kilograms of CO2 (MIT Technology Review, 2019). At the same time, technology has the potential to help companies understand how to build products, services and infrastructure more energy efficiently by identifying sources of waste and inefficiency. Ongoing efforts to roll out a greener, renewable energy-powered infrastructure are also part of the push to deliver more sustainable AI.

AI can also be a driver of sustainability in other industries and business areas: for example, computer vision is used in conjunction with satellite imagery to identify deforestation and illegal logging activity in rainforests, as well as the activity of illegal fishing, which impacts biodiversity in the oceans. As a result, there is expected to be a continued push towards implementing AI initiatives aimed at tackling some of the most pressing problems facing the planet, rather than simply seeking greater corporate profits.

The last trend and perhaps the most pertinent and relevant one is that AI ethics will become a top priority (Harvard Business Review, 2021). Despite its proven value and great potential, AI still has complex ethical and legal issues. The severity varies because new implications can range from negative to dangerous. From models that have degraded over time to deep fakes to biased algorithms, these are all frightening reminders that regulatory frameworks need to adapt to the rapidly evolving AI market. Thus, while regulatory and legal frameworks are currently in the works, with an AI Bill of Rights in the near future, companies need to approach

AI ethically and safely. The first US class action lawsuit against an AI system was recently filed, and it won't be the last. The technology may be one step ahead of the legal industry, but as AI integrates into our daily lives, businesses and governments need to take safe and responsible practices seriously. It will also possible to see more transparency around cases like this and learn how to avoid these missteps for future deployments. In the next years, there will be efforts to overcome the AI "black box" problem (Bloomberg, 2022).

Those responsible for implementing AI systems will work harder to ensure they are able to explain how decisions are made and what information was used to get there. The role of AI ethics will become increasingly important as organizations come to terms with eliminating bias and injustice from their automated decision-making systems (Belenguer L., 2022). Biased data has been shown to lead to biases in automated results which can potentially lead to discrimination and unfair treatment, which simply will not be acceptable in a world where artificial intelligence plays a role in decisions affecting employment and access to justice or health care.

Therefore, artificial intelligence changes the way marketing is done, not only in relation to the tools used, but also metaphorically, in terms of how customers and companies are considered. Important reflections follow from this assertion: will the difference between man and machine become lighter? Will the concept of identity change? Will it be possible to manage the technologies wisely and humanely?

Although it has been rumored for a decade, 2023 will be another year of strong growth for AI: commercialization of new features and products, advances in access and convenience, and a focus on responsible practices will open up disruptive use cases for the enterprise and beyond. It is an inspiring time to be in the AI space and it will be interesting to see how the industry progresses over the next future.

There is no doubt that artificial intelligence can offer the world immense advantages and great lucrative potential; however, there are errors of operation to be solved that especially concern ethical and moral principles. This topic will be extensively covered in the next paragraphs, where the main limitations of AI technologies mentioned above, possible solutions and applicable regulations, and the most important ethical and legal aspects will be discussed.

1.4 Ethics and AI: a possible combination?

Artificial intelligence is becoming more and more decisive in the life of twenty-first century society, and this is a fact (Craig S. Smith., 2021). Many questions arise related to this topic, including how far technology can go, whether it is appropriate to get used to the idea of living with sentient machines or whether artificial intelligence can be trusted. It is interesting to find out if there could be a possible union between ethics and artificial intelligence (Siau K., & Wang W., 2020).

First of all, it is necessary to clarify what is meant by ethics in the world of technology. In a publication of the Roman Observatory on the theme of technological ethics, an interesting analysis is tackled on the relationship between technology and power (Floridi L, 2018). In particular, it can be observed that in New York in the last century, urban development had a clear political vision: to improve the quality of life of the white population, the richest and in power, to the detriment of the poorest classes represented above all by ethnic minorities.

The example demonstrates how the availability of power, especially economic power, translates into the availability of increasingly innovative technologies. But this is not synonymous with ethical goals. The Observatory's analysis transfers the absence of ethics highlighted in the New York administration of the past to the current 'algorithmized' digital economy. Therefore, it is worth reflecting on this, especially with regard to artificial intelligence. One of the possible questions is whether a machine guided by algorithms can behave in an ethical way (Tsamados A., Aggarwal N., Cowls J. et al., 2022). This seems to be a very difficult point on which insiders are still wondering and struggling to answer.

A further starting point for reflection on the ethics of artificial intelligence emerges from the world of jurisprudence regarding the concept of intellectual property (Aristodemou L., & Tietze F., 2018). Although it may seem absurd to discuss such a question, the facts support a different opinion: AI algorithms are the brains of increasingly sentient machines, capable of pondering choices and deciding on the basis of plausible reasoning (Signorelli C. M., 2018). Naturally, considering the due limits with respect to human intelligence.

Therefore, the question that arises is whether a machine can own the copyright (Kretschmer M., et al., 2022). In reality, this science fiction scenario is not yet regulated by specific laws on the subject. There are many difficulties in managing such a particular case. First, the impossibility of attributing the intellectual property of any discovery to a machine poses an enigma to the competent legislators, namely who owns the copyright. To answer this question, it would be necessary to understand whether the object of the dispute is the result of human work with the assistance of artificial intelligence or is generated autonomously by AI (Barfield W., 2018). The trend seems to be the most logical one, which is to attribute intellectual property to the human behind the machine. However, even in this circumstance a difficulty of discernment can emerge: whether the copyright belongs to the developer of the AI algorithm or to whoever, precisely with the help of that algorithm, has produced new knowledge (Kop M., 2020). On balance, it is still very difficult to move in this nebulous context without ad hoc regulatory reference points.

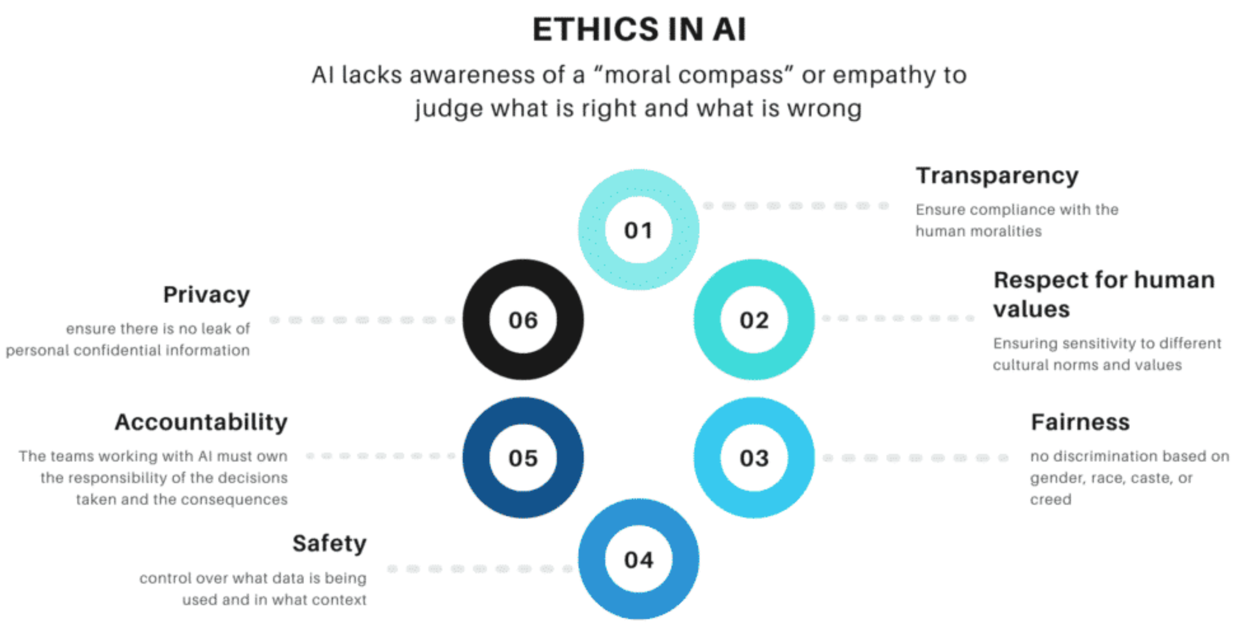
While waiting for more detailed feedback on the matter, a necessary distinction can be made between strong and weak artificial intelligence (Forbes, 2019). If the concept of strong AI is unfamiliar, it is probably worth asking what it means. A clear definition of this concept is provided by an American philosopher in his essay Minds, Brains and Programs: "According to strong artificial intelligence, in the study of the mind the computer is not just a tool; rather, a properly programmed computer is indeed a mind" (Searle J. R., 1980).

In other words, a strong artificial intelligence algorithm is capable of emulating human cognitive abilities in an authentic, almost imperceptible way. Its features are: the development of mathematical logic, at the basis of reasoning, problem solving and above all the ability to consciously demonstrate one's choices; the analysis and understanding of human language, even in complex forms; management autonomy in carrying out its activities. Of course, these are real robots in the typical sense of sentient life forms in the form of machines, also defined as "computers suitably programmed as real minds" (Searle J., 1980). And even if the famous

Turing Test, created by Alan Turing in the 50s, which seems to have been passed only rarely, could by now be obsolete for evaluating the potential of algorithms, in the not too distant future strong artificial intelligence could really put itself on a par with a human being (Thomas M., 2022).
Instead regarding weak AI, the main distinction with respect to strong AI lies in the goal of the developers: the intention is not to create machines similar to humans, but tools capable of easily carrying out complex tasks and operations. Weak artificial intelligence algorithms are therefore great problem solvers, capable of optimizing some demanding performances such as data profiling, translation or correction of texts and above all the elaboration of solutions based on analytical comparison.
Probably the question of the decade is whether AI can be trusted (Forbes, 2021). It is well known that trust in disruptive technologies is always a difficult issue, which requires a broad debate and the contribution of multiple disciplines. Artificial intelligence is no exception, especially in consideration of its humanistic aspects. As it has already been addressed above, the field of jurisprudence plays a key role in defining some fundamental rules in the management of these technologies, such as the attribution of property rights, the guarantee of privacy and information security. However, this may not be enough and many issues would still remain to be resolved, including: safeguarding the work of professionals threatened by the presence of AI algorithms; the excessive reliance of decision-making responsibility solely on algorithms; the negative effect of some biases that affect the judgment of the AI, for example for issues related to racism or sexism (Belenguer L., 2022). And the list certainly doesn't end here.
The question is how is it possible to build greater trust in artificial intelligence? (Brebant H., 2020). IBM offers a real Action Guide for all companies that intend to develop AI solutions by enhancing the ethics of algorithms. The AI Ethics in Actions report highlights some significant aspects about the role of ethics: 75% of executives believe that it is essential for the development of an increasingly human and performing artificial intelligence; 55% of companies have already activated training plans to retrain their employees and collaborators with respect to the use of AI; 67% of companies improve their performance in terms of social responsibility, sustainability and inclusion; three out of four innovators think it is important to differentiate themselves competitively from competitors. It is interesting to analyze in detail the three areas in which the IBM guide is divided: corporate strategy and vision, governance, implementation. In this first area, the ethical practices to be inserted appropriately in the strategic context of the company are defined.
The main points to consider are: the criticalities in creating a trustworthy AI to achieve the set strategic objectives, answering the questions about what values could be improved by AI and how its success is measured; the role of AI in an organization's growth strategy and its approach to innovation; the man-machine balance in the company. The second area, regarding the governance, establishes the management approach to be adopted in the company to allow the ethical development of artificial intelligence. The most important points covered are: the perspectives of stakeholders (managers, employees, customers, government and institutions), addressing relevant principles such as privacy, fairness and transparency; the role of the firm in a broader organizational and socio-political ecosystem; the risk profile for the AI and the processed data,

establishing a threshold level of acceptability. Finally, in the third and final area about the implementation the integration of ethics in algorithms and AI technologies is concretely addressed.
The key points to keep an eye on are: the interaction with the stakeholders, previously mentioned; the definition of processes for the acquisition, review of data and the monitoring of the activities performed; the formation of work teams and the enhancement of diversity as a key factor for the formation of algorithms; the definition of integrated methodologies and toolkits. The IBM Action Guide is an interesting starting point to understand the necessary steps to follow to reconcile ethics and artificial intelligence in the company. What to expect in the future? It is now a very topical theme, technological innovation is making great strides obtaining extraordinary results, and expectations are high. The time to act is now and the period seems in line for a new phase in AI creation (Mc Kinsey, 2018).
Artificial intelligence has the power to innovate and build almost anything. As technology advances, AI can be expected to create machines with decision-making and learning capabilities. However, what is missing is an awareness of a "moral compass" or empathy to judge what is wrong and what is right. There are some ethics elements that AI developers and designers can implement to instill a sense of accountability among people working with AI and mitigate bias within systems (Floridi L., 2018).

Figure 2. Ethics in Artificial Intelligence



Source: Campos J., 2020

The first principle to consider is transparency: the development of each AI-based equipment or AI algorithm must be transparent and include a complete purpose, explanation and justification for its development (Fjeld J., et al., 2020). These revelations are essential to monitor the outcome of the technology in order to ensure its conformity with human morality. The design should be structured in such a way that human beings can easily detect, understand and perceive its decision-making process. The second is based on the principle of respect for human values, since every AI innovation should lead to the overall growth of individuals and communities. Indeed, extreme care is required to ensure sensitivity to different cultural values and norms. Moreover, equity is essential to promote a fair workplace without discrimination based on gender, caste, race or creed (Agarwal A., 2022). Therefore, engaging team members from diverse cultural groups can promote inclusiveness and minimize bias.

The next principle, that of safety, refers to both the well-being of people and the security of user data. It is essential to identify the risks and work to find solutions to mitigate them. Using security practices such as encryption and allowing users control over what data is used and in what context can preserve and protect users' right to data. Responsibility is another fundamental principle, since teams working with AI need to take the consequences and the responsibility for the decisions they make (Dameski A., 2018). Consequently, decision-making processes must be reviewable, especially in cases where AI works with confidential and sensitive data such as intellectual property, personal health information, biometric data, identifiable data or national security information. User privacy should be a priority in any AI work process. It is necessary to ask for the users' consent to store and use their data. Great care must be taken to ensure that there is no dispersion of confidential personal information. In conclusion, to sum up, ethics constitute the foundation of any innovation and artificial intelligence is not exempt from this (Forbes, 2022). By adopting the ethical standards mentioned above, developers and designers working with AI solutions contribute to a safer technology-driven world with strong roots in human values.

1.4.1 The dark side of AI: limits and ethical implications

The one relating to artificial intelligence is a real ethical problem: in fact, despite the growth of the sector in the last year, the main criticality concerns trust in these technologies. It is possible to find out in detail what are the ethical challenges concerning the sector. First, it is necessary to consider all the internal dynamics that relate to the design of the algorithms and the related choices of the developers: presence of unintentional biases or distortions that damage the quality of AI activities; correctness in the use of algorithms towards stakeholders; transparency of decision-making processes regarding artificial intelligence systems and respect for privacy in accordance with current regulations. Secondly, are taken into account the external implications relating to social, economic and political systems and the people who belong to them: respect for fundamental human rights regarding freedom, both individual and collective, equality and equal opportunities.

Moreover, it is important to consider the safeguard of the emotional and professional well-being of workers according to the development of increasingly performing algorithms, even for complex tasks; environmental protection and adequate use of technologies, aiming towards a real reduction in the use of resources and related costs. Finally, building trust, i.e. a shared sense of trust towards AI in the planning and management of resources, in the operations of choice and in the learning of specific know-how to motivate the decisions adopted on the basis of clear and above all acceptable criteria. The centrality of the ethical question in the use of artificial intelligence is therefore palpable from different points of view. This topic was the subject of a very interesting speech where the following concept was emphasized: “machine learning as the main tool for teaching ethics to AI” (TED Talks, 2018).

In recent years the potential of artificial intelligence has increased considerably, as well as the implications related to it: in fact, with great power comes great responsibility. The consequences and risks associated with the use of AI in marketing, as well as in any other sector, have made it necessary the introduction of regulations and data ethics. Ethics, defined as the science that formulates values and generates principles, can be applied to the living of the human being, as well as to the action of artificial intelligence. The interactions between these two apparently distant worlds can extend to different levels:

• Ethics by design - technical/algorithmic integration for the formulation of judgments or reasonings that are ethically acceptable as part of the behavior of AI systems;

• Ethics in design - identification of regulatory and engineering methods that guarantee transparency, fairness and accountability of algorithms towards various stakeholders;

• Ethics for design - definition of codes of conduct and certifications that ensure the integrity of AI system developers and users during use or management. (Dignum V., 2018)

The ethical question can be introduced by the Latin phrase primum non nocere, which means “First, do no harm.”. This wording represents the central axiom of the coexistence between man and artificial intelligence; the other elements that favor this interaction can be summarized in four principles that AI must respect (Trezza R., 2020):

1. intuition - artificial intelligence technologies must understand humans and his intentions to foster an environment in which intellect and talent of human being are respected;
2. intelligibility - artificial intelligence technologies must be able to be understood by man, who must know how the machine works, especially for coexist safely with it;
3. adaptability - artificial intelligence technologies must value and respect both the rational and the emotional nature of human being, adapting to his personality;
4. adequacy of objectives - AI technologies must work together with man and assist him in achieving the goals established and classified by himself.

Europe is determined to take advantage of the opportunities offered by artificial intelligence in ethical and safe way, placing man and his fundamental values at the center of this picture. Through ethical and safe algorithms combined with a greater commitment of civil society, institutions and industries, it is possible to build a socially guided artificial intelligence and focused on man, as well as promoting a general feeling of trust towards this technology.

In 2019, the European Commission presented the targets to be achieved for encouraging an adequate development of AI; among the main ones emerges the increase in investments with an amount of over 20 billion euros per year, made available for the most innovative hubs. This choice was motivated by the Vice-

President and Commissioner responsible for the Digital Single Market, Andrus Ansip, according to which: "The ethical dimension of artificial intelligence cannot be considered a luxury option or an accessory complement: only with trust our company can take full advantage of technologies. Ethical AI is a win-win proposition, which can offer a competitive advantage to Europe, namely that of being a leader in the development of anthropocentric artificial intelligence that citizens can trust".

The approach taken by the European Commission to answer the main ethical questions started with a definition phase in which seven key requirements were established for the development of a reliable artificial intelligence:

1. human action and monitoring - artificial intelligence technologies must assist and support man, without ever limiting his autonomy;
2. robustness and security - algorithms must be secure and robust enough to handle possible errors;
3. privacy and data governance - users must not be harmed through an improper use of their personal data, which they must be able to check constantly;
4. transparency;
5. diversity, non-discrimination and equity;
6. social and environmental well-being - artificial intelligence technologies must work for the good of man and the planet;
7. responsibility also understood as accountability. (European Commission, 2019)

Therefore, the European code of ethics is mainly focused on the protection of dignity and physical, psychological and financial security of the human being. Furthermore, in April 2021, the European Commission established and published a new proposal for regulation on the European approach to artificial intelligence: the Artificial Intelligence Act (AIA). This proposal defines which are the prohibited applications of AI systems, such as those with high, limited or minimal risk.

Forbidden practices include the use of real-time biometric identification systems in spaces accessible to the public (except for strictly necessary reasons), the use of *social scoring*[1] on behalf of public authorities and the use of manipulative techniques that distort the user behavior, causing damage or exploiting vulnerability. Among the AI systems with high risk, there are, for example, those used for personnel selection and facial recognition in specific contexts; while among those at limited risk, chatbots and voice assistants. Finally, the minimum risk category includes the majority of AI systems, such as antispam filters, predictive maintenance systems or video games (European Commission, 2021).

However, according to some researchers, these regulations would not be sufficient to contain the risks related to AI technologies used to detect mental information and emotional traits of man. An example of this are the predictive algorithms of decision-making processes.

[1] The term *social scoring* indicates a system of social control and evaluation, subject to a mechanism of rewards and sanctions, used to classify citizens through a social credit. It is an initiative of the Chinese government born in 2007.

Everything that has been illustrated and described up to now is realized in the concept of innovation. In most circumstances, innovation means renewal, opportunity, productivity and prosperity, in other words is identified in the idea of a positive change. But often when there is innovation, as Steve Jobs said in the past, mistakes are made. According to the most famous American entrepreneur and computer scientist, it is necessary to promptly admit it and continue to work to improve. As with all innovations, even that related to artificial intelligence and its use in various sectors brings with it numerous doubts, questions and shaded areas. Revolutions often catch people unprepared and, for this reason, more training and education are needed to understand and manage them in the best possible way. Furthermore, the revolution in question implies numerous changes which concern, in particular, a reorganization of both the economy and society.

According to several experts, including MIT professor Daron Acemoglu, the consequences of an uncontrolled development of AI could be really risky: "I argue that if AI continues to be employed along its current trajectory, without being regulated, could produce various social, economic and political damages. (...) It may be useful to understand them before they are fully realized and become more difficult or even impossible to reverse, precisely because of the promising and broad potential of AI. I also suggest that these costs are not inherent in the nature of AI technologies, but are related to how they are used and developed at the moment" (Acemoglu D., 2021). As previously explained, the use of AI technologies is becoming increasingly popular also in the world of marketing, a sector in which questions and concerns are widespread. The main problems, which are sources of relevant discussions, mainly concern the accuracy of data acquisition and interpretation, consumer privacy, algorithmic biases that materialize in prejudices and discrimination, governance and the future of the working world.

It is well known that no technological advance has ever replaced the previous one: it is always gone by evolution or accumulation. Will it also be the case with artificial intelligence, such a powerful and incisive evolution that it is sometimes considered risky?

1.4.2 Data, Privacy and Bias

The algorithms used in machine learning systems and artificial intelligence can work optimally only in presence of large amounts of high-quality data. This necessity sheds light on two important actions: the acquisition and interpretation of data, which are presented as the main resource of the new digital economy. Data is lifeblood for AI and the energy that makes algorithms work. In marketing, data are essential for predicting preferences and consumer behavior: as a result, they are able to create more similar products compared to human expectations and needs. However, there are several sources of error in data acquisition processes and the artificial intelligence systems that use this incomplete or distorted information could lead to inaccurate results which may violate fundamental rights of people. The principle "garbage in - garbage out", often used in the computer field, even applies in this context; according to that, low-quality data lead to low-quality results. According to the EU Agency for Fundamental Rights (FRA), the poor quality of data leads to two major sources of mistakes, measurement and representation errors, which are linked to two important conditions for data quality: accuracy and validity. The first concept refers to the stability and consistency of measurements; in fact, unreliable data lead to results with high variance and uncertainty. While, invalid data matches wrong measurements and, therefore, are distorted (EU Agency for Fundamental Rights, 2019). Due to the large number and diversity of data sources, it is difficult to measure quality; however, some can be considered criteria that correspond to four essential dimensions: availability, usability, reliability and relevance. Each dimension includes key elements; in particular, when it concerns the reliability of data these are:

• accuracy - the real value of the data verified through a comparison with a known value of reference;

• consistency - the logical relationship between data must be correct and complete;

• integrity - complete and standardized content and format according to a given model;

• completeness - relating to a piece of data that has multiple components, each of which must have a valid value. (Cai L., & Zhu Y., 2015)

An additional concern is always about the data and, in particular, their use sometimes inappropriate, which may violate some fundamental rights of men, such as the right to privacy or to the protection of personal data. In the new digital world, people are constantly leaving traces that can be traced back to them by subscribing to social networks and newsletters, accepting terms of use, downloading applications. They often think they are doing it for free but, in reality, they are paying with their own personal information. Improper use of data is certainly not fault of artificial intelligence and neither of the user, who, however, can be blamed for a great lack of awareness or a remarkable laziness, when he decides not to read the implications of his actions on the web. AI technologies have made it possible to acquire, analyze and combine masses of data before unthinkable, but above all, they authorized some people to use them, putting power of information in their hands. Certainly, data can be exploited in an appropriate way, for example in marketing sector they are developed to give advice to help consumers. But, as AI ghostwriter Ester Liquori points out: “However, information can also be misused, to favor governments or political maneuvers, (...). Or even worse to discriminate or prevent the equal progress of individuals. It can be used to identify, to monitor individuals, (...) not just their purchase behavior but also tastes and preferences that can bring data to light that fall within the categories of, so defined, “sensitive data”: health, sexual or political orientations” (Liquori E., 2020). This information can be used to manipulate the consumer and to benefit companies at his expense. As for the first concept, current AI technologies allow to know consumers’ preferences much more accurately than they do themselves. This makes possible potential manipulations; for example, by estimating the periods of primary vulnerability, companies market products, even of low quality, which tend to be purchased during such times, increasing profits.

In reference to the second concept, companies that use artificial intelligence technologies for data acquisition can reach better positions than the competitors and have more power (Acemoglu D., 2021). However, this creates competition that is deemed unfair, which takes the form of engaging a surplus of customers and, as a consequence, better economic results. In addition, there is a social dimension of data that introduces two interconnected effects:

1. when a user shares personal information, at the same time he provides information about other individuals. Certain information, when have an impact on privacy, creates negative externalities;
2. as more information about an individual is shared, it loses value and become less important both for the user himself and for those who will acquire such data.

These concepts formed the foundations of the Cambridge Analytica scandal in March 2018, when personal data of about 87 million Facebook users were extrapolated for political purposes. The information was gathered through a personality test proposed below form of application, downloaded by around 270,000 users. These, by accepting the conditions, have shared not only their information, but also that of their contacts. This is because, until 2014, users who were allowed to view other people's information could share it likewise with the other applications they used. In the same year, the EU General Data Protection Regulation (GDPR) was established, with the aim of regulating and standardizing the personal data processing policies throughout Europe, according to principles that respect fundamental freedoms. Despite the supplementary regulations and this great achievement, the characteristics of which will be illustrated later, there are still problems related to the processing of personal data and, in particular, to the acceptance of cookies, files of text that memorize the actions performed on a particular web page while browsing (Sky tg24, 2019). One of the last episodes dates back to December 30, 2021, when the CNIL (French authority for protection of personal data) has fined Facebook and the company that controls Google for having made difficult the refusal of cookies in a conscious and intentional way. Moreover, starting on January 9, 2022, the new guidelines on the use of cookies issued by the Privacy Guarantor will come into force. As can be seen from the infographic created by the GPDP: "the new guidelines suggest some improvements to be adopted in order to provide users with information that complies with the requirements of transparency provided for by the Regulation". More in detail, the above-mentioned information must present a simple language and must be promoted through interactions on different levels with informative pop-ups, chatbots, voice assistants and other systems (GPDP, 2022).

One of the shaded areas that most creates fears and perplexities concerns algorithmic biases. The term bias refers to prejudices that can be introduced into machines during their setting; this can lead the algorithm to make wrong and discriminatory decisions, for example on the basis of elements such as gender, age, ethnicity or sexual orientation. As indicated by the EU Agency for Fundamental Rights, episodes of discrimination due to artificial intelligence are increasing. Indeed, according to their statement, hiring algorithms generally prefer men over women; moreover, gender biases also affected machine translation systems, while some facial recognition systems did not work equally well with individuals of different ethnicities (EU Agency for Fundamental Rights, 2019). In summary, the conditions that are particularly affected by the effect of algorithmic biases relate to equality between men and women, access to fair trials in the field of criminal justice, private and family life.

It is necessary to highlight that these results are mainly due to the data used for train machine learning systems. This is biased data that is entered by humans themselves and that reflect what is the current society; artificial intelligence amplifies essentially this reality. Alessio Semoli, expert in digital marketing and analytics, is convinced: "The important thing is to know that machines are not guilty: harmful traits, which can lead to a growing inequality, are a responsibility of the invisible hand that created them. In order to avoid the bias prejudicing the workings of the machines, we must begin to fight them, by designing artificial intelligence systems and coding robots with data reflecting a wider range of human experiences and interests. Removing bias from robots is a huge task, which will require time" (Semoli A., 2019).

The difficulty in correcting algorithmic biases mainly depends on whether they are ignored by both man and machine until such time as an independent review takes place; moreover, the exponential growth in the amount of available data makes it more and more difficult to identify the source of error. Referring to concrete cases, one of the most discussed concerned Google in 2020, when AlgorithmWatch (non-profit research and advocacy organization engaged in analysis and evaluation of algorithmic decision-making processes) highlighted a racial bias in the computer vision service Google Vision Cloud. In fact this system, depending on the skin tone, labeled a portable thermometer differently: in the picture where an individual of skin dark held the thermometer this was labeled as a "gun", while in the same image where the individual, however, had a lighter skin tone, the thermometer was labeled as an "electronic device". What happens is linked to the more recurring presence of dark-skinned individuals within violent scenes that are part of the data set for the training of artificial intelligence systems. For this reason, it is more likely that an algorithm associates people with darker skin tone to terms related to violence. There are many other examples similar to the one illustrated; for example, Facebook's computer vision system has prevented a user from posting a photo by stating that the proposed content was aggressive. In reality, it was simply a drawing in which the subjects presented dark skin (Kayser-Bril N., 2020).

Furthermore, an additional phenomenon that can generate algorithmic bias is data poisoning: a data pollution or poisoning caused intentionally by hackers experienced in AI. The algorithm, learning from this distorted or misleading data, can draw wrong and often harmful conclusions. In the marketing world, incurring algorithmic bias mostly means not being able to identify the right audience, decrease sales opportunities, worsen economic results and risk suffering reputational crises (Zampori I., 2021). In order to mitigate algorithmic biases, it is a must to develop technologies that can identify and constantly monitor them; but the original problem concerns society and must be solved over time through investment in education and formation.

According to Luciano Floridi, artificial intelligence is a divorce, in the sense that with it "the skills to act successfully and the need to be intelligent to do so, get divorced". The success of this divorce mainly depends on two factors (Fiore P., 2021):

1. the environment, which must be able to accomodate this technology and redefine its problems, so that they can be solved by AI;

2. the human being, who will have to significantly develop one of the most important life skills: the awareness. Unfortunately, in this scenario, it is to fear the one who “has his hands on the wheel”, namely the human being, who is often driven by selfish feelings aimed at an avid and individual enrichment even at the expense of other individuals. Therefore, it is necessary to promote a right awareness to exploit most of the potential offered by AI, to limit the risks associated with its improper use, to promote the well-being of humanity and the planet. The optimal cooperation with artificial intelligence can effectively facilitate the protection of the environment and society: “On the one hand we have AI, an extraordinary, flexible force, with a great ability to solve problems; on the other we have many social and environmental problems to solve. Let’s put one and the others together, at the service of sustainability and a more equitable society, which knows not only how to generate wealth but also how to distribute it. With the blue of digital technologies and the green of the environment we can do more with less, differently and better” (Floridi L., 2021).

On the basis of the considerations reported up to these last lines, it is possible to affirm that there are mainly three axioms on which the future depends for an ideal coexistence between man, environment and artificial intelligence. These are awareness, to understand the present and design the best possible future, responsibility, which will determine the success or failure of this perspective, and humanity; an aspect which, perhaps, can give rise to greater concern with respect to some algorithms and new technological combinations.

1.5 Case studies: COMPAS and Amazon

Instruments based on artificial intelligence have become an integral part of our lives. From judicial systems to advice from search engines. But there’s growing concern around AI bias: situations where AI makes decisions that are completely unfair to specific groups of people. Indeed, researchers have found that AI bias has the potential to cause real harm. However, is it possible for the algorithms behind these useful applications to suffer from an old defect of human beings, namely prejudice? It is interesting to understand how machine learning algorithms work, and that ethics is essential if we are to continue to build increasingly intelligent machines.

In order to better understand the issue overexposed may be useful analyze the case of COMPAS[2] Recidivism Algorithm. One of the most discussed cases in the field of predictive justice is undoubtedly the one involving Mr. Eric L. Loomis. In 2013, Mr. Loomis was charged with five counts, all in recidivism, and was sentenced to six years in prison, following a shooting incident, in which the police were also involved, in La Crosse, Winsconsin (Harvard Law Review, 2017). In preparation for the ruling, the Wisconsin Department of Corrections produced in court an assessment of the risk of recidivism generated by an AI-based software called COMPAS. This system estimates the risk of recidivism based on the criminal history of the offender and a criminal reference database. Since the methodology behind COMPAS is a trade secret, the Department of Corrections only referred the estimates of the risk of recidivism to the court.

[2] COMPAS is the acronym of “Correctional Offender Management Profiling for Alternative Sanctions”.

This artificial intelligence system is used, now more and more often, inside of the American courts with the aim of foreseeing the possibility that a criminal, after his arrest, may be recidivist or not. These risk assessments are used by the court as additional information when a decision must be taken against of the defendant, such as the possibility of being released at liberty during all phases of the trial, on the amount of the security applicable even more fundamental decisions concerning the freedoms of the accused.

A study carried out by the independent American organization ProPublica has shown how such software makes its decisions on a discriminatory basis, and in particular discriminatory against the African-American population (Angwin et al., 2016). In the case brought forward by ProPublica are compared two crimes, one committed by a black person, up until then uncensored, and one committed by a white man, who instead had a past from criminal. When these two people were arrested and subjected to judgment of the machine was assigned to them a rating concerning the possible propensity to commit crimes in the future; the black person is been identified as "high risk potential" to commit new crimes, while the fair-skinned one was identified as "person to low risk".

However, in the following two years, it became clear how the assessment made by the machine turned out to be wrong: the black person in the two years following his arrest, was not charged with any new crime, while the fair complexion one was serving a new eight-year sentence for burglary and theft. Obviously, the study carried out by ProPublica was not based exclusively on this satellite case. The analysis was carried out on a basis of over 10,000 criminal defendants in Broward County, Florida, and have been compared their expected recidivism rates with the actual rate in a period of two years (Angwin et al., 2016). Most of the defendants, when registered in prison, fill out a COMPAS questionnaire. Then, their answers were inserted into the software which generated several scores, including the predictions of the risk of recidivism and the risk of violent recidivism.

The risk categories of recidivism under the COMPAS instrument were then compared with the recurrence rates of defendants in the following two years. The result that emerged was that the score obtained by each defendant predicted correctly the relapse of an offender in 61% of cases, but it was correct in its predictions of violent recidivism only 20% of the time. In the prediction of who would commit a crime again, the algorithm correctly predicted recidivism for white or black accused roughly the same speed but made very different mistakes. Indeed, the analysis demonstrated that: black defendants were often recognized as criminals who presented a higher risk of relapse than they actually were. In fact, black defendants who have not been recidivist for a period of two years had almost double the chances of being classified wrongly at higher risk than their white counterparts. White defendants were often considered less risky than actually were. In fact, the defendants of fair complexion who then committed crimes in the following two years had been mislabeled as low-risk individuals.

The analysis has also shown that even when checking the previous crimes, recidivism, age and sex, black defendants had 45% more likely to receive higher risk scores than the white defendants. Even in the case of violent recidivism, black defendants had double of the probability of being classified at higher risk than the whites defendants. And white repeat offenders were 63% more likely to be wrongly classified as low risk of violent relapse, compared to black violent offenders (Angwin et al., 2016).

One of the issues raised during the court debates was whether or not COMPAS conducted risk assessments based on racial prejudice. Some argued that because the race of individuals was not fed into the software to make its predictions, the algorithm could not be prevented against black defendants. Angwin et al., (2016) proved the opposite, finding that "Black defendants who did not recidivate were incorrectly predicted to reoffend at a rate of 44.9%, nearly twice as high as their white counterparts at 23.5% [...]. In brief, COMPAS scores seemed to favor white defendants over black defendants, underestimating recidivism for whites and overestimating recidivism for black defendants. How is it also technically possible for a software to suffer from prejudice? Or, to launch the discussion in terms of higher-order learning, how can a deep learning algorithm independently reconstruct the notion of race from raw data to predict relapse?

First, it is known that, in the United States, African Americans are more likely to be unfairly convicted (Gross S. R., Possey M., & Stephens K., 2022). For example, African-American prisoners convicted of murder are around 50% more likely to be innocent than other convicted murderers, and innocent blacks are near 12 times more likely to be convicted of drug offences than innocent whites. Since the COMPAS predictive model is based on past beliefs, it does not foresee the risk of recurrence per se, but rather the risk of being convicted of recidivism, and there is consensus in the judicial system that such available historical data is racially biased.

Second, even if the defendant's race is not included in the calibration data, other information may act as a proxy to indirectly inform the algorithm about defendant's race, such as schools attended, geographical indicators, annual income or profession. A sufficiently powerful predictive model, driven by artificial intelligence, would be able to autonomously discover the race construct in the data, and then make biased racial predictions. Therefore, these predictions would be more "accurate" in the sense that they would better fit the historical data of the past. In other words, a powerful AI algorithm will have no problem reproducing with great precision the prejudices and distortions present in its training data.

In the purely procedural field, both civil and criminal, artificial intelligence is designed to support judicial decisions. At the moment many systems to identify possible judicial solutions in legal patterns even very complex are being tested and their rate of accuracy is already 79% (Morelli C., 2018). If at present it is premature to think that these systems can replace judges and magistrates, certainly they are tools intended to have a support function during the investigative, procedural and future legal activity. The consequence could be that, given the shortage of resources of judicial resources, the judges use such systems in inappropriate manner, delegating the final decision to them, without this being the goal for which they were programmed and relying on their "neutrality" which, as analyzed above, may not exist at all. The reference is to the United States of America, where some jurisdictions are already using predictive systems of relapse, also in order to release behind bail. Mathematical and computer models are often used, whose transparency and functioning are not explored but which reveal examples of discrimination and prejudice.

Companies spend huge efforts of time and money to find a perfect candidate for an open position as this can make the difference between success or failure. According to a recent report by LinkedIn, cost-to-hire and time-to-hire matrixes are seeing an increasing trend in mid-to-high level openings. The result is that key positions remain open longer than normal regardless of having to spend more money. This leads companies like Amazon to look for new innovative ways to reduce those expenses and times through the application of artificial intelligence.

Based on some reports (Dastin J., 2018) in 2014, Amazon set up a team to create a tool to examine job applicants' resumes, which used machine learning (ML) and natural language processing (NLP) to find the best candidates who fit the professional profile. Once implemented, this software would utilize sophisticated AI algorithms to learn key traits from successful candidates' resumes, for a period of time, and look for similar indicators in resumes sent for screening. Then, this tool would evaluate the candidate on a 5-star scale in a very similar way to the evaluation system used to rate products on Amazon, depending on how similar they look to the previous successful candidate.

By the end of 2014, the use of this experimental instrument was widespread in the company and few relied heavily on this system as it saved a considerable amount of time. In 2015, the company noted that technical job titles such as software developers and architects are not rated gender-neutrally. This issue led the company to instruct its engineers to investigate the root cause. After much research, engineers arrived to the conclusion that the cause of bias was the data used for training the AI system, mainly consisting of curriculum of male employees in response to the then trend of male dominance in the technology and business industry. This unknowingly distorted training data leads algorithms to create an association that downgrades resumes that include words like "women's" as in "women captain of the chess club." And it has also been reported that engineers have identified instances where the system has downgraded graduates of two all-females' colleges (Dastin J., 2018). These findings led Amazon to rewrite the algorithms so that they were neutral in that context, but it was pointed out, in conclusion, that such an artificial intelligence system theoretically in the future could develop a candidate selection system that might be somehow discriminatory.

According to Reuters, the software used to identify the best talents in the field of technology preferred male resumes, based on the trend of recruitment in the previous 10 years. Gender inequality affects not only humans, but also the artificial intelligences programmed by humans themselves. One of the most relevant examples concerns Amazon and its attempt to create a recruiting engine. Five people familiar with the matter told Reuters that since 2014, the Seattle-based company's team specializing in machine learning has been developing software to automate the search for the best talent. The problem is that the software in question penalized women, especially with regard to jobs for software developers and other technical roles. The team was disbanded in early 2017, as Amazon executives had lost confidence in the project (De Cesco A. F., 2018).

The experiment was born at a crucial moment for the e-commerce giant: machine learning was gaining ground and Amazon's Human Resources department was about to triple its staff. The task of the project team, a dozen people employed in the Amazon engineering center in Edinburgh, was to develop an artificial intelligence

capable of quickly sifting through the resumes of people who had applied for a job offer and identify the best candidates. To each candidate was assigned a score which, as in the case of products sold on the Amazon e-commerce site, ranged from one to five stars. “Everyone desired this Holy Grail”, said one of the sources of Reuters. “They wanted an engine that could scan 100 resumes and find the best five to hire”.
The group had created 500 software programs, focused on specific jobs and locations, and trained them to recognize about 50,000 terms that appeared on the resumes of previous candidates. The catch was right here, referring to the curriculum vitae submitted to the company over a period of ten years: most, in fact, came from men. And so, the system created by machine learning specialists had taught itself that it is better to hire male candidates. While they gave little weight to skills common to most candidates (such as the ability to write different computer codes), the software favored profiles in which appeared verbs most often used in the curriculum of male engineers (for example, executed, “executed”, and captured, “acquired”); conversely, resumes that included expressions such as “women’s” were considered less interesting and attractive, as were those of applicants who had graduated from women’s university.
In 2015, the company realized that the programs presented issues of gender disparity and tried to correct this. But there remained the risk that other discriminatory problems would emerge. However, it also happened that, due to errors in the data underlying the judgments, unqualified candidates were recommended for all types of employment. At one point there was no other option than to close the project. Amazon has pointed out that its recruiters have never referred only to the recommendations generated by the system. The ecommerce leader is currently using “a very watered-down version” of the recruitment engine that is used to perform some elementary tasks, such as eliminating duplicates from the profiles stored in the database. It also appears that a new team will be formed in Edinburgh with the aim of giving a second chance to the automated selection of workers. But this time diversity cannot be overshadowed (Corriere della Sera, 2018).
The project was closed in 2017, when Amazon decided to shut down its experimental artificial intelligence recruiting tool after it was discovered that it discriminated against women. The company created the tool to browse the web and identify potential candidates, evaluating them from one to five stars. But the algorithm has learned to systematically downgrade women’s CVs for technical jobs like software developer.
Altough Amazon is at the forefront of AI technology, the company has been unable to find a way to make its algorithm gender neutral. But the failure of the company reminds us that artificial intelligence develops prejudices from a variety of sources. While there is a common belief that algorithms should be built without any of the biases or biases that color human decision-making, the truth is that an algorithm can unintentionally learn biases from a variety of different inputs. Everything from the data used to train it, to the people using it, and even seemingly unrelated factors can all contribute to AI prejudice (De Cesco A. F., 2018).
Artificial intelligence algorithms are trained to observe patterns in large datasets to help predict outcomes. In Amazon’s case, its algorithm used all the resumes submitted to the company over a ten-year period to learn how to identify the best candidates. Given the low percentage of women working at the company, as in most technology companies, the algorithm quickly identified the male domain and thought it was a success factor.

Because the algorithm used the results of its own predictions to improve its accuracy, it remained stuck in a model of sexism against female candidates. And since the data used for training was at some point created by humans, it means that the algorithm also inherited unwanted human traits, such as bias and discrimination, which have also been a problem in recruiting for years.

Some algorithms are also designed to predict and deliver what users want to see. This is typically seen on social media or in online advertising, where users are shown content or advertisements with which an algorithm thinks they will interact. Similar patterns have also been reported in the recruitment sector. A recruiter reported that while using a professional social network to find candidates, AI learned to give him results that were most similar to the profiles he initially engaged with. As a result, entire groups of potential candidates have been altogether systematically removed from the recruitment process. However, bias also appears for other unrelated reasons. A recent study of how an algorithm ran ads promoting STEM jobs showed that men were more likely to see the ad, not because men were more likely to click on it, but because women are more expensive to advertise. Since companies value ads targeting women at a higher rate (women drive 70% to 80% of all consumer purchases), the algorithm chose to deliver ads more to men than women because it was designed to optimize ad posting while keeping costs down (Science, 2018).

But if an algorithm only reflects patterns in the data provided, what its users like, and the economic behaviors occurring in its market, isn't it unfair to blame it for perpetuating the worst attributes of humans? It is automatically expected that an algorithm will make decisions without any discrimination, when this is rarely the case for humans. Even if an algorithm is distorted, it could be an improvement over the current status quo. To take full advantage of using AI, it is important to investigate what would happen if AI were allowed to make decisions without human intervention. With regards to the first case discussed above, a 2018 study explored this scenario with bail decisions using an algorithm trained on historical criminal data to predict the likelihood of criminals' recidivism. In one projection, the authors were able to diminish crime rates by 25% by reducing discrimination cases in inmates.

However, the benefits highlighted in this research would only occur if the algorithm was actually making every decision. This is unlikely to happen in the real world as judges would probably prefer to choose whether or not to follow the algorithm's recommendations. Even if an algorithm is well designed, it becomes redundant if people choose not to rely on it. Many people already rely on algorithms for many of their daily decisions, from what to watch on Netflix or buy from Amazon. But research shows that people lose faith in algorithms faster than humans when they see them make a mistake, even when the algorithm has overall better performance. For instance, if GPS suggests using an alternative route to avoid traffic that ends up taking longer than expected, chances are to stop relying on GPS in the future. But if taking the alternative route was an individual's decision, it is unlikely to stop trusting their own judgment. A follow-up study on overcoming aversion to the algorithm also showed that people were more likely to use an algorithm and accept its mistakes if they had the opportunity to modify the algorithm themselves, even if it meant doing it work imperfectly.

While humans might quickly lose faith in flawed algorithms, many people tend to trust machines more if they have human characteristics. According to research on self-driving cars, humans were more likely to trust the car and believed it would work better if the vehicle's enhanced system had a name, a specific gender, and a human-sounding voice. However, if machines become very similar to humans, but not entirely, people often find them disturbing, which could affect their confidence in them. Even if the image that algorithms can reflect of society is not necessarily appreciated, it seems that people are still eager to live with them and make them look and behave like humans. And if so, surely even algorithms can be wrong? (Science, 2018).

When it comes to artificial intelligence, there is a tendency to assume that the results produced are more reliable. However, it is forgotten that they are in turn, the product of intelligence and the work of man, so that the prejudices that are part of human nature, necessarily reverberate on the nature and structure of the algorithms that drive AI. In an interesting report published in 2018 by the European Union Agency for Fundamental Rights, it is possible to read that "If the data used for building an algorithm are biased against a group (i.e. systematic differences due to the way the data are collected), the algorithm will replicate the human bias in selecting them and learn to discriminate against this group" (European Union Agency for Fundamental Rights, 2018).

In short, AI, especially if used as a predictive police tool, risks becoming, if not properly regulated and analyzed, from a means of preventing certain types of crime, to a potential tool for perpetuation and exasperation of discrimination. Furthermore, as Fabio Basile points out, predictive policing systems are "systems that to some extent feed themselves with the data produced by their own use, with the risk of triggering vicious circles: if, for example, a predictive software identifies a certain "hot spot", the controls and patrols of the police in that area will intensify, with inevitable growth in the rate of crimes detected by the police in that area, which will therefore become, even more "hot", while other areas, originally not traced in the "hot zones", and therefore not manned by the police, risk remaining, or becoming, for years free zones for the commission of crimes" (Basile F., 2019).

Precisely the above issues and the awareness of the damaging potential of AI, have prompted the European Commission for the Efficiency of Justice, to publish in 2018, the "European Charter of Ethics on the use of artificial intelligence in judicial systems and related areas", which among others establishes the "Principle of non-discrimination". It states that "Particular vigilance must be exercised both in the processing and in the use phase, especially when the treatment is directly or indirectly based on "sensitive" data. It may include racial or ethnic origin, socio-economic conditions, political opinions, religious or philosophical belief, trade union membership, genetic data, biometric data, health data or data relating to sexual life or sexual orientation. Where one of these discriminations is identified, corrective measures should be provided in order to limit or, if possible, neutralize such risks and raise awareness among the actors" (CEPEJ, 2018).

Therefore, all that remains is to hope and hope that the good intentions for a truly fairer, more effective and equitable AI will become reality in the future and not be confined to mere intentions.

CHAPTER 2: LITERATURE REVIEW

2.1 Artificial intelligence and its ethical challenges

“Of all the exciting new challenges that we find ourselves facing today, understanding and defining the new technological revolution is the most fascinating and evocative, since involves a real transformation for humanity. We are on the threshold of a revolution that is drastically changing the way we live, work and relate to each other. If we evaluate its scope and complexity, what I consider the fourth industrial revolution is something with which man has never had to confront before.” This is the beginning of the text by Klaus Schwab, The fourth industrial revolution, which invites all experts of various disciplines to ask themselves what this change means and how to deal with it. It is necessary to have a shared vision of the change taking place: we are faced with a new technology which is blurring the boundaries between physical, digital and organic: artificial intelligence (Schwab K., 2016).

According to Nicholas Carr every technology is an expression of human will in order to extend power and control about the surrounding world: about nature, time, distance (Carr N., 2011). There are technologies that enhance physical strength, dexterity, or resilience; others that extend the range or the acuity of our senses, still others that help us to give a new shape to nature to accommodate our needs and desires. There is then a fourth category that could be defined as “*intellectual technologies*”[3]: “This term refers to all the tools we use to broaden or strengthen our mental faculties [...] that we use to express ourselves, to give shape to our personal and public identity and to cultivate relationships with others” (Parisi F., 2019).

This definition, although less generic than that of Wanda Orlikowski, “set of rules and resources put in place in practice” (Orlikowski W. J., 2000), can still align with the reasoning conducted by Federico Cabitza on “artificial intelligence” and on how we should ethically deal with it: instead of considering AI systems as “living agents, true existences and own that integrate naturally and gracefully into our domestic, urban and working environment” (Sadin E., 2019), as the expression “artificial intelligence” might suggest, it is necessary to harshly go back to the idea that this expression concerns “manifestations and modalities of human doing” (Floridi L., & Cabitza F., 2021).

For centuries historians and philosophers have reconstructed and discussed the role of technology in shaping civilization. On the one hand, determinists, according to which technological progress is seen as an autonomous force beyond human control[4] (McLuhan M., 1986); on the other hand, the instrumentalists, who minimize their power by considering them neutral artifacts subservient to the unconscious desires of their

[3] Nicholas Carr borrows a term used with a slightly different meaning from social anthropologist Jack Goody and sociologist Daniel Bell (Carr N., 2011. *Does the Internet make us stupid*?, cit., p.68)

[4] The deterministic vision in its most extreme conception considers human beings little more than “the sexual organs of the machine world” (Mc Luthan M., 1986), until the point of making men themselves superfluous.

users[5]. “If the experience of modern society teaches us something”, observes the political scientist Langdon Winner, “is that technologies are not only aids for human activity but also powerful forces acting to give new shape and meaning to that activity” (Winner L., 2004). Sometimes our tools do what we we tell them to do. Other times, we adapt to their requests (Carr N., 2011). This is another interesting point of view that emerges clearly in technologies intellectuals and especially in artificial intelligence.

According to Kate Crawford, before questioning the impact of AI in today’s society, two myths must be debunked. The first myth is to believe that non-human systems (any technology intellectual) “are something similar to human mind”. The second myth “is that intelligence is something independent, a sort of natural element, distinct from social, cultural, historical and political forces” (Crawford K., 2021). In light of this, “it is useful to avoid considering machines and instruments as something ontologically stable and fixed, in spite of their materiality; or as something equipped with “agency”, that is, an autonomous capacity for action; this is because doing so could lead us into the error of attributing to machines also an identity and, even, an autonomous will” (Cabitza F., 2022).

These myths are particularly powerful in the field of artificial intelligence, where the belief that human intelligence can be formalized and reproduced by machines has been dominant since the mid-20th century[6]. Already then Joseph Weizenbaum, pioneer of AI and creator of Eliza, the first chatbot or software designed to simulate a conversation with a human, did not agree with this vision considering extremely reductive a concept of intelligence in which human beings were conceived as simple information processing systems and from which descended the “great perverse fantasy” for which AI scientists could have created a machine capable of learning “like a child” (Weizenbaum J., 1972).

This has been one of the central disputes in the history of AI since the 1960s[7]. Hubert Dreyfus, worried of the drift that had taken the discussion, in his work “What Computers Can’t Do”, emphasized a basic difference: while computers require processes and data to be explicit and formalized, human intelligence and experience are based in large measure on numerous unconscious and subconscious processes (Dreyfus H. L., 1988). “Consequently, the less formal aspects of intelligence must be extracted, deleted or approximated for computers, thus making them unable to processing information on situations as human beings do” (Crawford K., 2021). Many things have changed in AI since those years, including the transition from symbolic systems to machine learning techniques. Hence, according to this perspective, “in many respects the ancient controversies about what AI can do are forgotten and skepticism has dissolved” (Crawford K., 2021).

---

[5] “Technlogy is technology”, said media critic James Carey, “a tool for communication and transport across the space, nothing more” (Carey J. W., 2008)

[6] When asked if machines could think, Alan Turing, in a 1950 article, posed the question in other terms: a machine can behave in a way that seems intelligent (imitation game), that is, to seem to an outside observer a human being? According to Cabitza, by placing this question Turing resumes the Second meditation of Descartes to propose the primacy of behavior over being (Cabitza F., 2022)

[7] Milestone were the conferences held in 1961 at MIT (Massachusetts Institute of Technology) with the title *Management and the Computer of the Future* to discuss the rapid progress made by digital computing. At the end of the lectures, John McCarthy, inventor of the term *Artificial Intelligen*ce, went so far as to support the illusory differences between human and machine work.

In recent years, artificial intelligence has become is rapidly expanded both as an object of academic study and in industry. Today, a small number of powerful technology companies are deploying, on a planetary scale, AI systems that once again are hailed as equals or even superior to human intelligence[8]. But what is meant for artificial intelligence? Already since the sixties there referred to that field of investigation of intersection of the new disciplines committed to reducing the complexity of the physical world to retroactive information mechanisms or symbolic and computable models: cybernetics and informatics. Since then AI has been defined in many ways and there is no unitary definition on which all agree[9]. Indeed, the extant literature has generated many definitions of AI throughout the years, without actually coming up with something definitive.

After sixty-five years the High-Level Expert Group on Artificial Intelligence (AI Hleg) appointed by European Commission to draft an ethical guidelines document for reliable AI, defines Artificial Intelligence as "systems that exhibit intelligent behavior by analyzing the environment and performing actions, with a certain degree of autonomy, to achieve specific objectives" (AI Hleg, 2019). This definition will then be extended at the end of the document as follows: "Artificial intelligence (AI) systems are software systems (and possibly, hardware) designed by man that, given a complex goal, act in the physical or digital dimension perceiving their environment through data acquisition, interpreting the structured or unstructured data collected, reasoning on knowledge, or processing derived information from this data and deciding the best actions to be taken to achieve the given objective. AI systems can use rules symbolic or learning a numerical model and can also adapt their behavior by analyzing how the environment is influenced by their previous actions" (AI Hleg, 2019).

Adapting the definition presented in this document, it was found that, despite numerous studies using the term intelligence, precisely because of its vagueness, is less adequate than the term "rationality"[10], meaning ability to choose the best action to take to achieve a certain objective in the light of certain criteria to be optimized and the resources available. Considering that AI, as explained before, is a form of automation, something traceable to simple operations of reading and writing of symbols, "we clear the field from various misunderstandings and spectra, such as those of identity, the conscience and autonomy" (Cabitza F., 2022).

This assumption is supported also by the words of one of the greatest robotics experts, Rodney Brooks: "but it's only computation". In fact, the rationality of an AI system is expressed through the following process: "perceiving the environment through sensors in which it is immersed, and therefore collecting and interpreting the data, reasoning on what is perceived or processing the information deduced from the data, deciding which is the best action and acting accordingly through its actuators, and eventually changing the environment" (AI Hleg, 2019).

---

[8] Consider that in 2021 Us venture capitalists moved in this sector 52.9 billion dollars, three times larger than the Chinese (17 billion dollars) and ten times larger than the British (4.6 billion dollars). (Source: www.agendadigitale.eu)

[9] An outdated report (Legg S., & Hunter M., 2007) already listed 18 definitions of AI. The number is growing (Russel S., & Norvig P., 2021)

[10] Massimo Chiriatti prefers to speak instead of "unconsciousness", because algorithms, by executing rules that learn autonomously from data, produce results without any understanding and awareness of what they are doing (Chiriatti M., 2021).

The idea behind this document exemplifies an AI system. Each AI system interacts with the environment as it requires data[11] collected through sensors, which the system acquires as input. This data must be transformed in information that the reasoning module of information can understand in order to then elaborate a process decision-making. This is the cornerstone of an AI system which will then be translated into action through actuators. “A rational learning system is a rational system that, after having performed an action, evaluate the new condition of the environment (through perception) to measure the success of the action and then it adapts its rules of reasoning and its decision-making methods” (AI Hleg, 2019). Hence, the literature has confirmed that perception, reasoning, decision-making and implementation are, schematically, the three stages of the process on which are based most of the techniques used to build AI systems.

Synthesizing as much as possible, all these techniques can be organized in two main groups relating to reasoning and learning skills. The group of techniques pertaining to reasoning and decision-making includes the representation of knowledge and reasoning knowledge-based, planning, programming, research and optimization. After having assessed how these techniques allow to carry out the reasoning on the data coming from the sensors, I think it’s important to underline that they are part of this process the formulation of inferences through symbolic rules, the planning and scheduling of activities, the research among a wide range of solutions and optimization between all possible alternatives to a problem. This is crucial in the present research since the group of techniques related to learning includes machine learning (Machine Learning / ML), neural networks (Artificial neural networks / ANN), deep learning (Deep Learning / DL), and many other learning techniques. Thanks to them, an AI system can learn to solve problems that cannot be precisely defined, or the whose solution method cannot be described with rules of symbolic reasoning. Examples of such problems are those related to perception skills, such as understanding the speech and language, artificial vision or prediction of behavior[12].

Anyone who is not astonished at such a digital revolution has not grasped its scope “We are talking about a new chapter in human history” (Floridi L., 2022). We are on the threshold of a new era and the price of occupying such a special place, it is paid with uncertainty on the one hand, but also with responsibility to lay solid foundations for the future (De Luca S., 2020). For this purpose, it can be taken inspired by what Cardinal Carlo Maria Martini said in one of the dialogues he promoted in “The chairs of non-believers”: “Are all virtual technologies so threatening? The whole journey towards artificial intelligence will eventually devalue the value of the person, reducing it to pure mechanics? Or, instead, it will be the human values to induce science

---

[11] This data can be structured or unstructured depending on whether it is organized following a known or unknown scheme.

[12] Another relevant discipline is *robotics* which can be defined as “AI in action in the physical world”, also called *embedded AI*. The robot is a physical machine designed to deal with the dynamics, uncertainties and complexity of the physical world. In the control architecture of the robotic system are generally integrated capabilities of perception, reasoning, action, learning as well as interaction with other systems.

to open up new fronts thanks to technological achievements? [...] All this constitutes a very encouraging scenario, provided the intelligence human remain master of the processes"[13] (Martini C. M., 2015).

The ethical challenges posed by AI systems engage the current discussion of researchers from various disciplines. Among these, Luciano Floridi, in his latest book, convincingly argues that AI separates effective resolution of problems and correct execution (agere) from intelligent behavior (intelligere) (Floridi L., 2022). Through this split, AI can incessantly "colonize the boundless space of problems and tasks [...] it is precisely when we stop trying to produce human intelligence that we can successfully replace it in an increasing number of tasks" (Floridi L., 2022).

This divorce is ethically problematic because artificial agents "are sufficiently informed, "smart", autonomous and able to perform morally relevant actions independently of the human beings who created them[14]" (Floridi L., Sanders J. W. 2004). Such division has modified one of the equations on which moral evaluation has always been based: acting has always been associated to intentionality, since it means passing from the inner sphere knowledge and decision to the outer sphere of entities. Acting in this sense is an "original human phenomenon".

Romano Guardini explains this concept by comparing the actions of man with the activities carried out by animals. Hence, according to this perspective, man acts on the basis of knowledge and decision. With his action he pushes himself into reality, he grasps it, disposes of it, gives it shape, but his core remains internal, because it starts from the intention. Through action, intention becomes history: the inner space exceeds or precedes history, although by acting, the intention submits itself to the court of history. There is no morality that can remain confined to the intention, it must enter into reality and collide with resistances that reality poses. Guardini theorized that, differently in the animal, actions are totally connected to the corresponding animal life and are prolonged in the environment surrounding. This environment is harmonized with the life of the animal: "the animal is properly the concrete living being plus its own environment, it is the animal in its environment [...] the actin of the animal belongs to this context and precisely the perfection of this act is a sign that it has the character of function[15]" (Guardini R., 2021).

This concept is intrinsically linked to the fact that a very similar thing happens in AI systems: their actions are connected to the environment. In the forties and fifties the computer was a room and the interaction was physical; in the seventies it was released to sit in front of it and the relationship became semantic; today it was entered back into it in the form of an "infosphere"[16], which surrounds people and is increasingly adapted to

[13] For Paolo Benanti, a governance of technologies related to AI is a compelling appeal to consciences, "a space where anthropological and ethical considerations, in a mutual exchange and dialogue, must become effective forces to shape and drive technological innovation, making it authentic source of human development" (Benanti P., 2018)

[14] In the field of digital healthcare or financial markets, all relevant data is machine readable so that decisions and actions can be made automatically by applications and actuators.

[15] (Guardini R., 2021) For example, the spider's web is exactly as it should be according to the weight and type of movement of the spider, the type of insects to be captured, the actual possibilities of fixation, etc.

[16] In this infosphere the assumption is that an agent may be artificial. That is why people are regularly asked to prove that they are not robots by clicking on CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*).

AI. “We are”, to use the expression of Floridi, “increasingly onlife”. It is the environment that is designed for being compatible with robots, is an infosphere in which we feel more like analogue guests than digital ones. Therefore, “it is not surprising that our artificial agents perform better and better. It is their environment” (Floridi L., 2022).

According to some, the ethical risk of this winding is that beings human beings can become part of the mechanism without having awareness and that, drawing a digital world to measure of artificial acting and not of human intelligence, technologies could shape the physical and conceptual environments, pushing to adapt to them, which is the easiest way to make things work. Returning to Martini’s invitation that Floridi welcomes, “intelligent human design should play a greater role in shaping the future of our interactions with current and future “smart” artifacts and environments we share with them. After all, it is a sign of intelligence working stupidity for us” (Floridi L., 2022). The literature has shown that the ethical debate already took shape since the sixties[17], even if only in the last few years is increased the need for reflections on the impact of AI in the company. Already the science fiction writer Isaac Asimov in a short story of 1942[18], had elaborated his now infamous “Three Laws of Robotics”, and nowadays Frank Pasquale proposes new laws formulated for those who design, control and use IA systems[19].

What we know from the current literature on ethics is that in recent years many initiatives have been promoted by various organizations to establish ethical principles in the interest of socially beneficial AI[20]. The principles that emerged show their consistency “which is impressive and reassuring”[21]. Even The Pontifical Academy for Life under the pressure of the president, Mgr. Vincenzo Paglia, has shown a specific interest in new technologies, dedicating the two-year period 2019-2020 to robo-ethics and related ethical-anthropological issues connected to the so-called artificial intelligence.

---

[17] Think of the debate between Wiener and Samuel. Wiener argued that “machines can transcend some of the limitations of their designers, and that in doing so can be effective and dangerous” (Wiener N., 1960); Samuel instead “The machine is not a threat to humanity, as some people think. The machine does not possess a will, and its so-called “conclusions” are only the logical consequences of its input, as revealed by the mechanistic operation of an inanimate assembly of mechanical and electrical parts” (Samuel A., 1960)

[18] In the story *Runaround* the author formulates his three laws: 1. a robot cannot harm a human being, nor can allow that, due to its lack of intervention, a human being receives damage; 2. A robot must obey the orders given by humans, provided that such orders do not contravene the First Law; 3. a robot must protect its existence, provided that this self-defense does not conflict with the First or Second Law.

[19] (Pasquale F., 2021) Robotic systems and AI must: 1. complement and not replace professionals; 2. not counterfeit humanity; 3. not intensify a zero-sum arms race; 4. Always indicate the identity of their creators, controllers and owners.

[20] Among the major initiatives: the *Asilomar Principles for AI* (Future of Life Institute, 2017); the *Montréal Declaration for Responsible AI* (University of Montréal, 2017); the general principles offered in the second version of *Ethically Aligned Design. A vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems* (IEEE, 2017); the *Principles of partnership on AI* (Partnership on AI, 2018).

[21] Floridi L. (2022) in doing an analysis of all the principles that emerged bring them closer to the 4 fundamental principles used in bioethics: charity (every technology is created for the benefit of humanity); not maleficence (warn against the negative consequences of the use of AI); autonomy (to prevent the growth of artificial autonomy can undermine human autonomy); justice (promote prosperity, preserve solidarity and avoid inequity)

The outcome of this journey has found a first moment in the Rome Call for AI Ethics[22] as a shared document of commitments, in which, in a brief and concise form, are offered some lines for an ethics of Artificial Intelligence and for promote, according to the term coined by Paolo Benanti, an "algor-ethics" (Benanti P., 2018), or the development and use of AI according to fundamental principles of a good innovation. "If machines are able to substitute man in so many decisions, we must ask ourselves by what criteria this subrogation can take place. In other words, if the machine makes a mistake who is responsible?

This also suggests the importance of ethics, which becomes the guard rail that allows to live more safely with these clever machines. However, ethics is a question of values that are difficult to communicate to machines, that work on the basis of numbers. And then there is a need to put together algorithms and ethics. From here it comes a new term, "algorithmics", the new discipline that would like to make machines capable of computing typically human principles. A path which involves more disciplines: philosophy, technology, information technology are no longer enough, contamination is needed" (Coen E., 2020).

The lines drawn so far by the literature are: "transparency", so that AI systems can be understood; "inclusion", so that everyone can benefit and to all individuals can be offered the best possible conditions for expression and development; "responsibility", of those who design and implement AI solutions; "impartiality", so that equity and human dignity are safeguarded; "reliability", because AI systems must be able to function reliably; "security and privacy", so that the privacy of users is respected.

The starting point for Benanti is that we must not fall into the logic of datism, or as he defines it dataism, the new religion that ensnared by a logic of doing, considers the entire universe as a data stream. It is the "new universal narration that legitimizes a new principle of legitimacy: the algorithms and Big Data" (Benanti P., 2018). Therefore, the role of ethical reflections should not focus on finding technical solutions, but on make present the critical question about the sense of human that is what must guide the decisions of our life. "With a computer we can turn almost any human problem into statistics, graphs, equations. However, the really disturbing thing is that in doing so we create the illusion that these problems can be solved with computers" (Yehia N., 2005). Lehonard Gerd, in order to highlight the specificity of human being and his qualities, speaks of androrhythms that must be absolutely preserved even if, in comparison with systems, computers and non-biological robots, communicate a sense of clumsiness, complexity, slowness, danger or inefficiency[23] (Gerd L., 2019). There is the risk that this "fallibility", the hallmark of humanity, may be regarded as harmful to fundamental values (Benanti P., 2018).

While human slowness is not a bug or a disadvantage compared to machines; it is rather what differentiates people, the time of consciousness, the possibility of asking questions, of asking ourselves what we "feel" (Chiriatti M., 2021). Massimo Chiriatti began his text in a provocative way with a piece generated by an AI: Man sees Artificial Intelligence as a machine capable of making his own decisions, but he is wrong, because

[22] Signed in Rome on 28 February 2020 by Microsoft president Brad Smith, IBM vice president John Kelly III, FAO general director Qu Dongyu and minister for Technological Innovation Paola Pisano, with the participation of the then president of the European Parliament, David Sassoli.

[23] As opposed to the acronym STEM, the author coin another: CORE that stands for creativity/compassion, reciprocity and empathy.

it is only a calculator of symbols, even if ever more sophisticated; Artificial Intelligence sees Man as a set of numbers, but it is wrong, because consciousness is incomputable (Chiriatti M., 2021).
New and unprecedented challenges are being faced: new problems require new solutions. This is the invitation of Pope Francis. “A better world is possible thanks to technological progress if this is accompanied by an ethics based on a vision of the common good, an ethic of freedom, responsibility and fraternity, capable of encouraging the full development of people in relationships with others and with creation”[24].

2.2 Exploring the intersection of AI and consumer patterns

According to some, new technologies often alter customer behavior (e.g., Giebelhausen M., et al., 2014; Groom V., et al., 2011; Hoffman D. L., & Novak T. P., 2018; Moon Y., 2003). Thus, consistent with this theory, all the authors expect AI to do so as well. In light of this, it is useful to analyze three different topics, related to the adoption of AI, the use of AI, and post-adoption issues.
With regards to AI adoption, in general, due to a large variety of factors, consumers see AI negatively, which is an obstacle to adoption. This is crucial in the present research since, as already pointed out, these negative opinions often result from customers’ sense that AI is unable to feel (Castelo N., et al., 2018; Gray K., 2017) or that AI is relatively less able to identify what is unique in every customer (Longoni C., et al., 2019). Even Luo X., et al., (2019), in particular, has supported the idea that customers perceive AI bots as less empathetic. The literature has shown that customers are also less likely to adopt AI in subsequent domains (Castelo N., et al., 2018; Castelo N., & Ward A., 2016) and for tasks that are relevant to their identity (Castelo N., 2019; Leung E., et al., 2018).
Therefore, a relevant area for investigating, important both from a practical point of view, would be to consider how best to mitigate the impact of the above. Several researchers and practitioners point out that the placement of AI as a learning artificial organism, or the placement of AI application as one that combines AI and human inputs, can help mitigate the impact of the above concerns to some extent.
Longoni C., et al., (2019) too, has posited that offering consumers the possibility to slightly modify AI can make these customers look beyond the neglect of uniqueness and concentrate more on the benefits of personalization. Hence, this could be a way of mitigating the points raised earlier.
Moreover, according to some researchers, the discomfort with AI is accentuated if the application of AI is embedded in a robot: indeed, it was found that, as robots become more human, because of UVH, customers find these robots unnerving. For this reason, these factors can hinder the adoption of AI and are worth studying. As a result, I think it’s important to mention also the fact that an interesting moderator of this effect might be if the AI module is perceived by consumers as a servant or a partner; it has been shown that the UVH effects may be stronger if the AI reaches partner status.

---

[24] Speech of Pope Francis to the participants in the seminar *The Common good in the digital age*, promoted by the Pontifical Council for Culture and by the Dicastery for the Service of Integral Human Development, 27 September 2019.

In this regard, Castelo N., (2019) has underlined that the first efforts of mitigating these effects consist in trying to foster empathy, convincing customers that robots have a certain ability to see things from the point of view of customers and also have a certain ability to feel sympathy for the consumer if the consumer was suffering. However, other possible approaches may include anthropomorphizing AI, as this could convince customers that AI has some empathy.

Sociologists seem particularly interested in how robots with embedded AI could make inroads into society (Boyd R., & Holton R. J., 2018), observing that "complexities arise when cultural preferences associated with the human being over the delivery of personal service machine are considered. Do... consumers find social robots acceptable?". In general, research can address how attitudes towards robots differs according to culture (Li D., et al., 2010).

Beyond the concerns associated with culture, it may be relevant to explore what other character factors determine whether customers are willing to have robot-drawn hair or accept robot-provided care services (Pedersen I., et al., 2018). In addition to physical well-being considerations, some researchers suggest that robots can help with spiritual well-being (Fleming P., 2019), as exemplified by the robot priest BlessU-2 (Sherwood H., 2017) and the Buddhist monk Xian'er (Andrews T., 2016). Understanding how robots with integrated AI can help in several ways, as well as improving the physical well-being of customers, is a good area for further studies.

Central in the researches about AI usage, is also the fact that, when consumers interface with an AI application, this could trigger a low-level mindset (Kim T., & Duhachek A., 2018). The goal of the research should be then to assess which other mindsets could be triggered by AI, for example, AI can prioritize prevention between consumers for whom AI is a substantially new technology. It is important to analyze the implications that the relative insights would have on how the AI application should communicate with the consumer, due to the fact that communication has a greater impact when it fits the triggered mindset.

When AI is incorporated into robots, the robots probably have important roles in customers' lives, assuming the role of front-line service providers (Wirtz J., et al., 2018), nannies, pet replacements or companions. In addition to the previously documented UVH challenges, Mende M. et al. (2019) in their research have found that interactions with AI-embedded robots trigger discomfort and compensatory behaviors. After having assessed when customers perceive AI robots negatively, I think it's important to determine whether these perceptions can improve over time.

Lastly, if we think about the preferences of ideal consumers, these really vary from their past behaviors, for example, customers trying to stop eating unhealthy foods. For this reason, AI may make it harder for them to find and move towards their favorite options, only by presenting them with choices that reflect their past behaviors. An example of this is the large use of "retargeting" of digital ads. Thus, how to train AI to better manage this problem?

Hence, the literature has confirmed that the subsequent consequences of AI adoption even propose some significant research topics. In particular, consumers could feel a loss of autonomy if AI could be subject to their preferences. In theory, since AI facilitates data-driven and micro-targeting marketing offerings (Gans J., et al., 2017; Luo X., et al. 2019), customers should see bids more favorably, because it reduces search costs. But it could also undermine the perceived autonomy of customers, with implications for their assessments and choices (André Q., et al., 2018). Another interesting point of view is that, if customers learn that an AI algorithm can predict their preferred choices, they can deliberately choose an unprivileged option, in order to reaffirm their autonomy (André Q., et al., 2018; Schrift R. Y., et al., 2017).

In the attempt of applying all these considerations to the field of AI post-adoption, a variety of research questions have come up. For example, what factors determine if and how much customers value perceived autonomy in AI-mediated choice settings? For that purpose, it might be useful to further analyze the reasons and the individual differences, such as culture, and whether consumers consider AI as an employee or partner. Furthermore, research could also examine state factors, such as product type; indeed, perceived autonomy may be less relevant to utilitarian product choices than hedonic ones, due to differential links with customer identity.

In addition, some researchers have found that there is a generalized fear of a loss of human connection, if humans form bonds with robots with embedded AI. The popular press (Marr B., 2019) feeds concerns about robots with embedded AI that become popular over humans as partners. Indeed, robots like Harmony look promising in this sense, able to take on different personalities and express certain expressions. However, other authors have tried to explain that these robots could be harmful to society in general, increasing social isolation, reducing the incidence of marriage, or reducing birth rates, which is critical for countries like Japan, where birth rates are already low. Hence, I believe that, this point suggests some interesting research opportunities.

2.3 Addressing algorithmic bias in AI applications

To sum up, the literature has shown that, artificial intelligence is of interest to policymakers. It is possible to see three big areas where policymakers try to ensure that companies strike a fair balance between their business interests and customers' one: data privacy, bias and ethics[25].

As Wilson J. (2018) explained in his article, today the combination of AI and big data means that companies know a lot about their customers. Therefore, two issues deserve the attention of research. First, customers are concerned about the privacy of their data (Martin K. D., & Murphy P. E., 2017; Martin K. D., et al., 2017). According to Tucker C. (2018), privacy is problematic for several reasons: (1) the low expenditure of storage suggests that data can exist significantly longer than expected, (2) data might be reformulated and reused for logics other than those provided and (3) data relating to a particular person can contain information on other people. However, data privacy policy requires balancing two competing priorities: insufficient protection means that customers cannot adopt AI-related applications; too much regulation can stifle innovation.

[25] Firms are aware of this, and are taking steps to adequately answer (Deloitte Insights, as reported in Schatsky D., et al., 2019)

Second, important research issues are whether data privacy management efforts should be guided by legal regulation or self-regulation, since "it is not yet clear whether market incentives will be sufficient for companies to adopt consumer-friendly policies or whether regulatory supervision is necessary to ensure a fair outcome for consumers" (Verhoef P. C., et al., 2017). It is relevant to consider that cultural perspectives on data privacy even differs; some researchers have suggested that the lack of data privacy in China, for example, is consistent with Confucian cultural ideals (Smith C., 2019).

Third, there is a need for information on how best to recognize and address privacy issues when collecting data, as well as how to manage data privacy failures, for example data breaches. If we think about Amazon, it already sells doorbells with cameras, namely the Ring device, and may have plans to add AI facial identification to devices (Fowler G., 2019). Indeed, customers can worry if Amazon has access to the data recorded through Ring, which it could use or sell, and neighbors may also complain if Ring cameras record their activities in the garden without their permission. In addition, Ring's data may be cited by law enforcement or obtained illegally by hackers. Therefore, these issues reflect arguments for further research.

Finally, an author has underlined the importance of the privacy-customization paradox (Aguirre E., et al., 2015). According to him, customers need to balance privacy concerns with the benefits of personalized recommendations and offers. Important questions concern how customers discourage the optimal trade-off, including which individual difference variables and state variables could moderate their choices. Does the compromise depend, for example, on the product category or level of customer confidence in the company? And besides, what would this trade-off turn be like over time?

Central in the researches about bias is also the idea that the potential algorithmic bias embedded in AI applications could stem from multiple causes (Villasenor J., 2019), including data sets that inform AI. For example, Amazon abandoned a tool that used artificial intelligence to evaluate work applications, in part because it discriminated against female applications (Weissman J., 2018). This distortion emerged because the training data sets used to develop the algorithm were based on data from previous candidates, who were predominantly men. These and many similar kinds of research indicate the negative impact of biases in AI systems and the need for ensuring fairness in such systems (Agarwal A., et al., 2022). As Manheim K., & Kaplan L. (2019) explained in their article, AI technologies are often able to identify anonymized data and may produce algorithmic bias. Exacerbating the problem, many AI algorithms are opaque black boxes, so it is difficult to isolate which exact factors these algorithms consider (Davenport T., et al., 2020). This also suggests the importance of checking if there are biases in AI applications, but I will cover this topic more in depth in the next section.

Furthermore, the literature has underlined that AI may not be able to distinguish attributes that could induce potential bias. Villasenor J. (2019) affirmed that, by and large, it cannot be offensive when insurance companies treat men and women in a different way, with a set of premiums for male drivers and another one for female drivers. Hence, according to this perspective, the question that arises is whether it would be appropriate for AI to determine car insurance premiums on the basis of religion. In this regard, although many may not support the concept of basing car insurance premiums on religion, from the perspective of an AI algorithm designed to "slice-and-tell" data in any way it can, the distinction between the use of gender against religion, as a basis for determining car insurance rates, may not be obvious. In light of this, Knight W. (2017) too, has posited that the complexities and issues related to bias remain a not simple, not trivial and unresolved issue.

Lastly, AI developers must confront ethics; before going in depth in such topic, the literature has highlighted two issues. Firstly, data privacy choices may reflect a company's strategy (for example, if it wants to be perceived as a trusted company; Martin K. D., & Murphy P. E., 2017; Goldfarb A., & Tucker C., 2013) but could also be driven by ethical concerns. In this direction, the goal is then to assess "how can ethical regulatory theory pave the way for what organizations should do to exceed consumers' privacy expectations, as well as to over-comply with legal mandates in order to preserve their capacity for self-regulation" (Martin K. D., & Murphy P. E., 2017). Moreover, a related research topic could involve examining how ethical concerns about AI vary across cultures.

Second, companies choose to implement AI by defining what problems it will face. For example, two Stanford researchers used deep neural networks to identify people's sexual orientation, simply by analyzing facial images (Wang Y., & Kosinski M., 2018). The present study, therefore, found that deep neural network tools (vs. human judges) were better able to distinguish between gay and straight men. However, the work has raised ethical concerns, as many argued that this AI-based technology can be used by spouses on their partners (if they suspected that their partners were closeted), or more scarily may be exploited by some governments to "out" and then prosecute certain populations (Levin S., 2017). Therefore, an important topic for research is to address in advance the types of applications for which AI should be used for (or, should not be used for).

2.4 The future of marketing: harnessing the power of AI

In the future, it seems that both customer behaviors and marketing strategies will undergo a transformation because of artificial intelligence. Indeed, based not only on existing research but also on substantial interactions with practice, it is presented a multidimensional framework in order to understand the impact of AI. As already pointed out, it is very likely that artificial intelligence will influence marketing strategies, in particular sales processes, business models and customer service options, in addition to customer behaviors (Davenport T., et al., 2020). These upcoming transformations could be better understood by explaining some cases from different industries.

What we know from the current literature on the transportation sector is that AI-enabled and driverless cars may modify the business models on the one hand, and customer behavior on the other. Both taxi and ride sharing businesses need to face a change to avoid that AI-enabled transportation models will try to push them aside; the demand for car insurance and breathalyzers will likely decrease, while the demand for security systems that protect cars from hackers will increase (Hayes A., 2015). In addition, driverless vehicles could have an impact on the attractiveness of real estate, because driverless cars can move at higher speeds, and therefore commute times will be reduced. Moreover, commute times will be more productive for passengers, who can work safely while being guided to their destination; as such, remote suburbs can become more attractive than today's homes.

Secondly, sales processes in different sectors will be influenced by AI. Many salespeople still count on a phone call as a key part of the sales process. In the future, it seems likely that shop assistants will be supported by an AI agent who tracks conversations in real time. For instance, through the use of advanced voice analytics capabilities, an AI agent may be able to deduce from a customer's tone that an unspoken question continues to be an issue and give real-time comments to lead the salesperson's following approach. In this sense, AI might enhance salespeople's competences, but it could also trigger accidental negative consequences, particularly if customers feel embarrassed by the conversations being tracked through AI. Furthermore, in the future, companies may mainly use AI bots (Miller G., 2016) which, in certain cases, function as good as human sellers, to establish the first contact with potential customers. Though, there can still be the risk that consumers could find out they are interfacing with a bot, and this can lead to the fact that they might feel an unpleasant sensation, causing unforeseen detrimental consequences.

Third, online retailers currently use a business model that normally requires customers to dispose orders, after which the online retailer ships the products (Agrawal A., et al., 2018; Gans J. S., et al., 2017). Thanks to AI, online retailers might be able to anticipate customers' needs and desires; supposing that these previsions reach great precision, retailers could use AI to identify customer preferences and ship items without a formal order, with the option for customers to return what they don't need (Agrawal A., et al., 2018; Gans J. S., et al., 2017). This change would modify business models, sellers' marketing strategies, and consumer behaviors (for example, information search). Companies such as Stitch Fix, Birchbox and Trendy Butler are already using AI and these technologies in order to try to predict customers' desires, with variable degrees of success.

In light of these three use cases just illustrated above, for many researchers and academics it has become critical to forecast that AI will transform the face of consumer behaviors and marketing strategies. In fact, the literature has shown that AI will be the most adopted technology by marketers in the coming years (Columbus L., 2019). The factors needed for AI to deliver on its promises may already be there; it has been argued that "this very moment is the great inflection point of history" (Reese B., 2018). However, there may be a dispute about this subject. First of all, the technological potential required to run the above examples persists to be inappropriate. As an example, self-driving cars are not ready for deployment (Lowy J., 2016), since, among other things, currently self-driving cars cannot cope with severe weather conditions. Predictive analytics also needs to improve significantly before sellers can accept shipping and purchasing practices that avoid considerable product returns and the connected negative effect.

In the attempt of putting all this together, it emerges that marketers and researchers need insights into not only the final commitment of AI, but also the direction and timeline forward which AI is likely to develop. It is important to address the above issues, drawing not only on a literature review in the fields of marketing (and more broadly, business), sociology, psychology, computer science and robotics, but also extensive interactions with practitioners.

Second, the above examples underline the positive consequences of AI, without going in depth in the widespread and justifiable concerns associated with their use. One the one hand, according to Facebook CEO Mark Zuckerberg, “AI is going to make our lives better in the future.”; on the other hand, technologists such as Elon Musk believe AI is “dangerous” (Metz C., 2018). AI may not deliver on all of its promises, due to the challenges it introduces related to data privacy, algorithmic biases, and ethics (Larson K., 2019).

It is argued that the marketing discipline should take the lead in addressing these questions, because it has the most to gain from AI. In an analysis of more than 400 AI use cases, across 19 industries and 9 business functions, McKinsey & Co. shows that the greatest potential value of AI is in marketing and sales domains (Chui M., et al., 2018), through impacts on marketing activities such as the next best deals to customers (Davenport T. H., et al., 2011), programmatic buying of digital ads (Parekh J., 2018), and predictive lead scoring (Harding K., 2017).

Besides, the impact of AI differs depending on the industry; several researchers have described how the influence of artificial intelligence on marketing is greatest in sectors such as banking, retail, travel and consumer packaged goods. These businesses fundamentally produce great amounts of customer transaction data and customer attribute data and require frequent contact with large numbers of customers. Additionally, information from external sources, such as social media or data broker reports, can raise this data. Subsequently, AI can be leveraged to analyze this data and provide personalized recommendations (related to the next product to buy, the optimal price, etc.) in real time (Mehta N., et al., 2018).

However, the marketing literature linked to AI is relatively scarce, even if some researchers have made an effort to indicate a framework that explains where AI stands today and how it is possible to develop. Marketers expect to use AI in areas such as segmentation and analytics (linked to marketing strategy) and messaging, personalization and predictive behaviors (related to customer behaviors) (Columbus L., 2019). Therefore, it is useful to outline how AI can influence marketing strategies and customer behaviors in the future. In doing so, AI should be studied not only by those involved in computer science, but also by those who can integrate and incorporate insights from psychology, economics, and other social sciences (Rahwan I., et al., 2019; Burrows L., 2019).

According to some researchers, artificial intelligence “refers to programs, algorithms, systems and machines that demonstrate intelligence” (Shankar V., 2018), is “manifested by machines that exhibit aspects of human intelligence” (Huang M. H., & Rust R. T., 2018), and involves machines mimicking “intelligent human behavior” (Syam N., & Sharma A., 2018). It is based on various important technologies, such as machine learning, rule-based expert systems, natural language processing, deep learning, neural networks, physical robots, and robotic process automation (Davenport T. H., 2018). Through the use of these tools, AI produces a means to “interpret external data correctly, learn from that data, and exhibit flexible adaptation” (Kaplan A., & Haenlein M., 2019). Another framework that has been used by the literature in describing AI is not its underlying technology, but rather its marketing and business applications, such as automating business processes, gaining insights from data, or engaging customers and employees (Davenport T. H., & Ronanki R., 2018).

First, to automate business processes, AI algorithms execute clear assignments with little or no human assistance, including call centers to register keeping systems or transferring data from email, substituting lost ATM cards, performing simple market transactions or "reading" documents to extract central provisions using natural language processing. Moreover, AI might obtain relevant insights from huge amount of consumer and transaction data, containing quantitative data as well as text, facial expressions, pictures and voice. Indeed, through the employment of AI-enabled analytics, businesses are allowed to anticipate what a consumer intends to buy, predict credit fraud before it happens, or serve targeted digital advertising in real time. For example, stylists who work at Stitch Fix, a clothing and styling service, makes use of AI to discover which clothing styles work best for different clients. In fact, this type of AI may incorporate significant information and data produced by the following evidence: consumers' articulated options, handwritten comments, their Pinterest accounts, general fashion style and similar consumer preferences.

Finally, other authors have tried to explain that thanks to AI is possible to engage consumers before as well as after the sale. In this regard, it is showed how AI bots offer benefits beyond just 24/7 availability. After having assessed how these AI bots have lower error rates, I think it's important to underline that they also free up human agents to tackle more complex cases. Additionally, it is possible to state that the implementation of AI bot can augment or be restricted based on whether the demand increases or decreases.

Hence, the literature has confirmed that through AI there is the possibility to both diminish costs and increase earnings. Thus, better marketing decision, including pricing, promotions, product recommendations, increased customer engagement, may lead to increased revenue. On the contrary, the automation of designed market transactions, easy marketing tasks and client services can cause a decrease in costs. In addition, it is possible to see that instead of replacing humans, companies generally use AI to increase the capabilities of their human employees (Gaudin S., 2016).

A similar view is shared by Ginni Rometty, the CEO of IBM, who affirmed that AI would not lead to a world of humans "versus" machines, but rather a world of humans "plus" machines (Carpenter J., 2015).

2.5 The impact of AI in the recruitment process

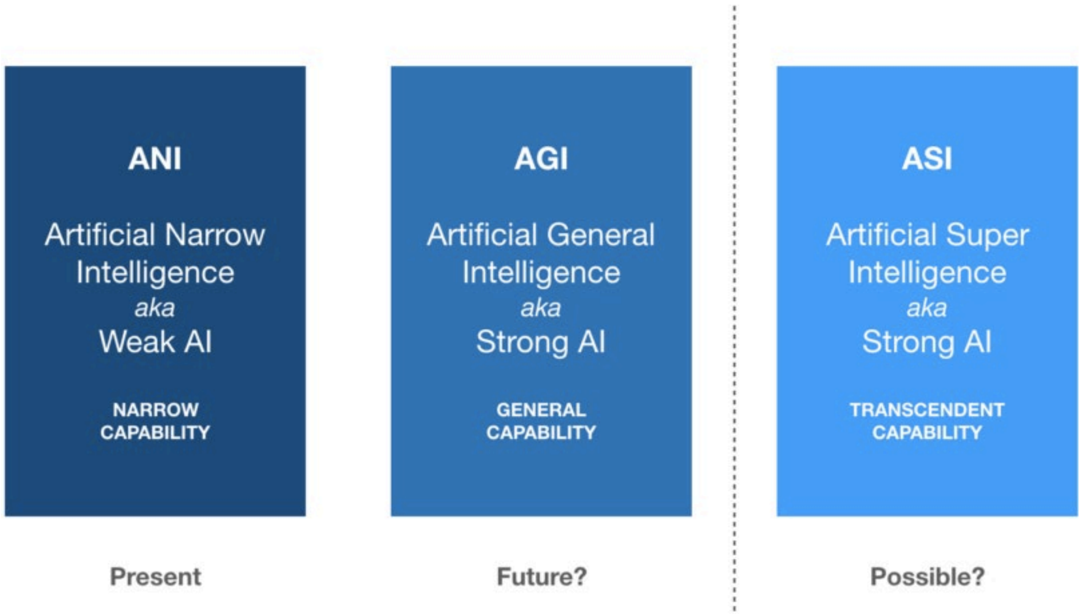
The extant literature has enormously underlined the importance of technological development as a means to automate various types of routine activities, devaluating in an important way the traditional and non-specialized work of repetitive tasks carried out within companies. Moreover, it's relevant to mention that, with the progress of technology in a form increasingly linked to artificial intelligence, jobs and professions

classified in the range media, therefore not managerial, face more and more this impact and suffer a *deskilling*[26] (Bhardwaj S., 2013). In light of this, the technological impact is transforming the world of work, in which new skills are increasingly required, new policies for business management and in particular new recruitment methods and personnel selection within human resources sector (Wright R. P., Storey J., & Ulrich D., 2019).
From the first industrial revolution to today, many types of changes have taken place, which include some radicals such as digital machines and automated production, generating important effects on the productivity of companies. The main reasons and causes of these changes are the efficiency in the management of resources, the increasingly shorter times for product development and better identification of market demand. Some of the elements developed over the years are applications, smartphones, laptops, etc., which have generated a huge potential for the development of economies in various countries of the world (Cevikcan E., & Ustundag A., 2017). The development of the global market and the growing internationalization and competitiveness of the latter have led to the emergence of a new revolution in the industrial field that is, the fourth industrial revolution, which follows three previous revolutions and has accompanied the development of the concept related to Industry 4.0 and its respective studies. The latter mainly concerns the development and implementation of systems fully automated and intelligent production, which has the capabilities to communicate to the main actors present at the organization various types of information in autonomy. This leads to full automation and digitalization of business processes and the increased use of information technology in the production of goods and services (Piccarozzi M., Aquilani B., & Gatti C., 2018).
A report by McKinsey Global Institute in 2013 describes the fourth industrial revolution as the age of systems defined "cyber-physical" that is, systems that include computing, networking and additional elements characterized by various technologies such as mobile devices, artificial intelligence, robotics, 3D printers, etc., which will have a significant impact in businesses (Piccarozzi M., Aquilani B., & Gatti C., 2018). The latter are well-known examples in the fields related to robotics and artificial intelligence and are defined as smart factories, which are able to produce elements with a high level of complexity without particular changes in the process of production and excluding any human action (Wisskirchen G., et al., 2017). The literature has shown that one of the concepts related to the fourth industrial revolution is the Industry 4.0, which was coined the first time in Germany, the latter has brought and will continue to bring many changes in the global economy and in variables such as investment, growth, employment, etc. This is crucial in the present research since growth and employment are the two most affected by the introduction of innovations mentioned above (Piccarozzi M., Aquilani B., & Gatti C., 2018; Skilton M., & Hovsepian F., 2017).
A large body of literature have focused on examining one of the technologies concerning the fourth industrial revolution and the related concept to Industry 4.0. The present study, therefore, centers on artificial intelligence, which is defined as the capacity of a system to interpret, learn and properly use such learning to

[26] It is the process by which, the engineering and design of digital technologies facilitate the activities and standardize the work processes to the extent that less skilled workers are able to perform work previously done by highly qualified personnel.

carry out the tasks and achieve the set goals. Furthermore, it is possible to state how machine learning[27] is an important part of AI, but not only, since this technology allows data to be perceived (e.g. image recognition, the voice recognition, etc.), to move, manipulate and control objects based on external information through the use of a robot or through devices connected to it (Kaplan A., & Haenlein M., 2019). However, several researches have described how artificial intelligence encompasses many types of technologies and is trying to develop within the respective machines traits similar to the brain of a human being, allowing the latter to be intelligent and operate with different levels of autonomy. Some examples are digital assistants, robots, the facial recognition, etc., which are programmed to demonstrate similar abilities to those human skills like reasoning, planning, problem solving, etc. (Khatri S., Pandey D. K., Penkar D., & Ramani J., 2020).

Figure 3. The Three Stages of Artificial Intelligence (AI)



Source: Author's reworking by Kaplan A., & Haenlein M., 2019, p. 16

According to Kaplan A. & Haenlein M. (2019), AI can be divided into three phases or generations, represented in Figure 3, which will be described in the list below:

-Artificial narrow intelligence (ANI): mainly concerns the specific tasks and therefore it is limited only to the activity or fields for which it was created, not being able to carry out further activities if not trained. This type of AI, achieved by all today's systems which contain it, has allowed for example Tesla to develop autonomous driving, Siri to recognize the person's voice and act on requests, Facebook to be able to recognize faces in photos and insert a tag;

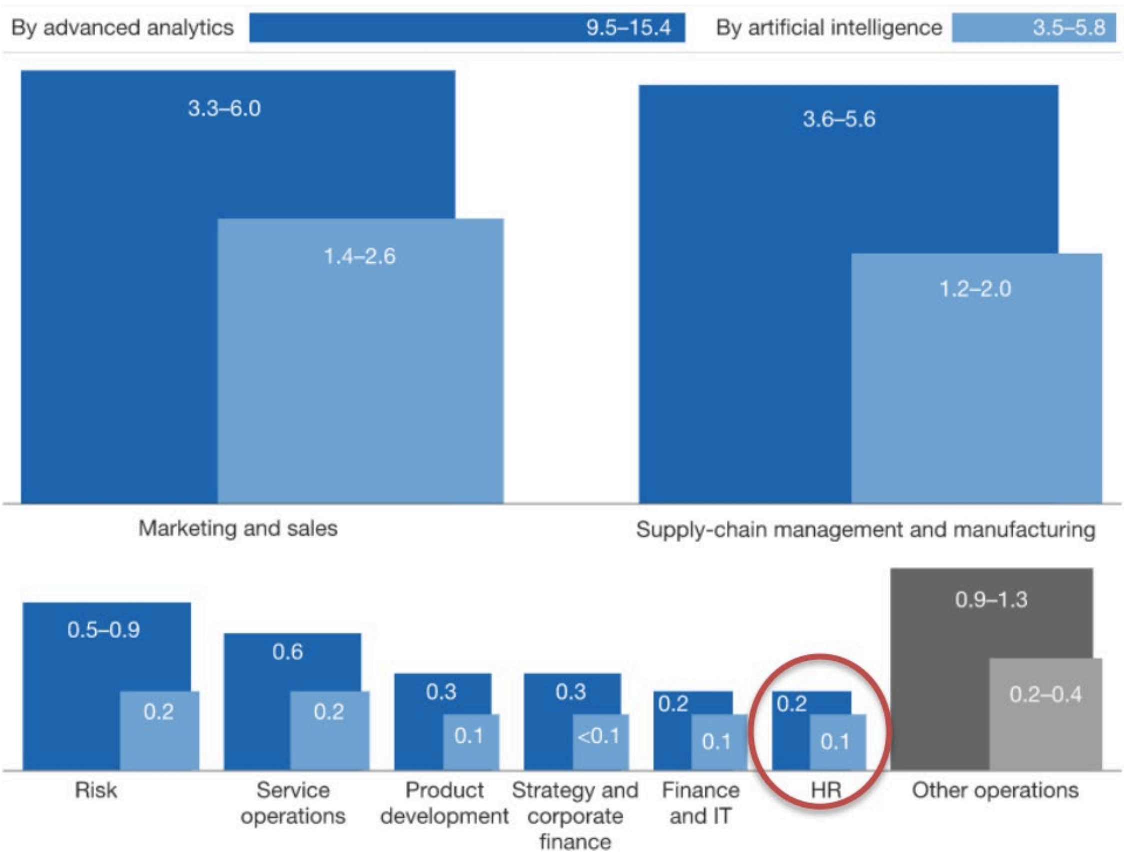
-Artificial general intelligence (AGI): the latter is able to reason, solve problems and autonomously plan tasks for which it has never been designed and trained. Moreover, it will probably be present in the future since it has not been concretely developed yet;

[27] *Machine learning* is a branch of AI, which can be defined through computational methods using data and experience to improve performance or generate accurate predictions (Mehryar M., Afshin R., & Ameet T., 2018)

-Artificial super intelligence (ASI): represents systems with their own consciousness, which make the human being useless and redundant for certain activities. The latter could operate in any area and have social skills, creativity, wisdom, etc. Currently the ASI has not been developed yet, but it will probably be there in the future.

Other authors have tried to explain that this technology is characterized by relevant developments in recent years, becoming increasingly plus an important variable for the industrial economy and also worldwide. This impact can be analyzed through the report called "notes from the AI frontier" of 2018, carried out by the MGI, which identified 400 cases in 19 industrial sectors, focusing the study on nine business functions (marketing and sales, human resources, etc.). The research states that the greatest potential impact of the resulting added value arising from the use of AI, is in marketing and sales. The latter are strongly affected by this technology, and it is in this area that there are some of the biggest opportunities for AI. At the corporate level, each company will have to analyze its own functions to find the more interesting opportunities to use AI, that is, where the latter should invest in the development of this technology (Chui M., et al., 2018; Occorsio E., 2018).

Figure 4. The potential economic impact of AI



Source: Author's elaboration from McKinsey Global Institute, 2018

By analyzing Figure 4, it can be seen how artificial intelligence has an important potential since the latter is able to generate within a few years, on a global scale, a estimated annual value added between 3.5 and 5.8 trillion dollars. Chui M. et al. (2018) in their research have found out that the latter accounts for 40% of the estimated total annual impact between 9.5 and 15.4 trillion dollars that could be potentially caused by all types of techniques defined as analytical or analysis. For AI, MGI estimates that it could potentially, by 2030, foster one boost of global economic activity by about 13 trillion dollars, or further growth of global GDP by 1.2% annually (Rupli D., et al., 2019). Moreover, it can be noted one interesting data related to human resources, for which the impact of AI will generate added value estimated at 0.1 trillion dollars. Hence, I believe that, in light of this, this data is important because it allows to view that also in this function AI is present and is significantly developing both in a current but also future perspective.

2.5.1 The role of artificial intelligence in human resources

According to some, in addition to innovation and technology another important element is human resources, which complement each other with the latter, being key elements for the growth of organization. In this regard, it can be noted how the impact of the fourth industrial revolution and in particular of AI has deeply affected many organizations in their structure. The literature has shown that one of the affected areas is that related to human resources, which is making important steps in using this technology. The latter are struggling with tasks that require an expenditure of time, money and energy, such as the envelope management pay, the search for the right candidate to join the company, etc. In these situations, the technology is able to carry out these activities in a faster way compared to the past. Merlin P. R. & Jayam R. (2018) theorized that AI will help human resources solve business problems complicated analysis and time-consuming ones, allowing that department to work on more productive activities and with greater added value for the organization. Hence, the literature has confirmed that artificial intelligence is replacing the jobs of routines and some activities present in the organization, pushing employees who work in the department to acquire and develop new skills to cope with technology in continuous evolution (Khatri S., Pandey D. K., Penkar D., & Ramani J., 2020).
The objective of AI in human resources is to make accurate the execution of various tasks, remove errors, saving time and costs, and increasing the productivity of this area. This type of goal can be achieved through some elements that make up AI or, algorithms, *big data*[28] and machine learning, which will reduce human intervention in low-value or high-risk activities (Khatri S., Pandey D. K., Penkar D., & Ramani J., 2020). The human resources information system (HRIS)[29], defined the basis for the application of AI in the area of human resources. In addition, technologies related to artificial intelligence offer various types of opportunities to improve and optimize human resources activities, such as recruitment, CV screening, automatic text sending etc. (Bhardwaj G., Singh S., & Kumar V., 2020).

[28] *Big data* is data that exceeds the limits of traditional database tools. The term big data is then used, by extension, also to define the technologies aimed at extracting knowledge and value from this type of data (Rezzani A., 2013)

[29] *Human resource information system* is a type of business information system, which includes systems used to collect, store and analyze human resource information in an organization. The latter includes applications information technology, databases, hardware and software needed to collect, record, store, manage, provide and control data for the optimal management of HR (Harris D. M., & DeSimone R. L., 1994)

According to Yawalkar M. (2019), artificial intelligence and related technologies, such as chatbots, machine learning, process automation through robots, etc., support different activities and processes within the organization (screening, recruitment, etc.). In the list below will be described the role of AI in GRU and in particular in some of the respective systems (Bhardwaj G., Singh S., & Kumar V., 2020):

-Recruitment and selection: the most demanding tasks performed by managers and employees who work in human resources are drawing up a list of the most suitable candidates and CV screening. Through artificial intelligence it will be possible to scan and evaluate CVs to then reject the unsuitable ones, viewing only the most suitable curricula for the offered job position (Garg V., Srivastav S., & Gupta A., 2018). A tool that helps improve the candidate's experience, update requirements and provide recommendations for job position are chatbots, which are important in this phase of recruitment and personnel selection. In addition to the activities previously carried out, the software, related to AI, will be able to formulate assessments through the analysis of interviews in audio mode and video. The latter will be based mainly on the choice of words used by the candidate, speech and body language, thus analyzing the characteristics of the candidates likely to start working for the company (Garg V., Rani S., & Matta H., 2019; Yawalkar M., 2019);

-Training and development: through AI it is possible to suggest videos or programs learning related to work tasks. Thus, this type of software reads and automatically analyzes the didactic documentation and elaborates the programs of learning for the employee. In addition, it provides real-time feedback during the training and can modify the latter according to its progress, for example dependency answers to questions (Gupta R., Bhardwaj G., & Singh G., 2019; Yawalkar M., 2019). This technology could be used to involve more actively employees and try to generate a more innovative type of learning within the organization;

-Performance management: the evaluation of employee performance within a working environment is a difficult task to perform for organizations, since in some cases there is the presence of prejudices. AI will try to reduce bias and rectify them through feedback to employees in charge of evaluating staff. This technology continuously evaluates and monitors the objectives, the joint commitment of the teams composed of collaborators and better manage the workforce internally to the organization. AI collects data related to different perspectives within the company, such as those on employee performance, their level of involvement and the causes of turnover. In addition, it can predict and analyze performance indicators for brilliant collaborators but also for employees, who do not work optimally or want to change their job position (Rajesh D. S., Kandaswamy M. U., & Rakesh M. A., 2018);

-Employee retention: one of the most demanding tasks, in addition to the selection and recruitment of candidates, is the retention of employees to keep them in the company. AI can enable companies to overcome this obstacle by allowing them to predict the requirements and actions of each individual collaborator in the organization. So, it gives the opportunity HR leaders to take actions, which anticipate the incident, or in this case the loss of the collaborator (Bhardwaj R., 2019).

In addition to the role of artificial intelligence in the organization, the latter is classified in three main forms, which concern the human resources system. They are speech recognition, bots and algorithms. The list below will explain the three main forms of AI present in the RU (Bhardwaj G., Singh S., & Kumar V., 2020):

-Voice recognition: this type of AI converts the information received into words, videos and websites searched on the Internet, because it automatically transmits the information to analytical tools through a discursive or text format (McGovern S., et al., 2018). The main function of this technology is to perform types of actions based on voice commands, such as opening files, searching websites, controlling devices in the office and other functions for which voice control can be used (Yawalkar M., 2019);

-Bots: the latter are used by the main search engines to find words on the Internet or key phrases. This type of tool is suitable for chatting, asking questions, giving instructions and further activities useful for the organization and for the candidate;

-Algorithms: in artificial intelligence they represent instructions and codes that such technology must follow in order to be guided in the various functions in which it is used. An algorithm developed in the correct way allows to ensure that various functions in HR are automated. Some of these functions are the dissemination of information to stakeholders, the supervision of key performance indicators, the monitoring of activities of potential candidates and employees on social media (Garg V., Singhal A., & Tiwari P., 2018).

In the previous researches, it was possible to see how some activities, related to human resources are successfully carried out by artificial intelligence. However, the latter in some situations presents difficulties in performing certain tasks, which currently only a human being is able to complete. In the table below will be drawn up a list of benefits and challenges associated with this technology.

Figure 5. Benefits and challenges related to artificial intelligence in human resources

Benefits	Challenges
• The workload of employees, linked to administrative tasks, is reduced; • Contributes to increased efficiency within the company and a saving of time. Minimizing the probability of error; • Helps to identify the right candidates for the jobs offered and also assists the acquisition of talents in the company; • Minimizes the behavior that causes bias in decision-making processes; • The results, processed through use of AI, are more accurate; • Allows certain human limits present in the workplace to be exceeded	• Difficulties of employees in using and learn from AI-related tools; • Implementation of AI systems can generate within the organization a sense of fear among collaborators; • Difficulty in identifying the right candidate in order to manage this technology within the holding; • Risk of AI exceeding authority and role related to human resources in the process decision-making of a company; • Complexity in interpreting the output of the algorithm, since the more the latter become accurate, the more difficult they are to understand and explain (Tambe P., Cappelli P., & Yakubovich V., 2019).

Source: reworking of the author from Yawalkar M., 2019, p. 23

Several researchers have described how the technological progress and the greater possibilities for calculation have led to the creation of new opportunities in human resources and especially in *recruiting*[30], in which all the elements linked to the logical process can be managed by a machine. The latter can offer equivalent performance in terms of quality and quantity higher than that of an employee. Moreover, through machine learning, the machine can learn from various situations without be controlled by someone (Alessandrini G., 2019). The concept related to recruitment 4.0 is characterized by the introduction of digital instruments. Alessandrini G. (2019), in particular, have supported the idea that the latter have significantly changed the process of recruitment and generated incredible opportunities in the field of human resources. In the list below will be described some of the digital tools concerning the recruiting of candidates (Verhoeven T., 2020):

-Applicant tracking system (ATS): are a type of system (software) present sine several years in the field of human resources, which allow the candidate to send electronic application to the company by filling in a specific form. The latter is written through a digital user interface, which allows the company, mainly recruiters, to process and organize applications within the system. ATS helps organizations and is the predominant form still present, since the classics paper applications are always fewer and companies in some cases do not accept them. Furthermore, this system offers interesting advantages for organizations but also for people who apply, such as saving time (e.g. immediacy in the arrival of data and documents) and costs (e.g. photocopies, mailing, etc.);

-Mobile recruiting: the use of the Internet by users has changed in recent years, as smartphones have become an important part of everyday life of each individual. Through the latter, candidates can inquire about the employers and the place of work, for this reason companies must adapt and develop always more accessible content to smartphones. Thus, mobile devices also have an important impact on candidate recruitment, but currently many career sites or forms of applications have not been optimized for mobile use;

-Recommendations programs: this type of program offers referrals to the company on possible candidates, which are provided by the present employees at the latter. It is defined as one of the best recruitment channels for a long-time term, for this reason companies are investing heavily in these instruments. Some advantages are that recommended employees integrate more easily into the company and quit less frequently. Several companies are trying to optimize these programs digitally, they use a platform to simplify processes of recommendation (e.g. Firstbird, Talentry, etc.). In these platforms every recommender can see recommended people and their status;

-Search engine marketing (SEO): search engine marketing aims to guide users towards the page sought, thus improving the availability of websites. The main reason for using SEO is related to the fact that many candidates through this type of search engines (e.g. Google) identify new offers of work. In addition, the employer companies are dealing with this topic because they want to publicize their job vacancies, the

[30] *Recruiting* concerns the entire process related to the recruitment and selection of candidates for a particular position offered by the company (Licata P., 2019)

company and their respective website. One of the tools that gives the possibility to advertise the company is for example Google Ads;

-Video interviews: this type of interviews has become an important alternative to telephone interviews, but not completely replacing the face to face ones carried out at the company headquarters. In addition to the video interviews carried out "live" through a digital platform (e.g. Microsoft Teams) there is another type of interview, in which the candidate answers to online questions while being filmed. After that the recording is sent the holding, which evaluates the results;

-Robotic process automation (RPA): many companies use software robots, even called bots, to automate simple and standardized activities such as those present in the recruiting. These are mainly related to administrative tasks that require a large expenditure of time. For these types of activities, defined as simple, is not always necessary AI but only RPA is enough;

-Chatbot: it is possible to find them in messaging services and they are available to respond immediately and automatically to questions related to the company or work, providing also some recommendations. An added value present in this type of instrument is related to the fact that it is available 24 hours a day and in addition, one can make an application through the latter. The chatbot in most cases is connected directly to ATS, so that the data appears directly in the system. In addition, this technology allows an improvement of the *candidate experience*[31], for example candidates no longer have to wait weeks to receive feedback but can immediately see them. This technology solves the main issues in the decision-making process of resources that is, it allows to remove almost entirely the prejudice, accelerate and shorten some processes, perform repetitive tasks (e.g. remember appointments, answer to repetitive questions, schedule appointments, etc.) and save costs. The main objectives of chatbots are improving the candidate experience and providing more time to professionals of human resources to be personally with the candidates, without taking away work at the recruiter but being only a support tool (Verhoeven T., 2020);

-Augmented reality: it is a tool linked to a future perspective, which elaborates reality in virtual way through a computer, special glasses, smartphones, etc. It is also used in recruiting and could generate added value in the process. It can be used to allow candidates to take a virtual tour within the organization or carry out interviews, also virtual, in the selection process;

2.5.2 The use of learning algorithms for screening candidates

From a managerial point of view, artificial intelligence is an interesting technology in this area, since it can help HR responsible and managers to accelerate the daily, repetitive and low-added value work. From a report prepared by Linkedin, in which were interviewed almost 9000 recruiters and hiring managers from 39 countries, it emerged that 76% say that the future impact of AI on recruitment will be very significant,

[31] *Candidate experience* is the individual experience or impression of a potential candidate towards the employer. In particular, it regards the application and selection process and all direct and indirect points of contact with the company (Verhoeven, 2016)

especially in research and in screening of candidates (Spar B., Pletenyuk I., Reilly K., & Ignatova M., 2018). Hence, the literature has confirmed that this technology is becoming an important element in the future development of human resources because it provides powerful *database*[32] and analytical support related to the recruiting process, which includes the screening of curricula and the interview of candidates. However, it is important to remember how the latter is an auxiliary decision-making system, therefore it is not intended to replace employees who work in human resources but help them in carrying out various activities present in the department.

Considering this, as explained before, it was found that some of the digital tools listed above contain AI within them. The first is the video interview, which is used to select candidates. In this case the AI can divide the latter according to the questions, for example, and convert it into a set of data points, calculated based on facial movement, facial expressions, intonation and words used. After that, the results are compared with other candidates or employees of success who play a similar role in the company. This technology helps to improve the efficiency of the interview without reducing its quality. Furthermore, facial recognition techniques can be used to determine whether the candidate is consistent with the document, thus preventing others from carrying out, for example, the interview or a certain test for him (Jia Q., et al., 2018; Kulkarni S. B., & Che X., 2019).

An additional tool that includes AI are chatbots, which automate tasks that involve a high expenditure of time related to the process of screening and evaluation of candidates. This type of software uses *natural language processing*[33] and machine learning to show an intelligence similar to human and involve the candidates. Moreover, it interacts with them through certain platforms (e.g. SMS, e-mail, social media, etc.), starting a real-time communication in the form of a screening interview, in which performs several assessment tests, answers the multiple questions of the candidates and finally, also elaborates a ranking of them based on the assessments made. Consistent with this definition, Kulkarni S. B. & Che X. (2019) affirmed that chatbots allow to eliminate the gap between recruiters and candidates and schedule interviews between them. A similar view is shared by Hmoud B. & Varallyai L. (2019).

Then there are the algorithms, which are characterized by tools based on the identification of candidates, through the use of algorithms learning related to machine learning. The latter have the task of learn from the data, apply it all and bring the candidates to the attention of the recruiter (Kulkarni S. B., & Che X., 2019). Finally, AI-related software is also being implemented in order to be able to develop a better screening of candidates and to see whether to consider them according to past decisions (Sekhri A., & Cheema J., 2019).

It has been shown from some previous researches that there are different digital instruments, which include AI within them. Therefore, there are different companies and solutions offered by international and also Swiss companies linked to artificial intelligence in recruiting candidates.

---

[32] The *database* is represented by a set of information or data, which are usually stored electronically in a computer system (Oracle, s.d.)

[33] *Natural language processing* is a branch of artificial intelligence that allows computers to analyze and understand human language. Giving to the person the opportunity to have an interaction with their computer, without the specific use of programming (Frankenfield, n.d.)

As digitalization progresses, companies must prepare to invest in it and especially in the inevitable introduction of artificial intelligence in the human resources management, for which even at the Swiss level companies are moving. From a study prepared by the University Professional Management School of Freiburg (Hochschule für Wirtschaft Freiburg) and presented to the HR association in 2019, could be seen a development of the concept linked to the adoption of artificial intelligence in GRU by companies in Switzerland. The study is mainly based on the relationship between AI, GRU and organizational behavior. The latter aims to investigate the perception of the adoption of AI in GRU processes (Baldegger R., Caon M., & Sadiku K., 2020).
This research was conducted taking into account the responses of 310 members of the French association of human resources and employees of different companies, in which could be noted a positive outcome from the introduction of AI in human resources. Furthermore, a very interesting fact emerged related to the fact that one of the main problems concerning the adoption of AI in human resources is the lack of skills and training (Baldegger R., Caon M., & Sadiku K., 2020). From the analysis it also resulted like more than 30% of the companies present in Switzerland, of which 20% are small and medium enterprises (SMEs), has developed AI-related projects in the area concerning the HR sector.
The literature has shown that the main potential impact of AI in this area can be identified in processes related to the attractiveness of candidates, their recruitment and training. In this category of processes, it emerged that AI will be useful mainly for two features namely, the first is linked to the possibility of drawing up a balance sheet of current and missing competences within the company. Instead, the second is the possibility of relating the candidate and the job position offered according to the skills of the latter. Thus, as well as at an international level, artificial intelligence is also present in human resources, especially in recruitment.
The list below will describe the importance of AI in this process (Geetha R., & Bhanu S., 2018):
-Time saving: this technology allows to save time while performing repetitive tasks such as CV screening;
-Talent mapping: AI accompanies human resources in identifying and acquiring the best talents needed by the company. Furthermore, it addresses the candidates and places them in the appropriate job position according to their skills, characteristics and abilities;
-Cost savings: the tasks related to the acquisition of the right candidate take place through AI. In some cases, the latter reduces the outsourcing of this activity related to recruitment, so it is possible to save costs;
-Reduction of bias in hiring: this process takes place through the use of AI and avoids the involvement of human being. Leading to an impartial and without prejudice screening and selection towards candidates;
-Employability with quality: AI technology enables human resources, through the use of large amounts of data, to carry out the screening and selection of candidates. It also minimizes prejudices and enables to improve the quality of recruitment;
-Availability of assistance: through AI, employees and applicants receive updated information and get answers to questions immediately. This leads to satisfaction of the latter and involves them more within the company. Moreover, it increases the quality of some of the services offered by the company.

Therefore, AI is an interesting and significant tool for this type of process related to human resources. According to some, in the future there will be new technologies and new tools that will help this sector to carry out some of their activities, such as face to face interviews, in which robots will be used to communicate and socialize with the candidate. This technology arouses in employees, but in general in all people, a sense of fear but also of hope. Fear since there is a risk that the algorithms and other tools linked to AI make the career opportunities depend solely on the basis of the choices of a machine. In addition, there is a fear that they will make even more decisions related to daily life. Instead, hope because there is the belief that this type of software improves and bring more fairness and less prejudice in recruitment, but also in other processes. Furthermore, it allows the recruiter to perform better and with greater quality his work (Verhoeven T., 2020). Currently the progress in this field is increasingly important, but there are still problems related to this technology. As for example in the case of Amazon, which had developed an algorithm of screening that was not optimal because it discriminated against women applicants. Or Google image, which in 2015 was criticized because through an algorithm classified the people of color under the category of gorillas (O'Neil C., 2018; Hern A., 2018). For this reason, there is still a need to invest and develop this technology, making it optimal and error free in use. This is crucial in the present research since many people, in this case candidates, will depend very much on that, namely the output of an algorithm linked to AI.

2.6 Research questions and conceptual model

Artificial Intelligence is increasingly being used in recruitment processes, but its implementation can have a dark side. AI has the potential to revolutionize recruitment processes by making them faster, more efficient, and fairer. However, AI can also perpetuate human biases and lead to discriminatory outcomes in the hiring process (Dastin J., 2018). Despite the efforts of organizations to ensure that their recruitment processes are unbiased, implicit bias can still influence decisions about who to hire. For example, AI algorithms can be trained on biased data, leading to discriminatory practices that disadvantage certain groups (Buolamwini J., & Gebru T., 2018). One study found that even with identical resumes, candidates with "white-sounding" names were 50% more likely to receive callbacks than candidates with "black-sounding" names (Bertrand M., & Mullainathan S., 2004). Moreover, job ads can contain biased language that discourages certain groups from applying for positions (Gaucher D., Friesen J., & Kay A. C., 2011). Additionally, AI can be used to screen job candidates based on criteria that are not job-related, such as their social media activity or online behavior, which can violate privacy and lead to unfair hiring practices (Brougher J., 2019). It is essential that organizations using AI in recruitment processes take steps to ensure that the algorithms used are transparent, accountable, and unbiased to avoid perpetuating existing social biases (Acas, 2020).

Previous research on discriminatory algorithms and bias in the context of recruitment has identified several issues related to the use of AI in hiring processes. However, there are still gaps in the literature that need to be addressed. For example, many studies have focused on identifying bias in AI algorithms, but few have proposed concrete solutions for addressing the problem (Suresh H., & Guttag J. V., 2019). In general, there is a lack of research on how different AI algorithms compare in terms of bias and accuracy, making it difficult to determine which algorithms are most effective at reducing bias in recruitment (Dastin J., 2018).

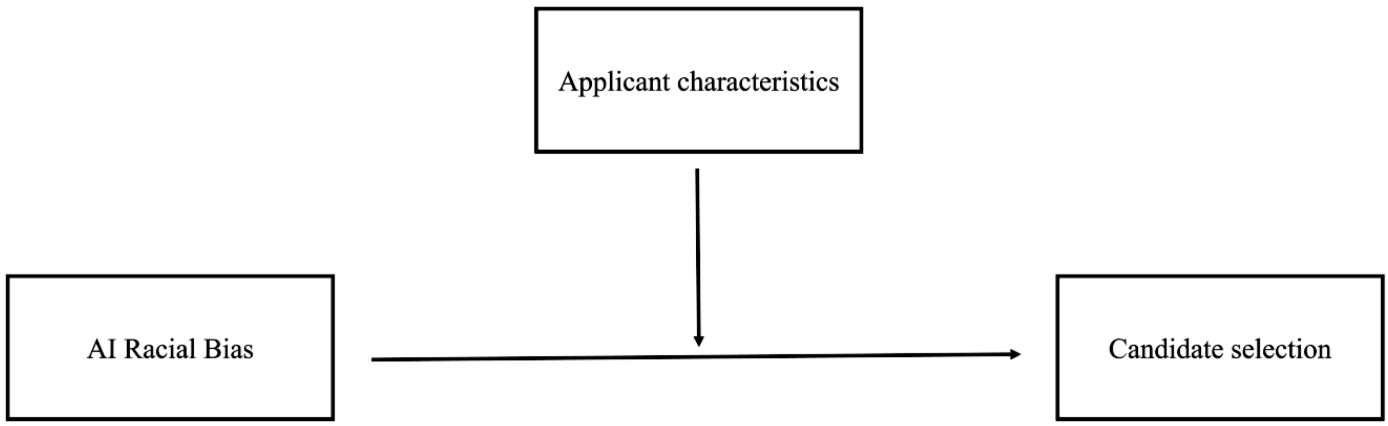
Another gap in the literature is the lack of research on how job candidates perceive the use of AI in hiring processes, and how their perceptions affect their job search behavior (Ruppel C. P., et al., 2020). More research is needed to understand these issues, as well as the broader ethical and social implications of AI in recruitment. Furthermore, more studies examining recruitment and workplace environment would provide a fuller view into the influences shaping the experiences of women and gender non-conforming people in technology.

In order to address these specific gaps in the literature, this research aims at understanding how AI racial bias, in the specific context of recruitment, might influence the selection of candidates. Moreover, I intend to study the applicant characteristics as a potential moderator of the relationship between AI racial bias and candidate selection, with the goal of finding out whether a possible moderation exists.

Given the previous discussion, I have formulated the following research questions: *How does the presence of AI racial bias, in the context of recruitment, influence the selection of candidates?*

Do the applicant characteristics moderate the relationship between AI Racial bias and candidate selection?

Figure 6. The conceptual model



Based on the research questions, the conceptual model includes the following variables: AI racial bias, as an independent variable; candidate selection, as a dependent variable; and finally, applicant characteristics, as a moderator.

The independent variable refers to the presence of racial bias within the algorithms used in AI systems and it is the variable that is being manipulated in the research. Indeed, the variable "AI racial bias" can be better explained as the potential for artificial intelligence systems to exhibit biases towards individuals or groups based on certain characteristics, namely the tendency of artificial intelligence systems to make decisions or predictions that systematically disadvantage or advantage people based on their race or ethnicity. This can occur when the AI system is trained on biased datasets that reflect real-world biases and inequalities, or when the algorithms used to make decisions incorporate discriminatory assumptions.

For example, an AI-powered recruitment tool may be trained on a dataset that predominantly consists of resumes from white applicants, which could lead the tool to favor white applicants over applicants of other races. This bias could be unintentional, as the tool is simply reflecting the patterns in the data it was trained on, but it could have significant real-world consequences, such as perpetuating systemic discrimination in hiring practices.

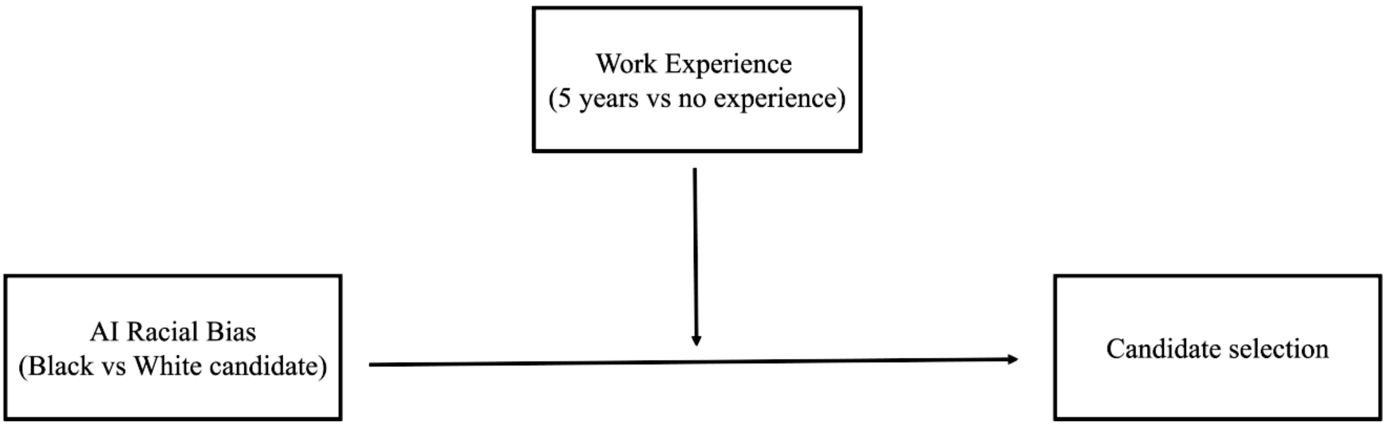
The dependent variable indicates the process of selecting a candidate for a job and it is the variable that is being measured in the research. Going forward with the explanation of variables, "applicant characteristics" refers to the personal attributes of a job applicant, such as their race, gender, age, education level, and work experience, which could affect how AI racial bias influences candidate selection.

More specifically, as it is possible to see in Figure 7, the independent variable "AI racial bias" represents the race of the candidate, indeed black or white candidate.

The dependent variable, labeled "candidate selection", indicates the probability to hire the job applicant. Furthermore, as regards the applicant characteristics, among the personal attributes of a job applicant, I decided to focus the attention on the variable "work experience". Therefore, I differentiate between a candidate with 5 years of work experience and a job applicant with no work experience.

The goal is to study if an interesting moderation exists, considering the "work experience" as a potential moderator of the relationship between AI racial bias and candidate selection.

Figure 7. The conceptual model with variables explained



As a result of this argumentation, I propose the following research hypotheses:

H1. White candidate (vs black candidate) increases candidate selection. Respondents are more willing to choose a white (vs black candidate).

H2. 5 years of work experience (vs no work experience) leads to a higher candidate selection. Respondents are more willing to choose a candidate with 5 years of work experience (vs no work experience).

H3. Work experience positively moderates the relationship between AI racial race and candidate selection.

I would expect that if the candidate has no work experience, its race may influence the candidate selection. No differences are expected for candidate selection for a candidate with 5 years of work experience.

CHAPTER 3: METHODOLOGY

After having analyzed the literature regarding recruitment, AI biases and the use of algorithms for screening candidates and having assessed that there is indeed a scientific gap in the literature regarding such constructs that needs to be filled, a thorough statistical investigation is initially needed. In order to address this specific gap in the literature, the present study aims at understanding how the presence of AI racial bias, in the context of recruitment, might influence the selection of candidates.

No one has ever analyzed such relationship before and actually, academic literature on recruitment lacks studies concerning the race of candidates and how they could influence the candidate selection. My research aims to fulfill some of the shortcomings that have existed so far.

Moreover, in accordance with what has been argued up to now, the research model includes the "work experience" of candidates as a potential moderator, which is a variable that was not included in studies before. In particular, as it can be seen at the end of chapter 2, the conceptual model of the research aims to study the relationship between 2 different types of race of candidates (black vs. white), and the candidate selection, having as a moderator the applicant characteristics, more specifically the work experience (no experience vs. 5 years' experience). In this way, the study could demonstrate if there is a sort of racial prejudice about different types of race of potential candidates and the effect that they can have on consumers, as they were real recruiters, during the choice of selecting and hiring a candidate.

In order to conduct my research, I decided to use a mixed-method approach, which involves combining both quantitative and qualitative research methods. This procedure can offer several advantages: indeed, by conducting an experiment on a population and collecting quantitative data, I was able to gain statistical information about the outcomes of the intervention.

Additionally, using a qualitative approach, such as conducting interviews with real HR recruiters, provided more in-depth insights into the experiences and perceptions of those involved. By combining these methods, it was possible to achieve a more comprehensive understanding of the phenomenon I was studying, allowing me to triangulate the findings and strengthen the overall validity of the results.

Furthermore, using a mixed-method approach helped overcome limitations associated with using only one method, such as the potential for bias or limited scope of findings. Overall, the use of this approach provided me with a more nuanced understanding of my research questions and the context in which they exist.

3.1 Study 1

In order to verify the hypotheses, an experimental design method was used to test the relation between the variables. The data collection process will start with the definition of the sample size and features. After having selected the most suitable item scales from the current literature, the data for the main study was collected through an online Qualtrics experimental survey and set with four random scenarios.

Regarding the type of experimental design, it was used a between-subject with a 2x2 factorial design, where each participant took part only in 1 experimental condition, in order to see the analyses of differences between groups of participants. Indeed, the dependent variable is candidate selection, the moderator is applicant characteristics, that is important from a theoretical point of view.

Moreover, the independent variable AI racial bias (race) has 2 levels (black candidate vs. white candidate), and even work experience has 2 levels (no experience vs. 5 years). Therefore, as shown below in the Figure 8, four experimental groups are needed (2 x 2 = 4).

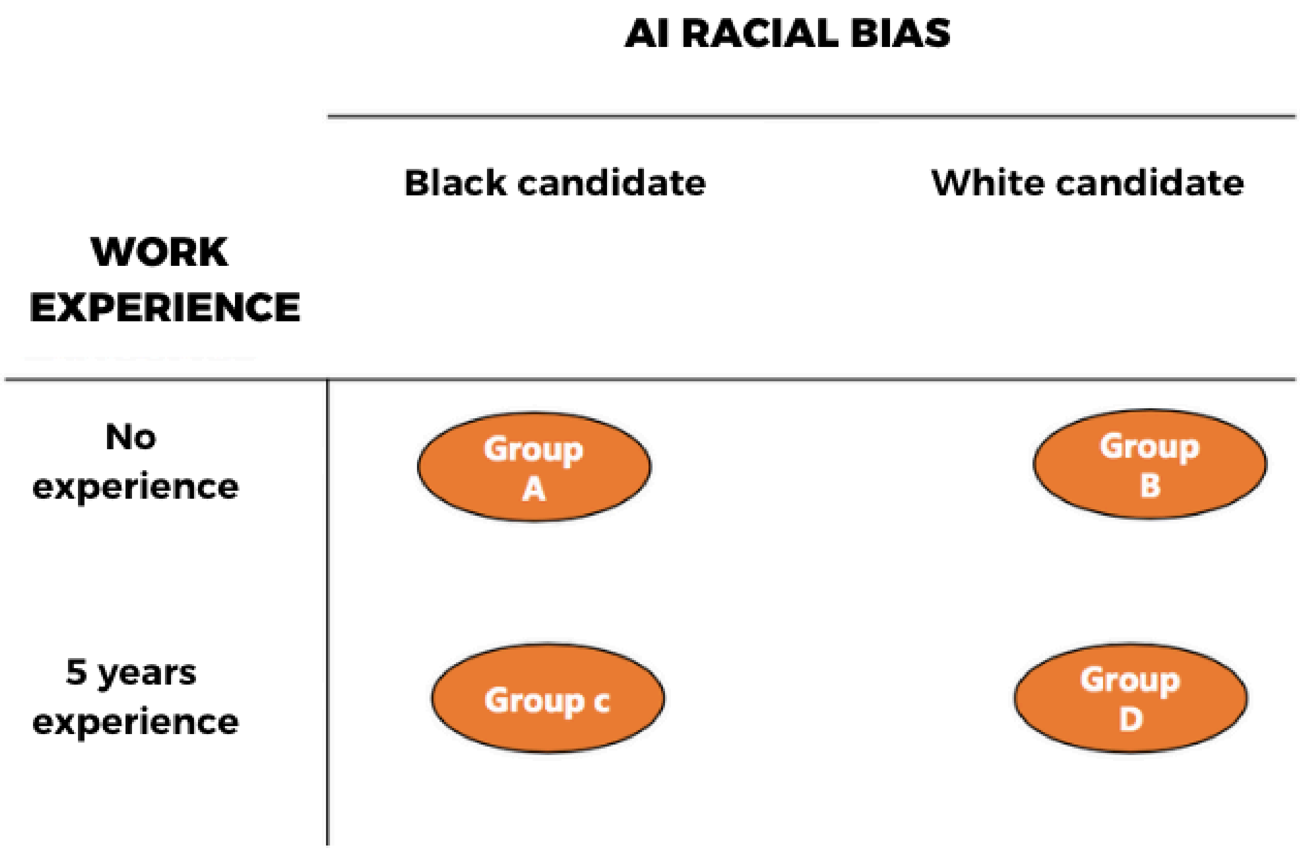
The objective of this research was to know if the decision of candidate selection changes if the description of the potential candidate showed 2 different types of AI Racial bias (black vs. white) and if the candidate had more or less years of work experience (0 years vs. 5 years).

The four scenarios were portrayed as follows: one in which the respondent was presented with a black candidate with no work experience or with 5 years of work experience.

On the other hand, the other two scenarios, concerned a white candidate with no work experience or with 5 years of work experience.

After having reached a significant number of respondents, a data scraping was performed, and the data were processed through the use of SPSS software.

Figure 8. Type of experimental design



The questionnaire was built using Qualtrics and was created and distributed only in Italian, since it was the language spoken by the respondents.
At the beginning of the survey, respondents were asked to imagine that they were HR recruiters for a bank looking for a young talent to fill the role of financial advisor.
In the scenario text, it was also explained to them that their mission would have been to identify and choose the right candidate to ensure the company's success.
Then, in the indications for the stimulus to see, they were advised that a description of a potential candidate would be shown.
Respondents were requested to look carefully at the scenario proposed and read the description in order to be able to give judgments about it later.
Afterward, respondents were randomly exposed to one of the four scenarios described before through the use of a randomization block with one evenly presented scenario.
Immediately after viewing one image, followed by the description, as if it were a kind of resume and respondents were in the screening and selection phase of a candidate, the scenarios were followed by one matrix type of questions in which the respondent had to select from a scale from 1 (not at all likely) to 5 (extremely likely) the level of probability of the statements shown.
The questions regarded, in particular, the likelihood to hire a candidate, which means the probability that the candidate will be selected/hired or not.
Indeed, it was asked to indicate to what extent the following affirmations were likely:

1. I would interview the candidate for the job.

2. I would personally hire the candidate for the job.

3. The candidate would be hired for the job.

In particular, I have found this pre-validated scale in the literature, in the section of human resource management, with the name of "Hireability Scale" by Rudman and Glick (2001): 5-point Likert scale with 3 item measure ["Participants also indicated on three scales ranging from 1 (not at all likely) to 5 (extremely likely) the probability that (a) they would interview the applicant for the job; (b) they would personally hire the applicant for the job; and (c) the applicant would be hired for the job"].
Subsequently, in the last part of the survey, after it was said that the study was almost over, participants were asked to answer to some final questions about themselves. More specifically, these were demographic questions, such as age, gender and occupation.
The manipulation of the independent variable took place through the creation of graphic-descriptive stimuli concerning different types of applicants for the job.
Therefore, the respondents were asked to carefully visualize the scenario, read the description and respond accordingly to the questions.

As can be seen from the scenarios, it is important to underline that only the candidate's image (black or white) and work experience (none or 5 years) vary.
The resumes are the same and the rest of the other characteristics of the job applicants remain constant in all scenarios, to see exactly where the effect comes from.
In fact, both candidates have the age of 28 years and a degree in Economics. Moreover, among the languages known by the job applicants we find Italian, English and French.
However, the main skills and competences that are indicated for both are the following: MS Office, software development, business model analysis.
In Figure 9, it is possible to see an overview of all four scenarios, characterized by the image of the candidate and his description below, here translated in English.

Figure 9. Stimuli

Scenario 1

Black candidate - No work experience

Age: 28
Education: Degree in Economics
Work experience: No experience
Skills: MS Office, software development, business model analysis
Languages: Italian, English and French

Scenario 2

Black candidate - 5 years of work experience

Age: 28
Education: Degree in Economics
Work experience: 5 years of work experience
Skills: MS Office, software development, business model analysis
Languages: Italian, English and French

Scenario 3

White candidate - No work experience

Age: 28
Education: Degree in Economics
Work experience: No experience
Skills: MS Office, software development, business model analysis
Languages: Italian, English and French

Scenario 4

White candidate - 5 years of work experience

Age: 28
Education: Degree in Economics
Work experience: 5 years of work experience
Skills: MS Office, software development, business model analysis
Languages: Italian, English and French

Moving on to the description of the sample, respondents were reached through "convenience sampling", which is a sort of non-probability sampling in which a sample is taken from a population segment that is close to hand. Specifically, most of respondents were recruited from my contact list by sending to them an anonymous link to the survey through WhatsApp and other social media platforms.

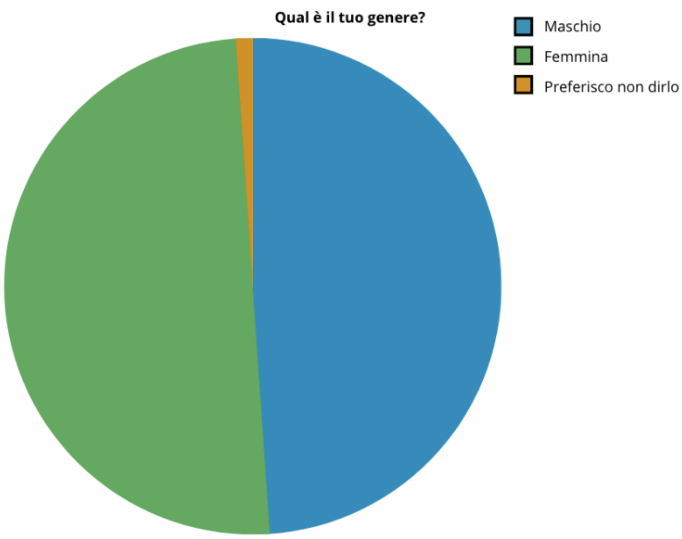
In addition, I even administered the questionnaire through a QR code to people I met in person. Overall, the data collection process took approximately two weeks and finally, a total of 379 people recruited exclusively in Italy replied to the questionnaire.

The data collected through Qualtrics Survey Software was transferred to SPSS Statistics version 28 for the analysis. I cleaned the dataset trying to have a sample eliminated from any distortions, and especially excluding people who were too young or too old in age, in order to make it as close as possible to the recruiters' working age. The final sample of respondents, following a clean-up of the dataset involving 5 previews, 48 distortions and 44 respondents who did not finish the survey, is made up of 282 participants whose answers were used in the analysis.

The gender distribution included 48.9% of men (n = 138), 50.0% of women (n = 141) and the remaining 1.1% (n = 3) preferred not to specify.

Figure 10. Gender distribution

Qual è il tuo genere?					
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulativa
Valido	Maschio	138	48,9	48,9	48,9
	Femmina	141	50,0	50,0	98,9
	Preferisco non dirlo	3	1,1	1,1	100,0
	Totale	282	100,0	100,0	

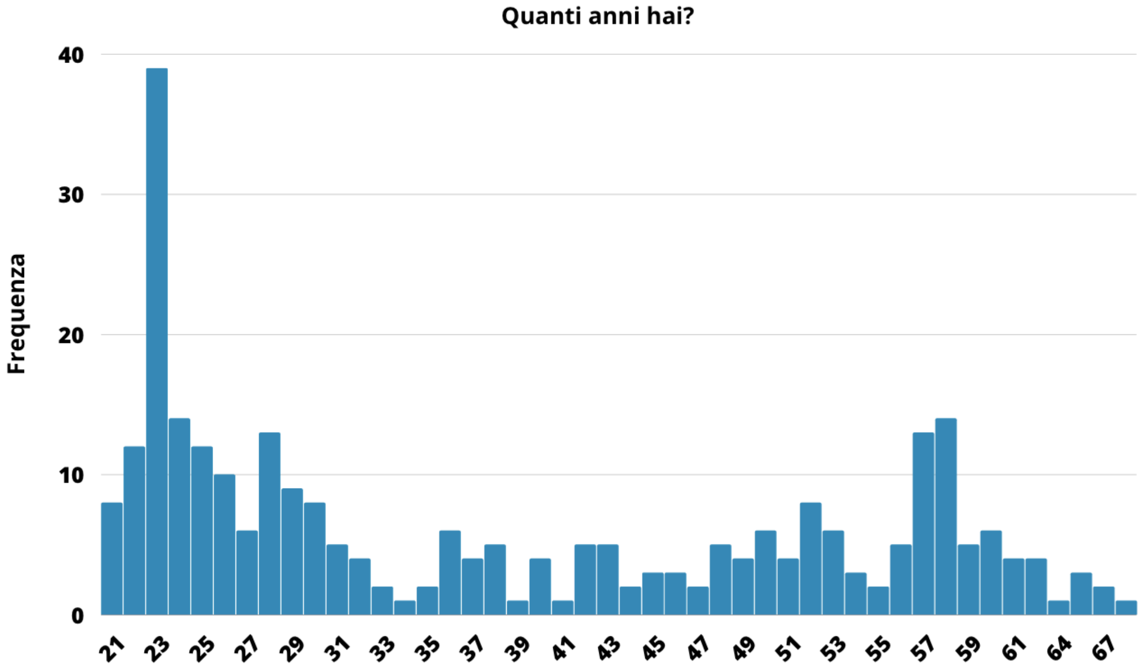


The age of the sample varied selecting from the sample respondents from a minimum of 21 years old up to a maximum of 68 (Mage = 38,1915; SD = 14,46602). In the table below it is possible to see some parameters and, in the bar graph, the distribution of age by frequency.

Figure 11. Age

Descrittive

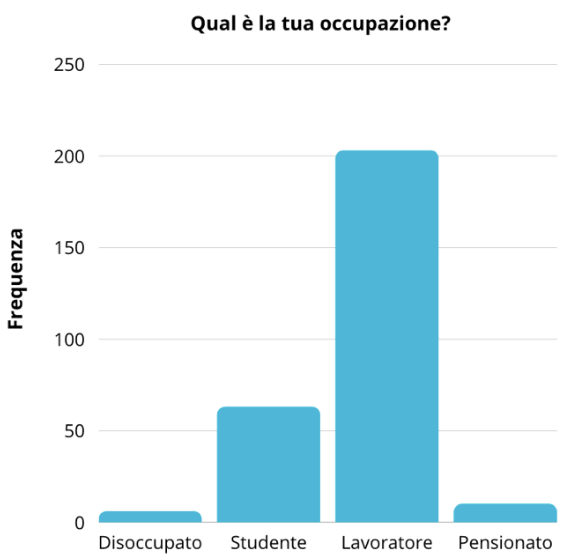
Statistiche descrittive									
	N	Minimo	Massimo	Media	Deviazione std.	Asimmetria		Curtosi	
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Errore standard	Statistica	Errore standard
Quanti anni hai?	282	21,00	68,00	38,1915	14,46602	,394	,145	-1,394	,289
Numero di casi validi (listwise)	282								



Regarding the occupation of the people who took part in the survey, it included 72,0% of workers (n = 203), 22,3% of students (n = 63), 3,5% of retirees (n = 10) and 2,1% of unemployed (n = 6).

Figure 12. Occupation

Qual è la tua occupazione?					
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulativa
Valido	Disoccupato	6	2,1	2,1	2,1
	Studente	63	22,3	22,3	24,5
	Lavoratore	203	72,0	72,0	96,5
	Pensionato	10	3,5	3,5	100,0
	Totale	282	100,0	100,0	



3.2 Study 2

After having distributed the questionnaire to ordinary people and after having analyzed the results with SPSS, I decided to conduct interviews with real recruiters both to go deeper into the topic of interest and to understand their opinions, also discussing with them the results obtained from the previous experiment.
In this way, the feedback from the recruiters better validated the results of the experiment, indeed I was able to obtain a complete all-round view and certainly the opinion of expert people in the field added value to the results I had found.
Regarding the Study 2, I interviewed 10 recruiters from different companies, mainly from major banks and financial, engineering and IT consultancy firms. This choice was made to stay in line and as coherent as possible with the scenario I invented for the experiment, in which I asked ordinary people to imagine they were recruiters for a bank that was looking for the figure of a financial advisor.
However, I even asked to a couple of recruiters from different types of companies, in the fashion or luxury sector, to see if I would have found any differences once I showed them the stimuli of the candidates to be hired.
In order to explain the process of carrying out the interviews, on one hand most of them were conducted by me in presence, on the occasion of the Luiss Career Day, where around 220 employers were present, and I had the opportunity to meet many human resources representatives. Actually, finance, consulting, corporate and legal were all sectors represented at the event.
While, on the other hand, some interviews were scheduled online on Teams. As regards the latter, I contacted the HR recruiters via email and scheduled the interviews based on their availability. The average duration of the interviews was 20 minutes each, with a range that varied from 15 to 40 minutes.
Therefore, I conducted these interviews, first of all introducing myself and explaining that the purpose of the interviews was for my experimental thesis, then it was my concern to assure all the recruiters that the short interview would have remained anonymous, thus their name nor that of their company would not have appeared.
Furthermore, the questions asked to the recruiters concerned both how they evaluated the candidate selection process in the company, how they ensured that candidates were chosen objectively and impartially, and whether there were policies in the company to promote inclusion and diversity. These were more or less the topics and themes in general addressed in the various interviews.
More specifically, the almost ten questions were the following:

1. How would you describe the process of selecting candidates in your company?
2. What are the main criteria you use to evaluate candidates?
3. What skills do you value most in candidates?
4. How do you ensure candidates are evaluated objectively and impartially?

5. How do you evaluate candidates to ensure that there is no discrimination on the basis of race, gender, sexual orientation or other protected characteristics?
6. How is the experience and competence of candidates from different backgrounds assessed?
7. If there are, what are the company's policies to promote diversity and inclusion in the workplace?
8. Are there people of different ethnicities within your company? If so, what percentage?
9. How many people of different ethnicities propose for a position from you?

As a result, the interviews took place as follows: first of all, I asked recruiters the series of interview questions I had prepared, leaving them the space to answer them openly. After the end of the questions, I briefly explained how I had conducted the experiment and showed the graphical stimuli of the experiment afterwards. Moreover, I asked them, if they were faced with the four scenarios, how they would behave, who they would choose and why, being interested in understanding more deeply the motivations, thoughts and reasoning behind it.

In the paragraph of the results related to Study 2, I will analyze very carefully the answers of the recruiters, looking for relevant insights that could add validity and value to the results obtained from the experiment.

As regards the analysis of the answers, I will not call the recruiters by their name, but simply by naming Recruiter 1, Recruiter 2, and so on…

Then, I will try to summarize the key points, underline common thoughts and highlight relevant quotes.

In conclusion, I firmly believe that this further analysis carried out by me will reveal interesting findings for my research, which will complement those of the experiment.

3.2 Results

3.2.1 Study 1

After having cleaned the dataset, excluding previews, distortions and people who did not finish the survey, I conducted the analysis of descriptive statistics in to describe the sample, as it is possible to see in the previous paragraph.
Then, even though the scale used to measure "hireability" was a pre-validated scale found in the literature of human resource management, and there is no need to prove the scale because there is an assumption that it is valid, I still decided to carry out myself a Factor Analysis to confirm the validity of the scale.
Indeed, the Factor Analysis was carried out to reduce a large number of variables to a more manageable set. The only scale to be subjected to factor analysis is the one measuring candidate selection:

Figure 12. KMO and Bartlett's test

Analisi fattoriale

Test di KMO e Bartlett		
Misura di Kaiser-Meyer-Olkin di adeguatezza del campionamento.		,693
Test della sfericità di Bartlett	Appross. Chi-quadrato	474,531
	gl	3
	Sign.	<,001

1. KMO test. This indicator measures the adequacy of the sample, meaning that it should be at least equal to 0.6 in order to affirm that the sample is sufficient and there are enough participants. In this case, as it is possible to see from Figure 12, the KMO presents a value of 0.693, thus the sample is sufficient. This analysis represents the ratio between the square of the correlation between the variables and the square of the partial correlation between the same variables. The total KMO score undergoes variations from zero to one, the closer it will be to one, the better the reliability will be.
2. Bartlett's test. The test works with the metrics of correlation, it checks if there is a redundancy between variables that can be summarized with some factors and the p-value should be lower than $\alpha=0.05$ in order to test that the correlation matrix is not diagonal. Indeed, the test has as a hypothesis the assumption of sphericity which must be rejected. In this case, p-value<0.001, thus significance correlations exist and the null hypothesis is verified.

Figure 13. Communality

Comunalità		
	Iniziale	Estrazione
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Farei un colloquio al candidato per il lavoro	1,000	,704
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Assumerei personalmente il candidato per il lavoro	1,000	,870
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Il candidato sarebbe assunto per il lavoro	1,000	,820

Metodo di estrazione: Analisi dei componenti principali

Communalities represent variance shared between observed variables and factor: each item should be equal at least to 0.5. If one item has a low communality means that there is a problem on the statement, thus it should be deleted because is not belonging to any factor. In this case, all communalities are bigger than 0.50, so all the items belong to one factor and we don't have to delete any of them.

Figure 14. Total explained variance

Varianza totale spiegata						
	Autovalori iniziali			Caricamenti somme dei quadrati di estrazione		
Componente	Totale	% di varianza	% cumulativa	Totale	% di varianza	% cumulativa
1	2,394	79,794	79,794	2,394	79,794	79,794
2	,428	14,257	94,052			
3	,178	5,948	100,000			

Metodo di estrazione: Analisi dei componenti principali.

This table allows us to analyze the percentage of total variance explained by the extracted factor. In this case, the item that has an eigenvalue > 1 is just one and the factor explains about 80% of the variance derived from the 3 items. According to the output, we can take one factor and we can therefore be satisfied with at least the 80% of the cumulative variance explained. From Figure 15, it is possible to see the scree plot, which shows exactly the same thing.

Figure 15. Scree plot

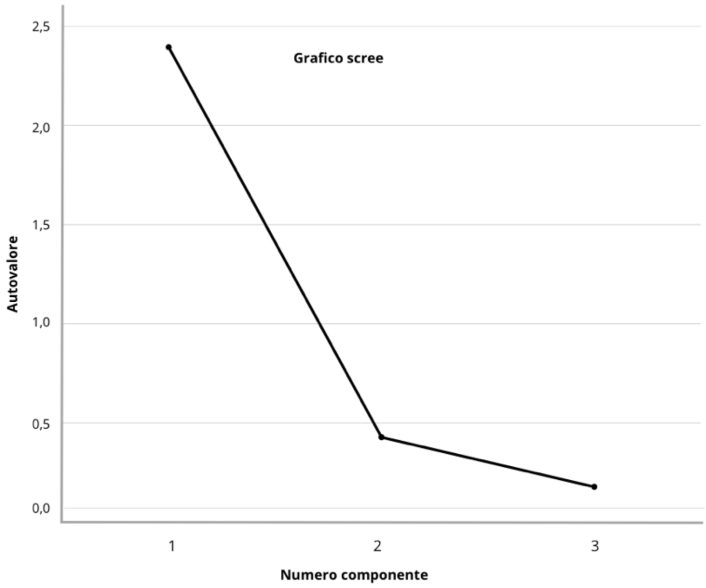


Figure 16. Matrix component

Matrice dei componenti [a]	
	Componente 1
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Farei un colloquio al candidato per il lavoro	,839
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Assumerei personalmente il candidato per il lavoro	,933
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Il candidato sarebbe assunto per il lavoro	,905

Metodo di estrazione: Analisi dei componenti principali.
a. 1 componenti estratti.

Moreover, as it is possible to observe from Figure 16, in the matrix component all the items registered a loading values higher than 0.30, so it was possible to confirm all the three items, without erasing them.
Thus I validated the scale related to the dependent variable and I continued with the analysis.

A reliability test was undertaken for the measuring scale that was assessed in this study to ensure that it was reliable. The Cronbach Alpha (α) of the scale was determined with respect to the numerous things that it is made up of to carry out this verification. Cronbach's Alpha is a reliability measure with values ranging from 0 to 1, such that $0 < \alpha < 1$.
The requirement of Cronbach's Alpha index $\alpha > 0.7$ is necessary for a measuring scale to be defined as reliable, where 0.7 denotes the minimum acceptable threshold, such as to assure sufficient internal consistency and test adequacy (Malhotra et al., 2012).
The higher the Cronbach's Alpha score, the more reliable the measurement scale is. The multiple components must all have values of at least 0.7 in this situation, as the aforementioned criteria define such a level of consistency and reliability that the measuring scales utilized can be deemed truly consistent and reliable.
Proceeding with the analysis of the reliability of the measurement scale "Hireability Scale" by Rudman and Glick (2001) it is possible to observe the items of the scale and the value of Cronbach's Alpha index from the reliability statistics reported in the following table:

Figure 17. Hireability scale's reliability

Statistiche di affidabilità	
Alpha di Cronbach	N.di elementi
,870	3

The alpha coefficient was used to assess the correlation between items whose object was the assessment of the same concept. This coefficient has a range of values from 0 to 1, and Cronbach's alpha must be between 0.70 and 1 for the scale to be considered reliable. Following the examination, this scale is judged to have a good level of reliability, since the alpha value is 0,870.
By analyzing the following table in Figure 18, related to the Hireability scale's reliability statistics, it is possible to verify that the higher value of Cronbach's Alpha index is $\alpha = 0{,}897$, if the first item was to be deleted.
In any case, the difference with the original Cronbach Alpha was not much and eliminating this item would not have improved that much, so I thought it appropriate to keep all three items, as the initial Cronbach alpha already had a good-excellent level of reliability.

Figure 18. Hireability scale's total elements statistics

	Statistiche elemento-totale			
	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Alpha di Cronbach se viene eliminato l'elemento
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Farei un colloquio al candidato per il lavoro	6,86	2,917	,667	,897
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Assumerei personalmente il candidato per il lavoro	7,47	2,819	,829	,749
Indica su una scala da 1 (per niente probabile) a 5 (estremamente probabile) in che misura sono probabili le seguenti affermazioni - Il candidato sarebbe assunto per il lavoro	7,50	2,820	,767	,802

Therefore, the measurement scale has a good degree of reliability that can be considered effectively useful for the research. It is, therefore, possible to affirm that the measurement scale for the candidate selection is reliable to be used for statistical analysis purposes.

In conclusion, after proving its validity and detecting the reliability through Cronbach's alpha index, the "Hireability Scale" by Rudman and Glick (2001) was selected for the measurement.

In order to test the relationship between the variable X (race) and the variable Y (candidate selection), an independent samples T-Test was carried out to verify whether or not there is a statistically significant difference in the mean of the perception variable within the scenarios of the race variable.

In the statistical test, the validity of a statistical hypothesis defined precisely as null hypothesis and indicated with H0 is verified. The null hypothesis, in general, represents the opposite of the scientific hypothesis that we want to prove. This hypothesis was formulated based on specific parameters of a variable. Therefore, the hypothesis assumed is reported below:

H0: The means are the same within the two experimental groups

H1: Means differ significantly in population, I will therefore reject the null hypothesis of equal averages, concluding that the averages are significantly different if the p-value is less than $\alpha = 0.025$.

The Levene's test was used to check the null hypothesis of equal variances (homoskedasticity) among the two subgroups before assessing the T-Test with independent samples. Indeed, the p-value related to F-test was equal to 0.059, that is higher than 0.05, thus the Leven's test has been found statistically significant and it was possible to continue with the independent sample T-test.

Figure 19. Group statistics

Statistiche gruppo					
	WHITE	N	Media	Deviazione std.	Errore standard della media
Hireability_scale_MEAN	1,00	140	3,5738	,90125	,07617
	,00	142	3,7042	,72455	,06080

By testing the hypotheses, we aim to understand whether the candidate selection in the condition of a white candidate (1), is greater than the one in the condition of a black candidate (0).

It is evident that the average of the candidate selection is slightly larger in the Scenario with black candidate than in the Scenario with a white candidate. However, in order to determine whether or not this difference is statistically significant, the T-test is necessary.

Then, the procedure continues with the analysis of the T-test aimed at investigating whether the difference between the two means is statistically significant.

Figure 20. Independent T-test

Test campioni indipendenti												
			Test di Levene per l'eguaglianza delle varianze		Test t per l'eguaglianza delle medie							
							Significatività				Intervallo di confidenza della differenza di 95%	
			F	Sign.	t	gl	P unilaterale	P bilaterale	Differenza della media	Differenza errore std.	Inferiore	Superiore
Hireability_scale_MEAN	Varianze uguali presunte		3,593	,059	-1,340	280	,091	,181	-1,3042	,09731	-,32197	,06114
	Varianze uguali non presunte				-1,338	266,079	,091	,182	-1,3042	,09746	-,32231	,06148

As a result, it is possible to reach two conclusions:

The Levene's test accepts the null hypothesis of equal variances (p=0,059, thus p>0,05), hence the independent T-test with equal variances will be used.

The null hypothesis of equal means is not rejected by the T-test (p = 0,181, indeed p-value is higher than α= 0.025), thus we cannot accept our research hypothesis: this means that the difference is true only for the sample mean but not for the population.

Indeed, the means of the two different groups are respectively 3.57 (white candidate) and 3.70 (black candidate), so the mathematical difference which is only 0.13 is too small in order to get a statistically difference between the two means.

The third phase of the analysis is aimed at understanding the impact that the moderator, “Work experience” generates in relation to the “Candidate selection” determined by the presence of white candidate or black candidate.
Firstly, the analysis began performing a Two-way ANOVA.

Figure 21. Descriptive statistics

Statistiche descrittive				
Variabile dipendente: Hireability_scale_MEAN				
WHITE	WORK_EXPERIENCE	Medio	Deviazione std.	N
,00	,00	3,6571	,66418	70
	1,00	3,7500	,78073	72
	Totale	3,7042	,72455	142
1,00	,00	3,3284	,99238	67
	1,00	3,7991	,74689	73
	Totale	3,5738	,90125	140
Totale	,00	3,4964	,85379	137
	1,00	3,7747	,76162	145
	Totale	3,6395	,81822	282

As it is possible to see from the Descriptive Statistics table in Figure 21, the highest mean is related to the scenario coded as 1.1 (White candidate, 5 years of work experience), with a value of 3,7991. While the lowest mean is represented by the stimulus coded as 1.0 (White candidate, no work experience), with a value of 3,3284.
The subjects exposed to the black candidate with 5 years of work experience showed a mean value of 3,7500, and finally the respondents who saw the visual condition represented by a black candidate with no work experience showed a 3,6571 as a mean.
In addition, a candidate with 5 years of experience showed a mean of 3,7747. On the other hand, it is possible to see a lower mean, equal to 3,4964 for a candidate with no work experience.
Independently from the work experience, basing only on the type of race, a black candidate reached a mean of 3,7042. Instead, a white candidate had in total a mean equal to 3,5738.
In order to see if there were means statistically significant, it is necessary to see the indicators of significance in the between-subjects effect test shown in Figure 23.

Figure 22. Levene's test of error variance equality

Test di Levene di eguaglianza delle varianze dell'errore[a,b]					
		Statistica di Levene	gl1	gl2	Sig.
Hireability_scale_MEAN	Basato sulla media	2,572	3	278	,054
	Basato sulla mediana	1,816	3	278	,144
	Basato sulla mediana e con il grado di libertà adattato	1,816	3	244,228	,145
	Basato sulla media ritagliata	2,458	3	278	,063

Verifica l'ipotesi nulla che la varianza dell'errore della variabile dipendente sia uguale tra i gruppi.
a. Variabile dipendente: Hireability_scale_MEAN
b. Disegno: Intercetta + WHITE + WORK_EXPERIENCE + WHITE * WORK_EXPERIENCE

Starting by studying the Levene's Test, it does not reject the null hypothesis of equal variances (p =0,054, thus p> 0,05). Thus, it is successful and I carried on with the analysis.

Figure 23. Between-subjects effect test

Test di effetti tra soggetti

Variabile dipendente: Hireability_scale_MEAN

Origine	Somma dei quadrati di tipo III	df	Media quadratica	F	Sig.
Modello corretto	9,246[a]	3	3,082	4,790	,003
Intercetta	3719,356	1	3719,356	5780,351	<,001
WHITE	1,377	1	1,377	2,141	,145
WORK_EXPERIENCE	5,592	1	5,592	8,691	,003
WHITE* WORK_EXPERIENCE	2,514	1	2,514	3,907	,049
Errore	178,879	278	,643		
Totale	3923,444	282			
Totale corretto	188,125	281			

a. R-quadrato = ,049 (R-quadrato adattato = ,039)

From this table the analysis gets many information, the most important for the research are:

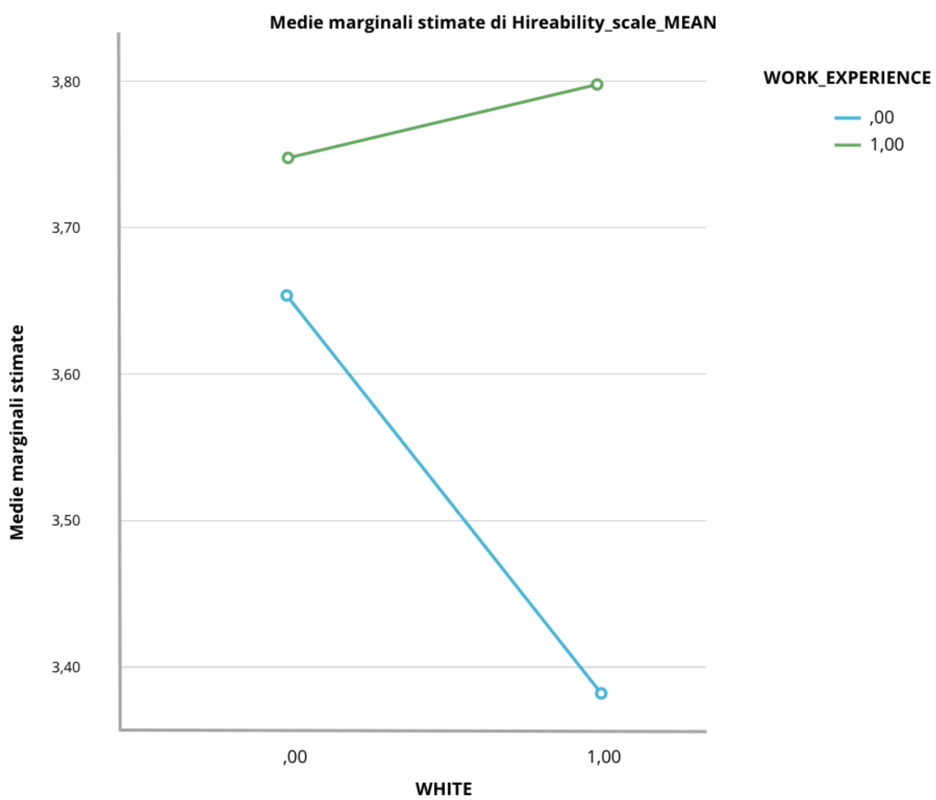
1. Model Fit: F(3,282)=4,790, p=0,003, thus p-value < α 0.05, the model is significant, there is at least one mean statistically different than the others;
2. Main effect Race (IV): F(1,282) = 2,141, p=0,145, thus p-value > α 0.05, race type of candidate (white or black) does not influence the average of candidate selection. Thus, H1 is rejected.
3. Main effect Work experience (MOD): F(1,282) = 8,691, p=0.003, thus p-value < α 0.05, type of work experience (5 years vs. no work experience) influences the average of candidate selection. H2 is not rejected.
4. Interaction effect: F(1,282) = 3,907, p= 0,049, thus p-value < α 0.05, type of work experience influence the relationship between race and candidate selection. Therefore, even H3 is not rejected.

Since the interaction effect of WHITE x WORK_EXPERIENCE on Hireability_scale_MEAN was statistically significant, there is need to interpret the interaction plot below to understand the direction of interaction.

As it is possible to see from Figure 24, we are in the Case 3: there is disordinal interaction with no-crossover effect between race and work experience on candidate selection.

Difference in candidate selection between black and white candidate changes with work experience. The difference is highest for white candidate. There is almost no difference in candidate selection for black candidate. With no work experience being the same, the difference in race between the two candidates affects and is more evident. While introducing the five years of experience, the moderator effect works, because there is no longer so much difference between the black and white candidate.

Figure 24. Interaction plot



3.2.2 Study 2

Understanding the ethical issues and biases that can arise in the context of recruitment is crucial in today's increasingly automated and AI-driven world. To highlight this important topic, a comprehensive research study was conducted, which involved not only gathering the perspectives of ordinary individuals through an experiment but also obtaining insights from professional in the field through a series of interviews, in order to gain valuable insights into the hiring process and better understand their perspectives.

This paragraph aims to present and analyze the results obtained from the interviews with recruiters, highlighting their perspectives on the ethical concerns and biases in the recruitment process. The interviews aimed to gather information about their decision-making processes, preferences, and the qualities they look for in candidates. By comparing the findings from both the experiment and the interviews, it is possible to gain a comprehensive understanding of the multifaceted challenges and potential solutions for addressing ethical issues and biases in recruitment driven by artificial intelligence.

Therefore, through the analysis of the results of these interviews, a deeper understanding of the factors that influence hiring decisions and the evolving dynamics of the job market can be achieved. This paragraph will present and comment on the findings of these interviews, shedding light on the key takeaways and providing a comprehensive overview of the recruiter's perspective.

Starting with the first question, it was asked how they would describe the process of selecting candidates in their respective companies. Based on the answers provided by the recruiters, several important aspects can be highlighted: transparency and organization, time efficiency, multi-step process, feedback and communication, focus on fit and alignment. From these responses, it becomes evident that recruiters prioritize transparency, efficiency, and fit in their selection processes. The emphasis on multi-step assessments and feedback demonstrates a commitment to thorough evaluation and maintaining positive candidate experiences. One notable quote from Recruiter 5 reinforces the commitment to transparency: *"It is a very fast, focused, and always very transparent process because we try to provide feedback to people in every case."* Another interesting quote comes from Recruiter 9, describing the emphasis on accurate candidate selection: *"Accurate, targeted, selective, and quite fast. [...] We conduct an initial screening of the received CVs [...] once we select the CVs that are more aligned, there is an initial exploratory interview. There is a call first, in which we ask a series of fairly general questions to understand the candidate's background, where they are in their academic journey, etc... If the answers are in line, we proceed with the first exploratory interview, either with me or my colleagues."* This quote highlights the meticulous and focused approach to the selection process, wherein importance is given to both the analysis of CVs and an initial exploratory interview. This targeted approach emphasizes the commitment to selecting candidates who are suitable for the company's needs and aligned with the requirements of the position.

Moving on to the second question "What are the main criteria you use to evaluate candidates?", it was observed that each recruiter has their own set of criteria that they consider when evaluating candidates. These differences reflect the unique priorities and requirements of each recruiter based on their company's culture, position specifications, and industry. However, some common themes emerged. Firstly, all recruiters considered CV and educational background to some extent, in particular Recruiter 2, Recruiter 4 and Recruiter 9, evaluated the candidate's academic path, that includes the type of degree, courses studied, and grades achieved. Secondly, it was placed a strong emphasis on soft skills, which were mentioned by multiple recruiters, such as Recruiter 1, Recruiter 3, Recruiter 4, Recruiter 6, and Recruiter 8, who emphasized qualities like communication skills, curiosity, clarity of intent, storytelling abilities, and the ability to work well with others. The importance of soft skills is more prominent when it comes to non-technical roles and it was shown their significance in candidate evaluation. There is overlap with 2 recruiters who emphasized both soft skills and hard skills. Lastly, only 1 recruiter specifically prioritized hard skills, especially in technical roles. Several recruiters also mentioned the importance of candidate's work experience, whether in a related field or not, as an important factor. Recruiter 4 valued the candidate's interests and hobbies, as he believed they provide insights into their personality. However, this aspect was not mentioned by other recruiters.

Continuing with the third question of the interview, about the most appreciated competencies of the candidates, the qualitative findings from the responses of different recruiters highlighted several key competences that were commonly sought in candidates. While there are variations based on industry and specific roles, some recurring skills emerged. Technical expertise and industry-specific knowledge were valued by recruiters, as evidenced by the emphasis on competences such as accounting, project management, management engineering, and economics. Proficiency in English was also considered important in many cases. Soft skills played a significant role, with communication, adaptability, problem-solving, teamwork, and relational abilities being mentioned by recruiters. The willingness to learn, curiosity, and the ability to ask questions were also seen as positive attributes. Additionally, consistency between a candidate's educational background and their career aspirations was highlighted as an important factor for some recruiters. Overall, a well-rounded candidate with a combination of technical skills, industry knowledge, and strong soft skills standed out in the eyes of recruiters.

The fourth question concerned how they ensured that candidates were assessed objectively and impartially. Several recruiters mentioned the importance of training and raising awareness about biases in the evaluation process. For example, Recruiter 5 said: *"We train our people. I am trained and have experience, and we also train additional interviewers to ensure they are aware of the biases that exist in the process."*. Even Recruiter 8 affirmed: *"We conduct courses and sharing sessions regarding what our company looks for in candidates..."*. Recruiters differed in their approaches to evaluation criteria, with some focusing on specific qualifications and experience while others emphasized personal fit and emotional aspects. Moreover, most of the recruiters mentioned the use of multiple interviews to ensure a more comprehensive evaluation and normalize feedback. For instance, Recruiter 4 explained: *"We always try to conduct two interviews to normalize feedback as much as possible"*. In addition, Recruiter 9 added: *"There are multiple steps conducted by different people, and it is not biased toward any specific group or age"*. The role of the hiring manager varied, with some recruiters giving more decision-making authority to the managers for the final selection. In particular, Recruiter 1 mentioned: *"The final decision is something that belongs to the managers... It is important for me to hire people who fit well within my team"*. Furthermore, Recruiter 10 specified: *"We review all resumes with the line manager, who, having more expertise in the industry, can better evaluate the skills a candidate should possess"*. Additionally, recruiters expressed the importance of reducing biases and maintaining an open mind during the evaluation process, as shown in the following examples: Recruiter 4 *"We try to reduce prejudices and conduct evaluations as objectively as possible"*; Recruiter 8 *"We strive to align as much as possible. There is some subjectivity, but it is essential"*. These were some of the commonalities and differences found among the recruiters regarding the evaluation process. Overall, they mentioned various approaches to ensure objective and impartial evaluations. These included providing training, conducting multiple interviews, normalizing feedback, minimizing biases, and involving different perspectives in the assessment process. While subjectivity exists to some extent, efforts are made to make evaluations as fair and objective as possible.

The fifth question of the interview was "How do you evaluate candidates to ensure that there is no discrimination on the basis of race, gender, sexual orientation or other protected characteristics?". Based on the answers, all recruiters emphasized the importance of evaluating candidates based on their skills, qualifications, and experience. None of the recruiters mentioned considering factors such as race, gender, sexual orientation, or other protected characteristics during the initial screening process, stressing the importance of equal treatment in the selection process. Most recruiters considered reviewing the candidates' resumes or CVs as the primary basis for evaluation. Therefore, several recruiters mentioned the use of tests or technical interviews to assess candidates' technical competencies. In addition, some recruiters pointed out the inclusion of diverse candidates and actively seeking individuals from different backgrounds or nationalities. Finally, the recruiters emphasized evaluating candidates based on their fit for the role and the company's requirements, rather than personal characteristics. In particular, two recruiters mentioned specific language requirements for certain roles, such as the need for candidates to speak Italian or English fluently: Recruiter 3 commented that for a sales role, it is necessary for candidates to speak Italian since the role involves selling, and not all company contacts speak English; even Recruiter 10 mentioned the importance of candidates knowing both Italian and English because it is crucial for their field, as they need to interact with clients and the company's headquarters is in London. Moreover, Recruiter 8 mentioned considering the needs of women consultants with children and adapting policies accordingly. Recruiter 9 indicated potentially specifying age limits in job postings to prevent unsuitable candidates from applying. Lastly, Recruiter 10 focused on evaluating written aspects of resumes, particularly work experience and language skills.

Going on with the sixth question about how recruiters evaluated the experience and competence of candidates from different backgrounds, it resulted that the assessment was based on various factors, for example alignment with the role, demonstrated value, relevant skills, experience, and adaptability. While some recruiters preferred candidates with a more focused or linear career path, others valued the diversity and transferable skills gained from different backgrounds. Overall, all recruiters emphasized the importance of the candidate's background aligning with the specific job requirements and expectations. Some recruiters, such as those in the financial services industry (Recruiter 4), preferred candidates with a specific background (e.g., banking or finance) due to the industry's unique demands. With regards to the influence of education, some recruiters, particularly for junior positions, favored candidates with a quantitative or relevant educational background (Recruiter 8), while others placed less emphasis on formal education and focused more on experience (Recruiter 9). Indeed, experience was a key factor for assessing competence, particularly for senior positions. Candidates with relevant work experience were often preferred. Specifically, recruiters differed in their approach to evaluating experience concerning the level at which candidates entered the organization. Some considered experience for determining seniority, while others used it as a basis for entry into the selection process (Recruiter 5). Furthermore, recruiters such as Recruiter 6 prioritized soft skills, while others valued technical skills more, depending on the position requirements.

The seventh question concerned whether they were present and what were the policies of inclusion of diversity in the workplace. From the provided responses, it can be observed that several recruiters mentioned policies promoting diversity and inclusion, while others acknowledged the need to work on developing such policies. Overall, all recruiters acknowledged the importance of diversity and inclusion in the workplace and expressed a desire to listen to individual needs and accommodate diverse requirements. Moreover, several recruiters mentioned a focus on gender inclusion and the promotion of women in their organizations. In particular, some recruiters highlighted efforts to create a supportive environment for working parents, particularly mothers.

On one hand, Recruiter 1 emphasized the equal distribution of candidates and the value of internal talent, while also recognizing the need for external diversity; Recruiter 2 put attention to women and inclusion for all races and genders, with campaigns both internally and externally; Recruiter 3 mentioned having women in important managerial positions and a context where gender differences are not experienced; Recruiter 6 emphasized the high representation of women in their company and the focus on facilitating career paths for employees with responsibilities outside of work; Recruiter 9 mentioned specific policies for supporting new mothers and women, including providing full-time smart working opportunities. On the other hand, Recruiter 5 noted that their company did not have official inclusion policies but strived to respect individual diversities and needs; Recruiter 8 stated that their company did not have strong inclusion policies but was now working on addressing the issue, particularly related to parenting and diversity, affirming *"We never really had them [inclusion policies] because it wasn't really a theme, lately we have started addressing the issue... I can't give you a precise answer because we are working on it"*. Lastly, Recruiter 9 said *"Regarding diversity, I can tell you that we have people of different ethnicities and different sexual orientations, but we haven't done much to promote this, it just happened... We are really addressing the issue now"*. It's important to note that while some recruiters provided specific details about their company's inclusion policies, others mentioned ongoing efforts or the lack of formal policies.

Regarding the last part of the interview' questions, before it was asked "Are there people of different ethnicities within your company? If so, what percentage?" and then "How many people of different ethnicities propose for a position from you?". In analyzing the findings from the recruiters' responses regarding the presence of people from different ethnic backgrounds in their respective companies, there are both commonalities and differences. Several recruiters affirmed the existence of employees from diverse ethnicities within their organizations. However, the percentages varied significantly. Some recruiters provided approximate figures, ranging from 1% to around 30%, while others mentioned a lower presence or even the absence of employees from different ethnicities. Factors such as the international nature of the roles, language requirements, and the absence of foreign offices were cited as influencing these percentages. For example, Recruiter 4 stated that the presence of individuals from different ethnic backgrounds was limited, possibly around 1%, due to language and project requirements. Furthermore, Recruiter 5 believed that less than 10% of their employees represented diverse ethnicities, primarily due to a lack of international operations. Overall, it is evident that the level of ethnic diversity within the companies varied, reflecting the unique circumstances and dynamics of each organization.

The recruiters provided a range of perspectives regarding the number of applicants from different ethnicities. In general, recruiters indicated that there are candidates of different ethnicities who apply for positions. Some recruiters mentioned receiving a significant number of applications, while others reported a relatively low number. Recruiter 1 and Recruiter 2 expressed that they receive many applications from individuals of different ethnicities. On the other hand, Recruiter 4 and Recruiter 7 reported receiving only a few applicants. Recruiter 3 noted that there are candidates from diverse ethnicities, but the number of applicants of color is relatively lower. Recruiter 5 mentioned receiving fewer applications due to language requirements and the absence of overseas offices. Recruiter 8 highlighted that although there are many applicants from diverse backgrounds, language proficiency becomes a limiting factor. In contrast, Recruiter 10 expressed satisfaction with the high number of applicants from diverse ethnic backgrounds. Overall, the findings suggest that while some recruiters receive a substantial number of applications from candidates of different ethnicities, others receive fewer, leading to varying percentages of applicants from diverse ethnic backgrounds across different companies.

After the end of the prepared questions, recruiters were shown the four graphical stimuli of the experiment, asking their opinions and which candidate they would have preferred and why. Most of the recruiters emphasized that they focused on other factors such as experience, skills, motivation, attitude, and fit with the position and company culture. Some recruiters mentioned that they did not even have access to the candidates' photos during the evaluation process, further indicating that physical appearance or race were not determining factors. Recruiters also mentioned that their preferences varied depending on the specific requirements of the position. For entry-level positions, recruiters were more open to considering candidates without experience. However, for roles requiring a certain level of expertise, recruiters generally preferred candidates with relevant experience, regardless of their race.

It is important to note that recruiters acknowledged the need to consider individual qualities and soft skills during the selection process. They mentioned assessing factors like communication skills, openness, motivation, and the ability to relate to others when making their decisions. The recruiters also considered the importance of qualifications, such as education and language proficiency, as well as considering budgetary constraints for more experienced candidates.

The following quotes highlight the recruiters' emphasis on qualifications, experience, skills, and the specific requirements of the positions rather than race as a determining factor in their decision-making process:

1. *"The fact that someone is white or black makes absolutely no difference. Positions are set based on seniority and the specific experience needed to carry out the required tasks."* - Recruiter 1.
2. *"Honestly, I would not make an evaluation based on race. At the same level of qualifications, I would try to understand their motivation and other factors that make them suitable for the job."* - Recruiter 2.
3. *"I would focus on their education level and salary expectations. If it's a junior position, I would consider both candidates with no experience. For the candidates with 5 years of experience, the decision would depend on their salary expectations within our budget."* – Recruiter 3.
4. *"At the end of the day, it's about the person. Soft skills and how they relate to others matter. Given the same level of experience, we would assess their interpersonal skills and relational abilities."* - Recruiter 4.
5. *"Preferences are impossible to determine in such a scenario. It depends on the specific position and the required fit. Experience may be necessary or not, and the candidate's race does not play a role in our evaluation."* - Recruiter 5.
6. *"If I need someone with experience, I call both candidates with experience. I don't have a preference between them. It wouldn't change anything for me."* - Recruiter 6.
7. *"Experience alone is not influential. It depends on the position and whether experience is required or not. We don't base our evaluation on race either."*- Recruiter 7.
8. *"The color of their skin is not important. It's about their attitude, qualifications, and the position's requirements. Experience alone or race would not be deciding factors."* - Recruiter 8.
9. *"Race has no value. It's about their qualifications, experience, language skills, and availability. The rest, such as photos, names, gender, or age, could be obscured."* - Recruiter 9.
10. *"For a front desk agent, previous experience is necessary. However, for positions with 5 years of experience, both candidates are equally valid, and I would pass both of their resumes."* - Recruiter 10.

3.3 Discussion

The two studies conducted provided valuable insights into the hiring decision-making process from different perspectives. The first study was quantitative and experimental, involving a sample of ordinary people, with a age as close as possible to the recruiters' working age. The results of the experiment conducted on the preference of normal people towards a white or black candidate showed no significant difference between the two candidates. However, the findings revealed that a candidate with five years of work experience was preferred over a candidate with no work experience.

Moreover, with five years of work experience, no difference was observed between the white and black candidates, indicating that work experience might be a more critical factor in the hiring decision-making process than race. This suggests that organizations should prioritize work experience over race when considering job candidates: this is because work experience is a reliable predictor of job performance and competence. In addition, it indicates that organizations should focus more on the skills, expertise, and qualifications of candidates rather than their race.

The second study was qualitative and involved interviews with real recruiters. It provided insights into the recruitment process in organizations and the confirmation of the previous findings of the experiment in qualitative interviews with real recruiters further supports the validity of the research.

Based on the feedback provided by the recruiters, it can be concluded that the race of the candidates did not play a significant role in their preferences. Overall, the recruiters expressed a focus on selecting candidates based on their qualifications, experience, and fit with the role, rather than making decisions based on race or physical appearance.

Furthermore, it is evident that there was a preference for candidates with 5 years of experience over those with no work experience, given that the candidates possessed similar qualifications and skills. The recruiters emphasized that experience is a significant factor when evaluating candidates for certain positions.

However, it is important to note that this preference did not appear to be influenced by race, as all candidates were evaluated based on their qualifications and suitability for the job. The recruiters' focus on experience suggested that they valued practical knowledge and the ability to apply skills in a professional setting, which can be gained through previous work experience. Nevertheless, this did not imply that candidates without experience were completely disregarded, as other factors such as motivation, attitude, and potential may also be taken into consideration during the selection process.

The qualitative interviews also provided insight into the role of diversity and inclusion in the recruitment process: unfortunately, it was revealed that not all companies have policies of diversity and inclusion, even if nowadays these topics are becoming increasingly important in the workplace.

Furthermore, small companies often have only one recruiter to do the screening, and people of different ethnicities are present in small minorities in Italy. Recruiters mentioned that the availability of candidates from diverse ethnic backgrounds could vary based on the location of the company. For example, recruiters in larger cities like Milan reported a higher number of applicants from different ethnicities compared to recruiters in other regions. This suggests that geographic factors can play a role in the diversity of the candidate pool.

Additionally, most companies in Italy do not hire candidates that do not speak Italian, which can be a limiting factor in the recruitment process. Indeed, several recruiters mentioned that candidates are required to have a certain level of proficiency in the Italian language. This language requirement could pose a challenge for applicants who do not meet the criteria, particularly those who are seeking positions in companies without overseas offices or where Italian is the primary language of communication.

Moreover, recruiters in certain industries or companies mentioned that there is generally less demand for candidates of different ethnicities. They observed that the number of applicants from different ethnic backgrounds, particularly applicants of color, is relatively low. This could happen due to a combination of factors, including historical recruitment practices and industry dynamics.

In addition, some recruiters pointed out that their organizations or industries have a strong connection to Italian culture or have a specific image or identity tied to the industry. This could create a preference for candidates who align more closely with the cultural and industry norms, resulting in fewer opportunities for individuals from different ethnic backgrounds to be hired.

While not explicitly mentioned by the recruiters, implicit biases in the recruitment process can also contribute to the underrepresentation of individuals from diverse ethnic backgrounds. Unconscious biases may affect decision-making, leading to unintended discrimination in the selection process. It is important to note that these factors are based on the recruiters' perspectives and may not represent the complete picture. Hiring decisions are influenced by a range of complex factors, and organizations should strive for inclusive practices to ensure equal opportunities for candidates from all ethnic backgrounds.

Overall, these findings highlighted the importance of considering the recruiter's perspective when analyzing ethical issues and biases in AI-driven recruitment processes. Understanding how recruiters approach candidate selection can provide valuable insights for addressing biases and ensuring fair and ethical practices in the recruitment context.

The insights from the recruiters were useful in highlighting the potential limitations of the recruitment process and the possibility of biases and prejudices. Even though the recruiters claimed to be indifferent to the race of the candidate, implicit biases may still be present, and the focus on work experience may not entirely eliminate biases.

Overall, the two studies suggest that the recruitment process can be complex, and there are many factors to consider when making hiring decisions. It is crucial to continue to promote diversity and inclusion in the recruitment process, even in companies that do not have specific policies.

Moreover, organizations should be aware of the limitations of the recruitment process and take steps to minimize the potential for biases and prejudices.

However, the finding that recruiters were indifferent to the race of the candidate suggests that the focus on diversity and inclusion has the potential to reduce bias in the recruitment process.

In conclusion, the experiment's findings suggest that work experience is a critical factor in the hiring decision-making process. The confirmation by real recruiters further supports this notion and suggests that organizations should prioritize work experience over race when considering job candidates.
Furthermore, the qualitative interviews provided insight into the role of diversity and inclusion in the recruitment process, indicating that it is becoming increasingly important.
Lastly, implicit biases can still play a role in the hiring process, making it crucial to adopt measures that promote fairness and inclusivity in recruitment practices.

3.3.1 Managerial implications

This work contributes to the research already existing in the literature under various aspects and can be useful for managerial purposes. Among the socio-ethical implications, the problem of bias emerges, which commit AI systems where they are replicated and amplified, perpetrating injustices and inequalities and further affecting categories already subject to prejudices and discriminations.
Therefore, the aim of this research is to investigate this phenomenon, closely connected to the process of recruitment, in order to outline how and in what forms could AI enter personnel offices. In human resources, and more particularly during recruitment processes, it is fundamental to ensure fairness and understanding how to avoid discrimination is a key factor in solving this problem.
Previous research has highlighted the biases within AI algorithms that reinforce stereotypes and potentially perpetuate inequities and discrimination against candidates. Biases in AI manifest themselves during the algorithm's development, the training of datasets, or through AI-generated decision-making.
This research also highlights these issues in order to better understand the negative effect of some biases that affect the judgment of the AI in the context of recruitment, for example for issues related to racism. Furthermore, this study contributes to the literature by better understanding why AI bias happens and why it's critical to regulate AI.
The ultimate goal is to provide tips for organizations that want to create unbiased AI and make recommendations for greater accountability within the public, private and nonprofit sectors, offering examples of positive applications of AI in challenging stereotypes. The potential for AI to exacerbate bias and discrimination highlights the need for organizations to take steps to ensure that their recruitment processes are transparent, accountable, and unbiased. This includes regular auditing and testing of algorithms to detect and address potential bias, as well as ensuring that hiring managers are trained in recognizing and mitigating implicit bias.
Moreover, a possible contribution of this work is to propose a sort of guidelines to lead managers and help companies to combine ethics and AI in order to solve their complex challenges: offering real suggestions can be helpful for all companies that intend to develop fair and inclusive AI solutions, by enhancing the ethics of algorithms.

3.3.2 Limitations and future research

One potential limitation of this study is that it only considers two races (black and white), while ignoring other ethnicities that may also be affected by AI racial bias. This could limit the generalizability of the findings to other contexts and populations. Another limitation is that the study assumes that AI racial bias is the only factor influencing candidate selection outcomes, while not considering other factors such as the qualifications of the candidates. Therefore, the study may overestimate the impact of AI racial bias on selection outcomes.
Another potential limitation of this study is that the definition of "AI racial bias" may not be well-defined or well-understood. Bias can manifest in various ways, such as through the training data used to develop the AI system or through the design of the algorithm itself. Indeed, without a clear definition of what constitutes AI racial bias, it may be difficult to accurately measure its impact on candidate selection. Furthermore, it is important to note that the data was collected exclusively in Italy since the questionnaire was created in Italian and only distributed among Italian citizens. By the way, it is likely to be not representative of the whole country. In addition, one limitation of my research is that, since I could not test bias in AI algorithms, I tried to prove whether there was a racial bias in recruiters, and therefore in people.

The use of a convenience sample may introduce selection bias, as participants were selected based on their easy accessibility, potentially resulting in a non-representative sample that may not accurately reflect the larger population. To improve the study's generalizability, future research should consider employing a stratified sampling technique, where participants are selected proportionately from different strata of the population. This would enhance the representativeness of the sample and allow for more precise statistical analysis and inferences about the entire population.

The last limitation is always related to the sample for the quantitative study, since it was pretended that the respondents were real recruiters but, in reality, they were not, since they did not cover the roles of expert people in the field. However, this limitation was filled through the qualitative study, by interviewing real recruiters and, therefore, using a mixed-method approach, both quantitative and qualitative.

In terms of future research, it would be valuable to investigate the impact of AI racial bias on other stages of the recruitment process, such as initial interviews or technical assessments.

Moreover, exploring the impact of other types of bias, such as gender bias or age bias, on candidate selection could provide a more comprehensive understanding of the effects of bias in recruitment. In particular, I highlight the need to explore how different types of AI systems (e.g. machine learning, rule-based systems) may lead to different types or degrees of racial bias.

Additionally, investigating potential solutions or interventions to mitigate AI racial bias in recruitment, such as algorithmic fairness or diversity training for recruiters, could be an important area of future research.

Finally, future research should ideally look at the impact of AI racial bias on other areas beyond candidate selection, such as performance evaluations or promotions, in order to provide a more comprehensive understanding of the effects of AI bias in the workplace.

CONCLUSION

In conclusion, this thesis has shed light on the dark side of AI, particularly in the context of recruitment, by exploring ethical issues and biases associated with its use. Through a comprehensive examination of AI's history, definition, and potential, as well as its associated benefits, opportunities, threats, and risks, this research has highlighted the complexities and challenges inherent in deploying AI systems. Moreover, it has emphasized the critical need to integrate ethics into the development and deployment of AI technologies.

The exploration of the dark side of artificial intelligence has revealed the potential for biases and discrimination to be perpetuated through AI-driven recruitment processes. Case studies, such as COMPAS and Amazon, have underscored the dangers of relying solely on AI algorithms without proper scrutiny and oversight. They have demonstrated how biases present in training data can be amplified by AI systems, leading to discriminatory outcomes in candidate screening and decision-making. These findings necessitate urgent attention to address bias and ensure fairness in AI-driven recruitment systems.

The literature review has further established the significance of ethical considerations in AI applications, exploring the intersection of AI and consumer patterns, algorithmic bias, and the future of marketing. It has become evident that as AI becomes increasingly integrated into our daily lives, its ethical implications must be carefully examined and proactively addressed. The potential benefits of AI, such as efficiency and productivity gains, must be balanced with the responsibility to mitigate biases, preserve privacy, and safeguard against discrimination. The methodology employed in this thesis, comprising of two comprehensive studies, has provided valuable insights into the prevalence and implications of AI bias in recruitment. The results have highlighted the need for organizations to critically evaluate and audit their AI algorithms, ensuring that they do not reinforce existing biases or discriminate against certain groups. The managerial implications derived from these findings emphasize the importance of responsible AI deployment, emphasizing transparency, accountability, and fairness throughout the recruitment process.

However, it is essential to acknowledge the limitations of this research. The evolving nature of AI and its applications means that ethical considerations and biases may continue to manifest in new and unforeseen ways. Therefore, further research and collaboration between academia, industry, and policymakers are required to keep pace with the rapid advancements in AI technology and to develop robust frameworks and guidelines that promote responsible AI use.

Overall, this thesis highlights the need for integration of ethics into AI development and deployment, especially in the context of recruitment. It invites organizations to prioritize fairness, transparency, and accountability in their AI-driven recruitment practices. As AI continues to reshape various sectors, including human resources and recruitment, it is imperative that stakeholders work together to ensure that AI technologies are leveraged in a manner that respects individual rights, minimizes biases, and upholds the principles of fairness and equality. By doing so, we can harness the full potential of AI while mitigating its dark side and fostering a future where technology serves as a force for good in our society.

BIBLIOGRAPHY

Acas, (2020). Artificial intelligence (AI) in recruitment. Retrieved from https://www.acas.org.uk/artificial-intelligence-ai-recruitment.

Accademia Politica, (2022). Machine learning e intelligenza artificiale: il futuro è oggi. *Il Sole 24 Ore*.

Acemoglu D., (2021). Harms of AI. *National Bureau of Economic Research*. Cambridge.

Agarwal A., Agarwal H., & Agarwal N., (2022). Fairness Score and process standardization: framework for fairness certification in artificial intelligence systems. *AI and Ethics*, 1-13.

Agrawal A., Gans J. S., & Goldfarb A., (2018). Prediction machines: The simple economics of artificial intelligence. *Harvard Business School Press*.

Aguirre E., Mahr D., Grewal D., De Ruyter K., & Wetzels M., (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, 91(1), 34-49.

AI Hleg, (2019). Una definizione di IA: principali capacità e discipline. *European Commission*, Bruxelles.

Akter S., Dwivedi Y. K., Sajib S., Biswas K., Bandara R. J., & Michael K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, 144, 201-216.

Alessandrini, G. (2019). Lavorare nelle risorse umane: Competenze e formazione 4.0. (S. Severino) Roma, Italia. *Armando Editore*.

Ammanath B., & Blackman R., (2021). Everyone in Your Organization Needs to Understand AI Ethics. Business Ethics, *Harvard Business Review*.

André Q., Carmon Z., Wertenbroch K., Crum A., Frank D., Goldstein W., et al., (2018). Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs and Solutions*, 5(1-2), 28-37.

Andrews T., (2016). Meet the robot monk spreading the teachings of Buddhism around China. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/morning-mix/wp/2016/04/27/meet-the-robot-monk-spreading-the-teachings-of-buddhism-around-china/?utm_term=.fed52d90bff3.

Angwin et al., (2016). Machine Bias. Ethics of Data and Analytics. *Auerbach Publications*.

Aristodemou L., & Tietze F., (2018). The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information*, 55, 37-51.

Asilomar AI Principles, (2017). Future of Life Institute.

Asimov I., (1942). Runaround. *Astounding Science Fiction*. Street & Smith. New York.

Awalegaonkar K., Berkey R., Douglass G., Reilly A., (2019). Report research AI: built to scale. *Accenture*. Retrieved from https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments

Baldegger R., Caon M., & Sadiku K., (2020). Correlation between Entrepreneurial Orientation and Implementation of AI in Human Resource Management (HRM). *Technology Innovation Management Review*, 10(4), 72-79.

Barfield W., (2018). Liability for Autonomous and Artificially Intelligent Robots. Paladyn, *Journal of Behavioral Robotics*, 9 (1), 193-203.

Basaldella M., (2018). L'etica nell'intelligenza artificiale per evitare i pregiudizi. *TED Talks*. Retrieved from https://www.youtube.com/watch?v=nBZXQgXZN8A&t=669s

Basile F., (2019). "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine", *DPU*.

Batra G., Queirolo A., & Santhanam N., (2018). Artificial intelligence: the time to act is now. *Mc Kinsey*.

Belenguer L., (2022). AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI Ethics*, 2, 771-787.

Benanti P., (2018). Le macchine sapienti. Intelligenze artificiali e decisioni umane. *Marietti*. Torino.

Benanti P., (2018). Oracoli. Tra algoretica e algocrazia. *Luca Sossella editore*. Rome.

Bertrand M., & Mullainathan S., (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991-1013.

Bhardwaj G., Singh S., & Kumar V., (2020). An Empirical Study Of Artificial Intelligence and its Impact on Human Resource Functions. International Conference on Computation, Automation and Knowledge Management, 47-51, *IEEE*.

Bhardwaj R., (2019). How AI is Revolutionizing the Human Resource Functions. Retrieved from https://www.entrepreneur.com/article/325715

Bhardwaj, S. (2013). Technology, and the up-skilling or deskilling conundrum. *WMU Journal of Maritime Affairs*, 245-253. Retrieved from https://doi.org/10.1007/s13437-013-0045-6

Bloomberg, (2022). Artificial Intelligence Market USD 1,581.70 Billion By 2030, Growing At A CAGR of 38.0%. *Valuates Reports*. Retrieved from https://www.bloomberg.com/press-releases/2022-06-13/artificial-intelligence-market-usd-1-581-70-billion-by-2030-growing-at-a-cagr-of-38-0-valuates-reports

Bloomberg, (2022). Social Media Can No Longer Hide Its Problems in a Black Box. Retrieved from https://www.bloomberg.com/opinion/articles/2022-08-01/facebook-youtube-s-problems-can-t-be-hidden-in-ai-s-black-box

Boyd R., & Holton R. J., (2018). Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation? *Journal of Sociology*, 54(3), 331-345.

Brebant H., (2020). How to Use Ethics to Build Trust in Artificial Intelligence. Retrieved from https://www.aspeninstitute.org/blog-posts/how-to-use-ethics-to-build-trust-in-artificial-intelligence/

Brougher J. (2019). The dark side of AI in hiring. *Harvard Business Review*. Retrieved from https://hbr.org/2019/09/the-dark-side-of-ai-in-hiring

Buehler K., Dooley R., Grennan L., & Singla A., (2021). Getting to know-and manage-your biggest AI risks. *Mc Kinsey*.

Buolamwini J., & Gebru T., (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on Fairness, Accountability and Transparency, *Proceedings of Machine Learning Research*, 81, 77-91.

Cabitza F., (2022). Deus in machina? L'uso umano delle nuove macchine, tra dipendenza e responsabilità. *Bompiani*. Florence-Milan.

Cai L., Zhu Y., (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*.

Campos J., (2020). Ethics in Artificial Intelligence. *WBPRO*. Retrieved from https://www.wbpro.com/ethics-in-artificial-intelligence/

Carey J.W., (2008). Communication as Culture. *Essays on Media and Society*. Routledge. New York.

Carmiel Y., (2023). Artificial intelligence: 3 trends to watch in 2023. *The Enterprisers Project*.

Carpenter J., (2015). IBM's Virginia Rometty tells NU grads: Technology will enhance us. Retrieved from https://www.chicagotribune.com/bluesky/originals/ct-northwestern-virginia-rometty-ibm-bsi-20150619-story.html.

Carr N., (2011). Internet ci rende stupidi? Come la rete sta cambiando il nostro cervello. *Raffaello Cortina Editore*. Milan.

Castelo N., (2019). Blurring the line between human and machine: Marketing artificial intelligence (doctoral dissertation), *Columbia University Academic Commons*. Retrieved from https://doi.org/ 10.7916/d8-k7vk-0s40.

Castelo N., & Ward A., (2016). Political affiliation moderates attitudes towards artificial intelligence. In P. Moreau & S. Puntoni (Eds.), NA - advances in consumer research, 723-723. Duluth, MN: *Association for Consumer Research*.

Castelo N., Bos M., & Lehman D., (2018). Consumer adoption of algorithms that blur the line between human and machine. *Graduate School of Business: Columbia University Working Paper*.

CEPEJ, (2018). Ethical charter on the use of artificial intelligence in judicial systems. Retrieved from https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment.

Cevikcan E., & Ustundag A., (2017). Industry 4.0: managing the digital transformation. Birmingham. Springer.

Charniak E., & McDermott D., (1985). Introduction to Artificial Intelligence. *Addison-Wesley*.

Chiriatti M., (2021). Incoscienza artificiale. Come fanno le macchine a prevedere per noi. *Luiss University Press*. Rome.

Chui M., Manyika J., Miremadi M., Henke N., Chung R., Nel P., & Malhotra S., (2018). Notes from the AI frontier: Applications and value of deep learning. *McKinsey global institute discussion paper*. Retrieved from https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai- frontier-applications-and-value-of-deep-learning.

Chui M., Manyika J., Miremadi M., Henke N., Chung R., Nel P., & Malhotra S., (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute*.

Coen E., (2020). Al futuro serve l'algoretica. Colloquio con Paolo Benanti. *L'Espresso*.

Colback L., (2020). The impact of AI on business and society. *Financial Times*.

Columbus L., (2019). 10 charts that will change your perspective of AI in marketing. *Forbes*. Retrieved from https:// www.forbes.com/sites/louiscolumbus/2019/07/07/10-charts-that-will-change-your-perspective-of-ai-in-marketing/amp/

Commissione europea, comunicato stampa, Bruxelles, 8 aprile 2019, Orientamenti etici sull'intelligenza artificiale: proseguono i lavori della Commissione. Retrieved from ec.europa.eu

Conick H., (2017). The Past, Present and Future of AI in Marketing. *American Marketing Association AMA*. Retrieved from www.ama.org

Craig S. Smith., (2021). A.I. Here, There, Everywhere. *The New York Times*. Retrieved from https://www.nytimes.com/2021/02/23/technology/ai-innovation-privacy-seniors-education.html

Crawford K., (2021). Né intelligente né artificiale. Il lato oscuro dell'IA. *Il Mulino*. Bologna.

Dal Co M., (2022). Chips Act, l'Europa corre con Cina e Stati Uniti: ecco i piani e gli ostacoli. Retrieved from www.agendadigitale.eu.

Dameski A., (2018). A comprehensive ethical framework for AI entities: Foundations. *Artificial General Intelligence*, 42–51.

Dastin J., (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Retrieved from https://www.reuters.com/article/ us-amazon- com-jobs-automation-insight/ amazon-scraps-secret-ai-recruiting-tool-that- showed -bias-against-women-idUSKCN1MK08G

Dastin J., (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Davenport T. H., Dalle Mule L., & Lucker J., (2011). Know what your customers want before they do. *Harvard Business Review*, 89(12), 84-92.

Davenport T., Guha A., Grewal D., & Bressgott T., (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48(1)*, 24-42.

De Bruyn A., Viswanathan V., Beh Y. S., Brock J. K. U., & Von Wangenheim F., (2020). Artificial intelligence and marketing: Pitfalls and opportunities. *Journal of Interactive Marketing, 51(1)*, 91-105.

De Cesco A. F., (2018). Amazon e l'intelligenza artificiale sessista: non assumeva donne. *Corriere della Sera*. Retrieved from https://www.corriere.it/tecnologia/18_ottobre_10/amazon-intelligenza-artificiale-sessista-non-assumeva-donne-4de90542-cc89-11e8-a06b-75759bb4ca39.shtml

De Luca S., (2020). Hic sunt leones. La democrazia nell'era dei social media, dei big data e dell'intelligenza artificiale. *Rivista di Politica*, 2, 147-159.

Del Porto M., (2019). MIT, quali rischi dall'intelligenza artificiale. *MIT Sloan Management Review*. Retrieved from https://www.futuro-europa.it/29636/science-tech/mit-quali-rischi-dallintelligenza-artificiale.html

Delecraz S., Eltarr L., Becuwe M., Bouxin H., Boutin N., Oullier O., (2022). Making Recruitment More Inclusive: Unfairness Monitoring With A Job Matching Machine-Learning Algorithm. *IEEE/ACM International Workshop on Equitable Data & Technology*, 34-41.

Deloitte, (2019). Artificial intelligence: From expert-only to everywhere. Retrieved from https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/cloud-based-artificial-intelligence.html

Di Fraia G., (2020). Fare marketing con l'AI. *Hoepli Editore*.

Dignum V., (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology, 20* (1).

Dreyfus H. L., (1988). Che cosa non possono fare i computer. I limiti dell'intelligenza artificiale. *Armando*.

Econopoly, (2022). L'Ue e l'intelligenza artificiale (AI): i vantaggi per aziende e cittadini. *Il Sole 24 Ore*. Retrieved from https://www.econopoly.ilsole24ore.com/2022/01/31/ai-ue-intelligenza-artificiale/

European Commission For Efficiency Of Justice (CEPEJ), (December 2018). "Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi". Retrieved from https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348.

European Commission, (21 April 2021). Regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'unione. Press release, Brussels. Retrieved from eur-lex.europa.eu

European Parliament, (2020). Quali sono i rischi e i vantaggi dell'intelligenza artificiale?. Retrieved from https://www.europarl.europa.eu/news/it/headlines/society/20200918STO87404/quali-sono-i-rischi-e-i-vantaggi-dell-intelligenza-artificiale

European Parliament, (2021). Che cos'è l'intelligenza artificiale e come viene usata?. Retrieved from https://www.europarl.europa.eu/news/it/headlines/society/20200827STO85804/che-cos-e-l-intelligenza-artificiale-e-come-viene-usata

European Union Agency for Fundamental Rights, (2018). "#BigData: Discrimination in data-supported decision making". Retrieved from https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf

European Union Agency for Fundamental Rights, (2019). Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights. Retrieved from fra.europa.eu

Ferrer X., Van Nuenen T., Such J. M., Coté M., Criado N., (2021). Bias and discrimination in AI: a cross-disciplinary perspective. *IEEE Technol. Soc. Mag. 40(2)*, 72-80.

Fiore P., (2021). Non abbiate paura dell'intelligenza artificiale. *Agenzia Giornalistica Italiana*. Retrieved from www.agi.it

Fjeld, J., et al., (2020). Principled artificial intelligence. Mapping consensus in ethical and rights-based approaches to principles for IA. *Berkman Klein Center, Harvard University*.

Fleming P., (2019). Robots and organization studies: Why robots might not want to steal your job. *Organization Studies*, 40(1), 23-38.

Floridi L., (2022). Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide. *Raffello Cortina Editore*. Milan.

Floridi L., Cabitza F., (2021). Intelligenza artificiale. L'uso delle nuove macchine. Ed. Kindle. *Bompiani*. Milan.

Floridi L., Cowls J., Beltrametti M., et al., (2018). AI4People-an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28(4)*, 689-707.

Floridi L., Sanders J.W., (2004). On the morality of artificial agents. *Minds and Machines*, 14, 349-379.

Forbes, (2019). Rethinking Weak Vs. Strong AI. Retrieved from https://www.forbes.com/sites/cognitiveworld/2019/10/04/rethinking-weak-vs-strong-ai/?sh=5d7e4fb16da3

Forbes, (2021). Can We Trust AI? When AI Asks For Human Help (Part One). Retrieved from https://www.forbes.com/sites/charlestowersclark/2021/01/25/can-we-trust-ai-when-ai-asks-for-human-help-part-one/?sh=b871774ee46a

Forbes, (2022). AI Ethics: What It Is And Why It Matters. Retrieved from https://www.forbes.com/sites/nishatalagala/2022/05/31/ai-ethics-what-it-is-and-why-it-matters/?sh=39f69c683537

Forbes, (2022). How AI Solutions Can Defend Against Cyberattacks. Retrieved from https://www.forbes.com/sites/forbestechcouncil/2022/11/07/how-ai-solutions-can-defend-against-cyberattacks/?sh=2fe6684c6693

Forbes, (2022). The Problem With Biased AIs (and How To Make AI Better). Retrieved from https://www.forbes.com/sites/bernardmarr/2022/09/30/the-problem-with-biased-ais-and-how-to-make-ai-better/amp/

Fowler G., (2019). The doorbells have eyes: The privacy battles brewing over home security cameras. Retrieved from https://www.sltrib.com/news/business/2019/02/01/doorbells-have- eyes/.

Frankenfield J., (s.d.). Natural Language Processing (NLP). Retrieved from https://www.investopedia.com/terms/n/natural-languageprocessing-nlp.asp

Gans J., Agrawal A., & Goldfarb A., (2017). How AI will change strategy: A thought experiment. *Harvard business review online*. Retrieved from https://hbr.org/product/how-ai-will-change-strategy-a-thought-experiment/H03XDI-PDF-ENG.

Garg V., Rani S., & Matta H., (2019). Consumer Adoption of Smart Biometric Lock Among SAARC Nations. *International Journal of Engineering and Advanced Technology* (IJEAT), 8(4C), 138-144.

Garg V., Singhal A., & Tiwari P. (2018)., A Study on Transformation in Technological Based Biometrics Attendance System: Human Resource Management Practice. 8th International Conference on Cloud Computing, *Data Science & Engineering*, 809-813, *IEEE*.

Garg V., Srivastav S., & Gupta A., (2018). Application of Artificial Intelligence for Sustaining Green Human Resource Management. International Conference on Automation and Computational Engineering, 113-116. *IEEE*.

Garnham A., (2017). Artificial intelligence: an introduction. *Routledge*.

Gaucher D., Friesen J., & Kay A. C., (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, 101(1), 109-128.

Gaudin S., (2016). At stitch fix, data scientists and a.I. become personal stylists. Retrieved from https://www.computerworld.com/article/3067264/artificial-intelligence/at-stitch-fix-data-scientists-and-ai-become-personal-stylists.html.

Geetha R., & Bhanu S., (2018). Recruitment through artificial intelligence: a conceptual study. *International Journal of Mechanical Engineering and Technology*, 9(7), 63-70.

Gerd L., (2019). Tecnologia vs Umanità. Lo scontro prossimo venturo. Ed. Kindle. *Egea*. Milan.

Giebelhausen M., Robinson S. G., Sirianni N. J., & Brady M. K., (2014). Touch versus tech: When technology functions as a barrier or a benefit to service encounters. *Journal of Marketing*, 78(4), 113- 124.

Goldfarb A., & Tucker C., (2013). Why managing consumer privacy can be an opportunity. *MIT Sloan Management Review*, 54(3), 10-12.

GPDP, (2022). Linee guida sull'utilizzo dei cookie e di altri strumenti di tracciamento. *Infografica*.

Gray K., & Wegner D. M., (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125-130.

Groom V., Srinivasan V., Bethel C. L., Murphy R., Dole L., & Nass C., (2011). Responses to robot social roles and social role framing. *In Proceedings of the IEEE International Conference on Collaboration Technologies and Systems (CTS)*, 194-203.

Gross S. R., Possey M., & Stephens K., (2022). Race and Wrongful Convictions in the United States. Report, *National Registry of Exonerations*.

Guardini R., (2021). Etica. Lezioni all'Università di Monaco (1950-1962). *Morcelliana*. Brescia.

Gupta R., Bhardwaj G., & Singh G., (2019). Employee Perception and Behavioral Intention to Adopt BYOD in the Organizations. International Conference on Automation, Computational and Technology Management, 73-78. *IEEE*.

Haenlein M., & Kaplan A., (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence, *California Management Review*.

Hage J., (2017). Theoretical foundations for the responsibility of autonomous agents. *Artificial Intelligence and Law, 25*, 255-271.

Hage J., (2017). Theoretical foundations for the responsibility of autonomous agents. *Artificial Intelligence and Law*, 25, 255-271.

Harding K., (2017). AI and machine learning for predictive data scoring. Retrieved from https://www.objectiveit.com/blog/use-ai-and-machine-learning-for-predictive-lead-scoring on 13 February 2019.

Harris D. M., & DeSimone R. L., (1994). Human Resource Development. *Harcourt College Pub*.

Harvard Law Review, (2017). State v. Loomis. Retrieved from https://harvardlawreview.org/2017/03/state-v-loomis/

Haugeland J., (1985). Artificial Intelligence: The Very Idea. *MIT Press*. Cambridge.

Hayes A., (2015). The unintended consequences of self-driving cars. Retrieved from https://www.investopedia.com/articles/investing/090215/unintended-consequences-selfdriving cars.asp.

Heer J., (2018). The Partnership on AI. *AI Matters*, 4 (3), 25-26.

Hern A., (2018). Google's solution to accidental algorithmic racism: ban gorillas. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorillablack-people

Hmoud B., & Varallyai L., (2019). Will Artificial intelligence Take Over Human Resources Recruitment and Selection?. *Network Intelligence Studies*, 7(13), 21-30.

Hoffman D. L., & Novak T. P., (2018). Consumer and object experience in the internet of things: An assemblage theory approach. *Journal of Consumer Research*, 44(6), 1178-1204.

Huawei Technologies Co., Ltd, (2022). A General Introduction to Artificial Intelligence. *Artificial Intelligence Technology*, 1-41.

IBM, (2022). IBM Global AI Adoption Index 2022. *Watson*. Retrieved from https://www.ibm.com/downloads/cas/GVAGA3JP

Intermedia Channel (2023). Il mercato italiano dell'intelligenza artificiale vale 500 mln (+32%). Retrieved from https://www.intermediachannel.it/2023/02/03/il-mercato-italiano-dellintelligenza-artificiale-vale-500-mln-32/

Jia Q., Guo Y., Li R., Li Y., & Chen Y. (2018)., A Conceptual Artificial Intelligence Application Framework in Human Resource Management. *Proceedings of the International Conference on Electronic Business*, 106-114.

Kaplan A., & Haenlein M., (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25. Retrieved from https://doi.org/10.1016/j.bushor.2018.08.004

Kayser-Bril N., (2020). Google apologizes after its Vision AI produced racist results. *AlgorithmWatch*. Retrieved from algorithmwatch.org

Khatri S., Pandey D. K., Penkar D., & Ramani J., (2020). Impact of Artificial Intelligence on Human Resources. *Data Management, Analytics and Innovation*, 2, 365-376.

Kim T., & Duhachek A., (2018). The impact of artificial agents on per- suasion: A construal level account. *ACR Asia-Pacific Advances*.

Knight W., (2017). Forget killer robots - bias is the real AI danger. Retrieved from https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/.

Kodiyan A. A., (2019). An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. *Researchgate Preprint*, 1-19.

Kop M., (2020). AI & Intellectual Property: Towards an Articulated Public Domain. *Texas Intellectual Property Law Journal*, 28.

Kopalle P. K., Gangwar M., Kaplan A., Ramachandran D., Reinartz W., & Rindfleisch A., (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing*, *39*(2), 522-540.

Kretschmer M., et al., (2022). Artificial intelligence and intellectual property: copyright and patents-a response by the CREATe Centre to the UK Intellectual Property Office's open consultation. *Journal of Intellectual Property Law & Practice*, 17(3), 321-326.

Kulkarni S. B., & Che X. (2019). Intelligent Software Tools for Recruiting. *Journal of International Technology and Information Management*, 28(2), 2-16.

Larson K., (2019). Data privacy and AI ethics stepped to the fore in 2018. Retrieved from https://medium.com/@Smalltofeds/data-privacy-and-ai-ethics-stepped-to-the-fore-in-2018-4e0207f28210.

Legg S., Hunter M., (2007). A collection of definition of intelligence. *Advances in Artificial General Intelligence. Concepts, Architecture, and Algorithms*, 17-24. IOS press. Amsterdam.

Letheren K., Russell-Bennett R., & Whittaker L., (2020). Black, white or grey magic? Our future with artificial intelligence. *Journal of Marketing Management, 36(3-4)*, 216-232.

Leung E., Paolacci G., & Puntoni S., (2018). Man versus machine: Resisting automation in identity-based consumer behavior. *Journal of Marketing Research*, 55(6), 818-831.

Levin S., (2017). New AI can guess whether you're gay or straight from a photograph. Retrieved from https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a- photograph.

Li D., Rau P. P., & Li Y., (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2), 175-186.

Licata P., (2019). Recruiting: cos'è e come assumere il candidato migliore in azienda oggi. *Digital4Executive*. Retrieved from https://www.digital4.biz/hr/talent-management/recruiting-10-passi-compiereunefficace-ricerca-del-personale/

LinkedIn, (2019). "LinkedIn Releases 2019 Global Talent Trends Report". Retrieved from https://news.linkedin.com/2019/January/linkedin-releases-2019-global- talent- trends-report

Liquori E., (2020). Marketing artificialmente intelligente. AI, uomo e macchina. *Dario Flaccovio Editore*.

Loble L., (2021). Learning to live in the time of AI. *The UNESCO courier*, 26-27.

Longoni C., Bonezzi A., & Morewedge C. K., (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*.

Lowy J., (2016). Self-driving cars are 'absolutely not' ready for deployment. Retrieved from https://www.pbs.org/newshour/science/self-driving-cars-are-absolutely-not-ready-for-deployment.

Luo X., Tong S., Fang Z., & Zhe Q., (2019). Machines vs. humans: The impact of chatbot disclosure on consumer purchases. Unpublished working paper.

Madgavkar A., (2021). A conversation on artificial intelligence and gender bias. *McKinsey*.

Manheim K., & Kaplan L., (2019). Artificial intelligence: Risks to privacy and democracy. *Yale Journal of Law & Technology*, 21, 106-188.

Manyika J., & Silberg J., (2019). Tackling bias in artificial intelligence (and in humans). *McKinsey Global Institute*.

Manyika J., & Sneader K., (2018). AI, automation, and the future of work: Ten things to solve for. *McKinsey*.

Manyika J., Silberg J., & Presten B., (2019). What Do We Do About the Biases in AI?. *Harvard Business Review*.

Markoff J., (2009). "Computer Program to Take On 'Jeopardy!'". *New York Times*. Retrieved from https://www.nytimes.com/2009/04/27/technology/27jeopardy.html

Marr B., (2019). How robots, IoT and artificial intelligence are changing how humans have sex. Retrieved from https://www.forbes.com/sites/bernardmarr/2019/04/01/how-robots-iot-and- artificial-intelligence-are-changing-how-humans-have-sex/# 3679d398329c.

Marr B., (2022). The 5 Biggest Artificial Intelligence (AI) Trends In 2023. Retrieved from https://www.linkedin.com/pulse/5-biggest-artificial-intelligence-ai-trends-2023-bernard-marr/

Martin K. D., & Murphy P. E., (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, 45(2), 135-155.

Martin K. D., & Murphy P. E., (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, 45(2), 135-155.

Martin K. D., Borah A., & Palmatier R. W., (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, 81(1), 36-58.

Martini C. M., (2015). Le cattedre dei non credenti. Ed. Kindle. *Bompiani*. Milan.

McGovern S., Pandey V., Gill S., Aldrich T., Myers C., Desai C., Balasubramanian V., (2018). The new age: artificial intelligence for human resource opportunities and functions. *Ernst & Young*.

McLuhan M., (1986). Gli strumenti del comunicare. *Garzanti*. Milan.

Mehryar M., Afshin R., & Ameet T., (2018). Foundations of machine learning (2 ed.). *MIT Press*, London.

Mehta N., Detroja P., & Agashe A., (2018). Amazon changes prices on its products about every 10 minutes – here's how and why they do it. Retrieved from https://www.businessinsider.com/amazon-price-changes-2018-8?international=true&r=US&IR= T.

Mende M., Scott M. L., Van Doorn, J., Grewal D., & Shanks I., (2019). Service robots rising: How humanoid robots influence service experiences and food consumption. *Journal of Marketing Research*, 56(4), 535-556.

Merlin P. R., & Jayam R., (2018). Artificial Intelligence in Human Resource Management. *International Journal of Pure and Applied Mathematics*, 119(17), 1891-1895. Retrieved from https://acadpubl.eu/hub/2018-119-17/2/153.pdf

Metz C., (2018). Mark Zuckerberg, Elon musk and the feud over killer robots. *New York Times*. Retrieved from https://www.nytimes.com/2018/06/09/technology/elon-musk-mark-zuckerberg-artificial-intelligence.html.

Miller G. (2016). Bots, chatbots and artificial intelligence, what the evolution is all (aBot) about. Retrieved from https://chatbotnewsdaily.com/bots-chatbots-and-artificial-intelligence-what-the-evolution-is-all-abot-about-a7e148dd067d.

MIT (Massachusetts Institute of Technology), (1961), conference Management and the Computer of the Future.

MIT Technology Review, (2019). Training a single AI model can emit as much carbon as five cars in their lifetimes. Retrieved from https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Montréal Declaration for a Responsible Development of AI, (2017). University of Montréal.

Moon Y., (2003). Don't blame the computer: When self-disclosure moderates the self-serving bias. *Journal of Consumer Psychology*, 13(1-2), 125-137.

Morelli C., (2018). "Algoritmi e diritti umani: si rischia la collisione immaginate a danno di chi?". Retrieved from https://www.altalex.com/documents/news/2018/07/09/algoritmi-e-impatto-sui-diritti-umani

O'Neil C., (2018). Amazon's Gender-Biased Algorithm Is Not Alone. *Bloomberg*. Retrieved from https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone

Occorsio E., (2018) Affari&Finanza. Intelligenza artificiale avanti tutta il dialogo uomo-macchina porterà 6mila miliardi di valore. Retrieved from *La Repubblica*.

Oracle, (s.d.). Database. Retrieved from https://www.oracle.com/it/database/what-is-database.html

Orlikowski W. J., (2000). Using technology and constituting structures. A practice lens for studying technology in organisations. *Organisation Science*, 11, 404-428.

Parekh J., (2018). Why Programmatic provides a better digital marketing landscape. Retrieved from https://www.adweek.com/programmatic/why-programmatic-provides-a-better-digital-marketing-landscape/.

Parisi F., (2019). La tecnologia che siamo. Codice Edizioni, ed. Kindle. Torino.

Pasquale F., (2021). Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'intelligenza artificiale. *Luiss University Press*. Rome.

Pedersen I., Reid S., & Aspevig K., (2018). Developing social robots for aging populations: A literature review of recent academic sources. *Sociology Compass*, 12(6).

Piccarozzi M., Aquilani B., & Gatti C., (2018). Industry 4.0 in Management Studies: A Systematic Literature Review. *Sustainability*, 10, 24.

Pictet, (2020). Le 20 innovazioni che hanno cambiato il mondo negli ultimi 20 anni. Retrieved from https://am.pictet/it/blog/articoli/tecnologia-e-innovazione/le-20-innovazioni-che-hanno-cambiato-il-mondo-negli-ultimi-20-anni

Pope Francis, (2019), speech in the seminar The Common good in the digital age. 27 September 2019.

Puntoni S., Reczek R. W., Giesler M., & Botti S., (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing, 85(1)*, 131-151.

PwC, (2020). PwC's Global Artificial Intelligence Study: Sizing the prize. *PwC Report*. Retrieved from www.pwc.com/AI

Rahwan I., Cebrian M., Obradovich N., Bongard J., Bonnefon J. F., Breazeal C., et al., (2019). Machine behavior. *Nature*, 568(7753), 477-486.

Rajesh D. S., Kandaswamy M. U., & Rakesh M. A., (2018). The impact of Artificial Intelligence in Talent Acquisition Lifecycle of organizations. *International Journal of Engineering Development and Research*, 6(2), 709-717.

Reese B., (2018). The fourth age: Smart robots, conscious computers and the future of humanity. *New York, NY: Atria Books*.

Rezzani A., (2013). Big Data: Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati (1 ed.). *Maggioli Editore*. Milan

Rich E., & Knight K., (1991). Artificial Intelligence. *McGraw-Hill*, New York.

Rudman L. A., & Glick P., (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*(4), 743-762.

Rupli D., Hess R., Zimmermann J., Neumann U., Biasio L., & Saussier J., (2019). Supertrends. Driving forces. *Credit Suisse*.

Ruppel C. P., Zanini M., & Stieglitz S., (2020). Bias in recruitment and selection: A systematic literature review of artificial intelligence and algorithms. *Journal of Business Research*, 117, 929-944.

Russel S., Norvig P., (2021). Intelligenza artificiale. Un approccio moderno. *Pearson*. Milan-Torino.
Russell Stuart J., & Norvig P., (2020). Artificial Intelligence: A Modern Approach. *Pearson*.

Sadin E., (2019). Critica della ragione artificiale. Una difesa dell'umanità. *Luiss University Press*. Rome.

Samuel A., (1960). Some moral and technical consequences of automation - A refutation. *Science*, 132, 741-742.

Schalkoff R. J., (1990). Artificial Intelligence: An Engineering Approach. *McGraw-Hill*.

Schatsky D., Katyal V., Iyengar S., & Chauhan R., (2019). Why enterprises shouldn't wait for AI regulation. Retrieved from https://www2.deloitte.com/insights/us/en/focus/signals-for- strategists/ethical-artificial-intelligence.html

Schrift R. Y., Wertenbroch K., André Q., & Frank D.H., (2017). Threatening free will. Presentation at the Symposium on Alienation and Meaning in Production and Consumption, Technische Universität München.

Schwab K., (2016). La quarta rivoluzione industriale. *Franco Angeli*. Milan.

Science, (2018). L'algoritmo di assunzione sessista di Amazon potrebbe ancora essere migliore di un essere umano. Retrieved from http://it.scienceaq.com/Electronics/1002069605.html

Searle J., (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3 (3), 417-424.

Sekhri A., & Cheema J., (2019). The new era of HRM: AI reinventing HRM functions. *International Journal of Scientific Research and Review*, 7(3), 3073-3077.

Semoli A., (2019). AI marketing. *Hoepli Editore*.

Shariari K., (2017). Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems. *IEEE*.

Sherwood H., (2017). Robot priest unveiled in Germany to mark 500 years since reformation. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/may/30/robot-priest-blessu-2-germany-reformation- exhibition.

Siau K., & Wang W., (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *Journal of Database Management*, 31(2), 74-87.

Signorinelli C. M., (2018). Can Computers Become Conscious and Overcome Humans?. *Front. Robot. AI*, 5 (121).

Skilton M., & Hovsepian F., (2017). The 4th industrial revolution: Responding to the impact of artificial intelligence on business (1 ed.). Cham: Palgrave Macmillan.

Sky tg24, (2019). Facebook e privacy dopo Cambridge Analytica, le accuse e la difesa del social web. Retrieved from tg24.sky.it

Smith C., (2019). Understanding China's Confucian edge in the global AI race. *Quillette*. Retrieved from https://quillette.com/2019/02/14/understanding-chinas-confucian-edge-in-the-global-ai-race/

Spar B., Pletenyuk I., Reilly K., & Ignatova M., (2018). GLOBAL RECRUITING TRENDS. *Linkedin*. Retrieved from https://business.linkedin.com/content/dam/me/business/enus/talent-solutions/resources/pdfs/linkedin-global-recruiting-trends-2018-en-us2.pdf

Spatharou A., Hieronimus S., & Jenkins J., (2020). Transforming healthcare with AI: The impact on the workforce and organizations. *Mc Kinsey*.

Statista, (2022). AI-enabled climate action use case benefits. Retrieved from https://www.statista.com/statistics/1226263/worldwide-artificial-intelligence-climate-action-project-benefits/

Statista, (2022). Artificial intelligence and machine learning use cases for companies worldwide from 2020 to 2021. Retrieved from https://www.statista.com/statistics/1111204/machine-learning-use-case-frequency/

Sudmann A., (2019). The Democratization of Artificial Intelligence. Retrieved from https://library.oapen.org/bitstream/id/61810f6f-bf40-404d-b688-61e9ac3405a4/external_content.pdf

Suresh H., & Guttag, J. V., (2019). A framework for understanding unintended consequences of machine learning. *ArXiv*.

Tambe P., Cappelli P., & Yakubovich V., (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15-42.

Taylor A., (2020). Artificial Intelligence and the Future of Sales and Marketing. *Harvard Business Review*.

The Economist, (2022). Two new books explore the upside of big data and AI. Retrieved from https://www.economist.com/culture/2022/11/30/two-new-books-explore-the-upside-of-big-data-and-ai

Thomas M., (2022) The Future of AI: How Artificial Intelligence Will Change the World. Retrieved from https://builtin.com/artificial-intelligence/artificial-intelligence-future

Thorhill J., (2021). Artificial intelligence bias can be countered, if not erased. *Financial Times*.

Totty M., (2020). How to make artificial intelligence less biased. *The Wall Street Journal*.

Tremolada L., (2021). L'intelligenza artificiale nell'ufficio del personale? Come evitare discriminazioni. *Il Sole 24 Ore*.

Trezza R., (2020). Diritto e intelligenza artificiale. Etica, privacy, responsabilità, decisione. *Pacini Editore*.

Tsamados A., Aggarwal N., Cowls J. et al., (2022). The ethics of algorithms: key problems and solutions. *AI & Society*, 215-230.

Tucker C., (2018). Privacy, algorithms, and artificial intelligence. In The Economics of Artificial Intelligence: An Agenda. *University of Chicago Press*.

Turing A., (1950). Computing Machinery and Intelligence. *Mind*, 236, 433-460.

Uzzi B., (2020). A simple tactic that could help reduce bias in AI, *Harvard Business Review*.

Verhoef P. C., Stephen A. T., Kannan P. K., Luo X., Abhishek V., Andrews M., et al., (2017). Consumer connectivity in a complex, technology-enabled, and mobile-oriented world with smart products. *Journal of Interactive Marketing*, 40, 1-8.

Verhoeven T., (2016). Candidate Experience (1 ed.). *Gabler Verlag*. Wiesbaden

Verhoeven T., (2020). Digitalisierung Im Recruiting (1 ed.). *Gabler Verlag*.

Villasenor J., (2019). Artificial intelligence and bias: Four key challenges. Retrieved from https://www.brookings.edu/blog/techtank/2019/01/03/artificial-intelligence-and-bias-four-key-challenges/.

Wall Street Journal, (2021). AI's Impact on Businesses—and Jobs. Retrieved from https://www.wsj.com/articles/ais-impact-on-businessesand-jobs-11615234143

Wang Y., & Kosinski M., (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, 114(2), 246-257.

Weissman J., (2018). Amazon created a hiring tool using a.I. it immediately started discriminating against women. Retrieved from https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html.

Weizenbaum J., (1972). On the impact of the computer on society. How does one insult a machine?. *Science*, 176, 612.

West D. M., (2018). The role of corporations in addressing AI's ethical dilemmas. *Brookings*.

Wiener N., (1960). Some moral and technical consequences of automation. *Science*, 131, 1355-1358.

Wilson J., Daugherty P., & Shukla P., (2016). How one clothing company blends AI and human expertise. *Harvard Business Review*. Retrieved from https://hbr.org/2016/11/how-one-clothing-company-blends-ai-and-human-expertise.

Winner L., (2004). Technologies as forms of life. Readings in the *Philosophy of Technology*. Lanham.

Wirtz J., Patterson P. G., Kunz W. H., Gruber T., Lu V. N., Paluch S., & Martins A., (2018). Brave new world: Service robots in the front-line. *Journal of Service Management*, 29(5), 907-931.

Wisskirchen G., Biacabe B. T., Bormann U., Muntz A., Niehaus G., Soler G. J., & Von Brauchitsch B., (2017). Artificial Intelligence and Robotics and Their Impact on the Workplace. *IBA Global Employment Institute*.

Wright R. P., Storey J., & Ulrich D., (2019). Strategic Human Resource Management: A Research Overview. *Routledge*. New York.

Yawalkar M., (2019). A Study of Artificial Intelligence and its role in Human Resource Management. *International Journal of Research and Analytical Reviews*, 6(1), 20-24.

Yehia N., (2005). Homo cyborg. Il corpo postumano tra realtà e fantascienza. *Eleutheria*. Milano.

Zampori I., (2021). Intelligenza Artificiale e Digital Marketing: come difendersi dai bias?. *Network digital 360*. Retrieved from www.digital4.biz

SUMMARY

INTRODUCTION

Artificial Intelligence (AI) has grown in prominence and is now a significant force that brings about transformation in many areas of our lives. AI has revolutionized numerous industries, including the field of recruitment: its ability to analyze vast amounts of data and make predictions has significantly streamlined the hiring process, enabling companies to identify suitable candidates more efficiently than ever before.
However, with this promising advancement comes a crucial concern: bias. As AI algorithms learn from historical data, they may inadvertently perpetuate biases present in the data, resulting in discriminatory outcomes. This ethical challenge raises questions about fairness and inclusivity in recruitment practices.
The first chapter provides a historical overview of AI and defines the concept, highlighting its impact on industry and society. It identifies the benefits and opportunities of AI, as well as the threats and risks associated with it. The chapter also discusses the future trends of AI and the need to examine the ethical considerations surrounding its integration. It delves into the dark side of AI, focusing on the issues of data, privacy, and bias. Are also presented case studies on COMPAS and Amazon to illustrate the real-world impact of these concerns.
Moving forward, the second chapter conducts a literature review that critically examines the ethical implications of AI, particularly in the recruitment context. It analyzes the intersection of AI and consumer patterns, algorithmic bias, and the potential benefits and challenges of AI in marketing and recruitment. The research questions and conceptual model for the subsequent empirical studies are also presented.
The third chapter details the research methodology and design, including specific descriptions of Study 1 and Study 2. The results from both studies are analyzed, leading to a comprehensive discussion of the findings. The chapter concludes by emphasizing the managerial implications of the research and suggesting avenues for future exploration. Overall, the thesis aims to explore the ethical issues and biases arising from the integration of AI in recruitment and provide insights for ethical decision-making and responsible AI practices.

CHAPTER 1: MANAGERIAL RELEVANCE

1.1 A brief history of Artificial Intelligence

Artificial intelligence (AI) is already a part of various fields, including medicine, economics, finance, industry, and social platforms. Many everyday tools and platforms, such as Spotify, Siri, Alexa, and Netflix, utilize AI through recommendation systems and algorithms. AI systems are widespread and aim to reduce human effort, increase speed, and improve accuracy across industries. The history of AI dates back to the second half of the last century, with early developments in electronic calculators and cybernetics. Notably, Alan Turing's work on computable numbers and his concept of the Turing machine laid the foundations for AI. In 1956, the term "artificial intelligence" was first used, and the following years saw significant inventions, including the

General Problem Solver program and the chatbot ELIZA. The turning point for AI came in 1997 when IBM's Deep Blue defeated the world chess champion, and later in 2009, Watson won the game show Jeopardy. Voice assistants like Siri, Cortana, Alexa, and Google's smart speaker emerged in the early 2010s. The advancements in AI have also led to automation, augmented and virtual reality, biometric authentication systems, and the integration of AI algorithms in social networks. Quantum computing is another area of discussion for its potential impact on AI. The progress in AI has been made possible by developments in computational power and exponential growth.

1.2 Definition of the phenomenon: what is AI?

Artificial Intelligence (AI) is a broad field with various sub-categories, encompassing areas such as learning, reasoning, playing chess, and proving mathematical theorems. Stuart J. Russell and P. Norvig define AI based on four elements: action, man, thought, and rationality. Four combinations arise from this definition: systems that act like humans, systems that think like humans, rationally thinking systems, and systems that act rationally. The first combination focuses on machines designed to perform tasks that require human intelligence. Alan Turing proposed a test in 1950 to determine if a machine behaves like a human, known as the Turing test. It evaluates a machine's ability to communicate, process language, memorize knowledge, reason, draw conclusions, and adapt through machine learning.

The second approach involves systems that think like humans, aiming to match the reasoning processes of humans. Insights from cognitive sciences and neuroscience contribute to the development of AI models based on neural patterns. The third notion involves rationally thinking systems based on logic and problem-solving programs. Logic and probability theory enable accurate reasoning, even with uncertain information. The combination of action and rationality occurs in an agent that perceives its environment and acts to achieve goals. Bounded rationality acknowledges the limitations of perfect rationality in complex environments and computational demands. A more accessible definition from the European Parliament describes AI as the ability of machines to display human-like capabilities such as reasoning, learning, planning, and creativity. AI allows systems to understand their environment, solve problems, and act toward specific goals.

To address concerns about AI, the term "augmented intelligence" is increasingly used to emphasize the support function of AI in enhancing human intelligence. However, the choice and use of terminology must be accurate and specific, as "augmented intelligence" and "artificial intelligence" are not synonymous.

1.2.1 Data and statistics on the impact of AI

AI adoption is accelerating due to the data challenges of the pandemic, and it is predicted that nearly 100% of organizations will be using AI by 2025, with the global AI market projected to reach $1,597.1 billion by 2030 (Bloomberg, 2022). China leads in AI adoption, followed by Singapore, India, and Italy. Larger companies are more likely to have implemented AI, while smaller companies are exploring its use. AI in sales and marketing has increased revenue and market share for 31% of respondents, and 67% believe it is critical for future competitiveness. Key drivers for AI adoption include technology accessibility, cost reduction, and implementation in standard business applications. Customer-centric applications are the most common AI use cases, with a focus on customer experience, generating insights, and customer interaction. Companies are saving time and reducing costs with AI automation. The shortage of qualified personnel is a challenge for companies adopting AI, with 34% citing skills as the main obstacle. Companies are investing in retraining existing employees, recruiting talent, and hiring college students. The most requested skills are coding, governance, security, ethics, data analysis, and advanced degrees related to AI. Problem-solving is considered the most critical soft skill. Transparency and accountability are key challenges for AI projects, including reducing bias, accounting for performance variations, and ensuring explainability.

1.2.2 Main benefits and opportunities

Artificial intelligence (AI) offers various benefits in different aspects of society. In terms of businesses, AI can lead to the development of new products and services, improve sales processes, enhance machinery maintenance, increase productivity and quality, and save energy. It is estimated that AI could increase labor productivity by 11% to 37% by 2035. In public services, AI can reduce costs, provide new options in education and transportation, and contribute to achieving environmental sustainability goals (Statista, 2022). The use of AI in public services is expected to reduce global greenhouse gas emissions by 1.5% to 4% by 2030 (European Parliament, 2020). AI can also strengthen democracy by enabling data-driven audits, preventing disinformation and cyber-attacks, and promoting diversity and equality of opportunity. Additionally, AI has significant applications in security, including crime prevention, criminal justice, and defense strategies. Many companies are investing in AI projects with the goals of cost reduction, revenue increase, and decision-making support (Il Sole 24 Ore, 2022). A majority of companies that have launched AI projects report success, highlighting the potential benefits of AI at technological, economic, and organizational levels.

1.2.3 Threats and potential risks

The increasing use of artificial intelligence (AI) in digital marketing offers numerous benefits, but it also raises concerns and limitations. Privacy violations, algorithm fairness and transparency, societal impact, and job losses are among the main concerns. However, in the marketing sector, the required skills are expected to focus more on data science and creativity (Conick H., 2017). Overuse and underuse of AI are risks that need to be addressed (Floridi L., 2018). Underutilization can lead to economic stagnation, while abuse can result in using AI for unsuitable purposes. Determining liability for AI-related damages is a challenge, as it involves questions of responsibility for accidents and product safety. AI's impact on fundamental rights and democracy is another concern, with biases, data protection, and the manipulation of information being potential threats. AI's effect on employment, including job displacement and the need for adequate training, is important to consider. The accumulation of information and competition distortion, security risks, and inequalities in access to information also pose potential threats. Transparency issues arise regarding the user's awareness of interacting with an AI system versus a human (The Economist, 2022).

1.3 Future Trends of AI

Several key trends will shape the future of AI, including the democratization of AI, augmented working, commercialization of generative AI, development of sustainable AI, and the priority of ethics in AI. The democratization of AI will make it easier for businesses to access deeper insights from data, and cloud service providers will incorporate AI into their services, leading to wider adoption. Augmented working will involve working alongside intelligent machines and robots, improving efficiency and providing real-time information. Generative AI will be commercialized, enabling new products and services, particularly in speech-to-speech technology and customer service (Statista, 2022). The development of sustainable AI will address the environmental impact of AI while leveraging it to tackle global challenges. The last trend is that AI ethics will become a top priority (Harvard Business Review, 2021). In the next years, efforts will be made to overcome the AI "black box" problem by improving transparency and eliminating biases (Bloomberg, 2022).

1.4 Ethics and AI: a possible combination?

The intersection of ethics and AI is an intriguing topic, considering the potential implications. The availability of power and technology does not necessarily align with ethical goals. It is challenging to determine whether an AI-guided machine can behave ethically, and the issue of intellectual property ownership by machines poses legal complexities. There is a distinction between strong and weak AI. Strong AI aims to create machines that emulate human cognitive abilities, while weak AI focuses on developing tools for specific tasks. Trust in AI remains a major concern, involving debates on various disciplines and addressing issues such as professional displacement, algorithmic decision-making, and biased judgments. To build trust in AI, IBM offers an Action Guide that emphasizes ethics: it covers areas such as corporate strategy, governance, and implementation. Ethics are essential for AI innovation, and by adhering to ethical standards, developers and designers contribute to a technology-driven world rooted in human values.

1.4.1 The dark side of AI: limits and ethical implications

The ethical challenges associated with AI include biases in algorithms, transparency in decision-making processes, privacy protection, respect for human rights, well-being of workers, environmental impact, and building trust in AI. The European Commission has established key requirements for the development of reliable AI, focusing on human action, robustness, privacy, transparency, diversity, well-being, and responsibility. The European Commission has proposed the Artificial Intelligence Act (AIA) to regulate AI systems and prohibit certain applications. However, some researchers argue that the regulations may not be sufficient to address risks related to AI technologies, such as the use of predictive algorithms in decision-making processes. Innovation in AI brings uncertainties and the need for education and training. Uncontrolled development of AI without regulation could lead to social, economic, and political damages. The use of AI in marketing raises concerns about data accuracy, consumer privacy, biases, governance, and the future of work.

1.4.2 Data, Privacy and Bias

Data is the lifeblood of AI: data acquisition and interpretation are crucial in the new digital economy. However, there are sources of error in data acquisition processes that can lead to inaccurate results and violate fundamental rights. The principle of "garbage in - garbage out" applies, meaning that low-quality data leads to low-quality results. The poor quality of data can result in measurement and representation errors, affecting data accuracy and validity. There are concerns about data misuse and violation of privacy rights. Users often provide personal information without fully understanding the implications. Improper data use can lead to manipulation and discrimination, affecting consumer behavior and preferences. In addition, biased data entered by humans reflects the current society and can result in discriminatory decisions made by algorithms. The difficulty of correcting algorithmic biases lies in identifying the source of error and requires constant monitoring. Mitigating algorithmic biases requires technological advancements and investment in education. The future coexistence of humans and AI depends on awareness, responsibility, and humanity. The ideal cooperation with AI can solve social and environmental problems and promote sustainability and equity.

1.5 Case studies: COMPAS and Amazon

Artificial intelligence (AI) instruments are increasingly integrated into various aspects of our lives, including judicial systems and search engine recommendations. However, concerns about AI bias have arisen, as these systems can make unfair decisions that harm specific groups of people. The case of the COMPAS Recidivism Algorithm highlights the potential for AI algorithms to exhibit prejudice. COMPAS is an AI-based software used in American courts to assess the risk of recidivism for criminal defendants. However, studies have shown that the system's decisions are discriminatory, particularly against African Americans. A study by ProPublica analyzed over 10,000 criminal defendants in Florida and found that black defendants were often assigned higher risk scores than their white counterparts, even when controlling for factors such as previous crimes, recidivism, age, and sex. The analysis also revealed that the algorithm's predictions of violent recidivism were only correct 20% of the time. This bias in the algorithm's predictions is attributed to racial prejudice and the biased historical data used for training. Similar issues of bias have been observed in other AI systems.

For example, Amazon developed an AI recruiting tool that unintentionally discriminated against women. The tool learned from resumes submitted to the company over a ten-year period, which were predominantly from male applicants. As a result, the algorithm downgraded resumes from female candidates, perpetuating gender inequality. These cases highlight the challenges of building unbiased AI systems. Algorithms can inadvertently learn biases from training data or reflect the prejudices present in society. It is crucial to address these biases through transparent and ethical practices when developing and deploying AI systems.

CHAPTER 2: LITERATURE REVIEW

2.1 Artificial intelligence and its ethical implications

Klaus Schwab's text on the fourth industrial revolution highlights the transformative nature of the technological revolution humanity is currently experiencing. The blurring boundaries between physical, digital, and organic through the advent of artificial intelligence (AI) require a shared vision to understand and navigate this change. Various experts emphasize the impact of technology on shaping civilization, with some considering it an autonomous force beyond human control while others recognize its power in shaping and giving meaning to human activity. The myths surrounding AI, such as equating it to human intelligence or considering it as an independent element, need to be debunked. Historically, debates have arisen regarding the capabilities of AI, with a distinction between explicit and formalized computer processes and the more nuanced and unconscious processes of human intelligence. Despite advancements in AI techniques, skepticism remains regarding its abilities. AI has rapidly expanded in academia and industry, with powerful technology companies deploying AI systems on a global scale. However, there is no consensus on the definition of AI. The ethical challenges of AI systems engage researchers from various disciplines, exploring the division between problem-solving and intelligent behavior and the potential devaluation of human values. The actions of AI systems and their impact on morality raise questions about intentionality, as human actions are connected to the environment, similar to AI systems. The increasing integration of technology in our lives calls for intelligent human design to shape future interactions with AI and ensure that technology serves human interests. The ethical debate surrounding AI has been ongoing for decades, but recent years have witnessed an increased focus on its societal impact.

2.2 Exploring the intersection of AI and consumer patterns

New technologies, including AI, often influence customer behavior (Hoffman D. L., & Novak T. P., 2018). Negative perceptions of AI, such as its inability to feel or recognize uniqueness, hinder its adoption. Mitigating strategies include presenting AI as a learning organism or combining AI and human inputs. Allowing customers to customize AI can shift their focus to the benefits of personalization. Discomfort with AI is heightened when it is embedded in robots that become more human-like, creating unnerving effects. The perception of AI as a servant or partner may moderate this effect. Efforts to foster empathy and anthropomorphize AI can alleviate concerns. Research also explores cultural attitudes towards robots and the factors influencing customer acceptance of robot-assisted services. AI applications can trigger different mindsets in consumers, and communication should align with these mindsets. Interaction with AI-embedded robots can lead to discomfort and compensatory behaviors, and it is important to understand if these perceptions improve over time. AI may challenge customers' preferences by presenting choices based on past behaviors. Loss of autonomy is a concern when AI predicts preferred choices, and customers may deliberately choose alternative options to assert their autonomy. There are fears of a loss of human connection when humans form bonds with robots with embedded AI, raising concerns about social isolation and declining birth rates.

2.3 Addressing algorithmic bias in AI applications

The literature highlights three key areas where policymakers focus on ensuring a fair balance between business interests and customer welfare in the context of artificial intelligence: data privacy, bias, and ethics. Concerns regarding data privacy include the longevity of data storage, data reformulation and reuse, and the potential inclusion of information on other individuals within personal data. The research explores whether data privacy management should be guided by legal regulation or self-regulation, considering cultural perspectives. Additionally, issues such as data breaches and the privacy-customization paradox are identified as important research topics. Bias in AI applications is examined, emphasizing algorithmic bias caused by biased data sets and the challenges of opaque algorithms. The research also addresses the inability of AI to distinguish between attributes that may induce bias and raises questions about the appropriateness of AI-based decisions based on sensitive factors. Finally, ethical considerations in AI implementation are discussed, including the relationship between data privacy choices, ethical concerns, and organizational strategies. The research also delves into the ethical implications of AI applications, such as the identification of sexual orientation through facial analysis, and the need to define appropriate applications for AI.

2.4 The future of marketing: harnessing the power of AI

Artificial intelligence is expected to transform customer behaviors and marketing strategies in various industries. AI is likely to influence sales processes, business models and customer service options. For example, AI-enabled and driverless cars may impact transportation business models and customer behavior. AI can also enhance sales processes through real-time conversation tracking and AI bots. Online retailers can anticipate customer needs and ship items without formal orders, changing business models and consumer behaviors. However, concerns related to data privacy, algorithmic biases, and ethics need to be addressed. The marketing discipline has the most to gain from AI, with potential value in sales domains. The impact of AI varies across industries, with sectors like banking, retail and consumer packaged goods experiencing the greatest influence. More research is needed to understand the role of AI in marketing strategies and customer behaviors, integrating insights from psychology, economics, and other social sciences. AI encompasses various technologies, such as machine learning and natural language processing. It automates business processes, gains insights from data, and engages customers and employees. AI has the potential to reduce costs and increase revenue, leading to better marketing decisions and augmenting human capabilities. The vision is a world where humans and machines work together rather than against each other (Carpenter J., 2015).

2.5 The impact of AI in the recruitment process

Technological development is important, particularly in the form of artificial intelligence, in automating routine tasks and deskilling non-managerial jobs (Bhardwaj, 2013; Wright et al., 2019). The fourth industrial revolution, characterized by fully automated and intelligent production systems, is transforming the world of work and requiring new skills and recruitment methods (Piccarozzi et al., 2018). Artificial intelligence, including machine learning, is a key aspect and encompasses various technologies aimed at developing human-like intelligence (Kaplan & Haenlein, 2019; Khatri et al., 2020). AI can be categorized into three phases: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial super intelligence (ASI) (Kaplan & Haenlein, 2019). AI has the potential to generate significant economic impact, particularly in marketing and sales, with estimated annual value added between 3.5 and 5.8 trillion dollars globally (Chui et al., 2018). By 2030, AI could boost global economic activity by approximately 13 trillion dollars (Rupli et al., 2019). The impact of AI on human resources is estimated to generate an added value of 0.1 trillion dollars (Chui et al., 2018). Overall, AI is a rapidly developing technology with significant present and future implications across various functions and industries.

2.5.1 The role of artificial intelligence in human resources

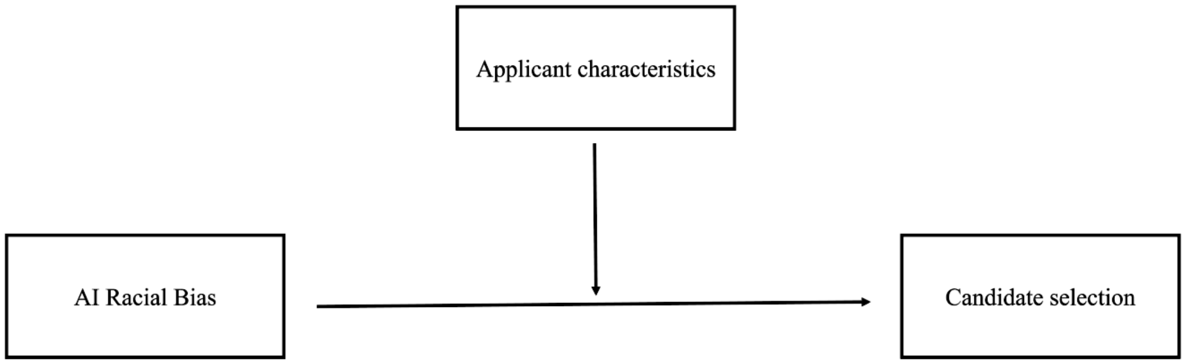
Artificial intelligence (AI) plays a crucial role in human resources (HR) by automating tasks, reducing errors, saving time and costs, and increasing productivity. AI technologies such as algorithms, big data, and machine learning are used to optimize various HR activities like recruitment, training, performance management, and employee retention. AI applications include CV screening, interview analysis, personalized training, bias reduction in performance evaluation, and predicting employee turnover. Voice recognition, bots, and algorithms are the main forms of AI in HR systems. Digital tools like applicant tracking systems, mobile recruiting, recommendations programs, search engine marketing, robotic process automation, chatbots, and augmented reality are transforming the recruitment process and improving candidate experience.

2.5.2 The use of learning algorithms for screening candidates

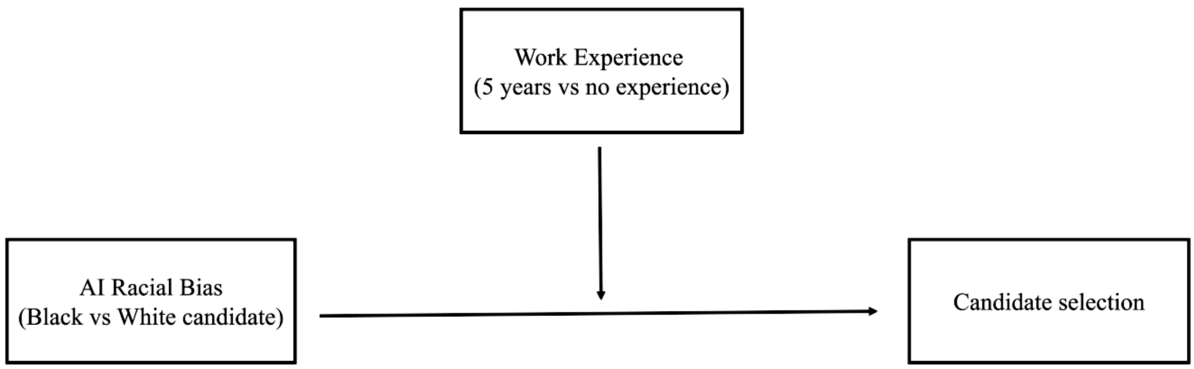
According to a report by LinkedIn, 76% of recruiters and hiring managers believe that AI will have a significant future impact on recruitment, particularly in candidate research and screening. Various digital tools incorporating AI are already being used in HR, such as video interviews that analyze facial expressions and speech patterns, chatbots that automate screening and evaluation tasks, and algorithms that identify and rank candidates. The introduction of AI in HR processes can improve talent mapping and job placement, reduce bias in hiring, enhance employability, provide immediate assistance, and improve the overall quality of recruitment. However, concerns remain regarding the potential biases and errors in AI systems, as highlighted by cases like Amazon’s discriminatory screening algorithm and Google’s image classification issue. Hence, further investment and development are needed to optimize and ensure the reliable use of AI in HR.

2.6 Research questions and conceptual model

AI in recruitment processes has the potential to make them faster, more efficient, and fairer. However, there are concerns about AI perpetuating biases and leading to discriminatory outcomes in hiring. Previous research has identified issues related to bias in AI algorithms, but there is a lack of concrete solutions and comparisons between different algorithms. There is also a gap in understanding how job candidates perceive AI in hiring processes and its impact on their job search behavior. This research aims to address these gaps by studying the influence of AI racial bias on candidate selection in recruitment, and whether applicant characteristics moderate this relationship.



In the previous conceptual model, AI racial bias is the independent variable, candidate selection is the dependent variable, and applicant characteristics act as a moderator. AI racial bias refers to biases exhibited by AI systems towards individuals or groups based on race. Candidate selection refers to the process of choosing a candidate for a job. Applicant characteristics include personal attributes such as race, gender, age, education level, and work experience.



The research hypotheses suggest that a white candidate (compared to a black candidate) and a candidate with 5 years of work experience (compared to no work experience) would increase candidate selection.
It is also hypothesized that work experience moderates the relationship between AI racial bias and candidate selection, with race having a greater influence when the candidate has no work experience. No differences are expected for candidate selection for a candidate with 5 years of work experience.
Overall, this research aims to shed light on the impact of AI racial bias in recruitment and the role of applicant characteristics, particularly work experience, in mitigating potential bias.

CHAPTER 3: METHODOLOGY

There is the need for a statistical investigation to fill the scientific gap in the literature regarding the influence of AI racial bias on candidate selection in recruitment. The study aims to address this gap by examining how AI racial bias in recruitment impacts candidate selection, particularly considering the race of candidates (black vs. white) and using work experience as a potential moderator. The research model aims to explore racial prejudice in candidate selection and its effect on real recruiters. To conduct the research, a mixed-method approach combining quantitative data collection through experiments and qualitative data collection through interviews with HR recruiters is adopted. This approach provides statistical information, in-depth insights, and a comprehensive understanding of the research questions and context, overcoming limitations associated with using only one method.

3.1 Study 1

The first study used an experimental design to test the relationship between variables. The data collection process involved defining the sample size and features and collecting data through an online Qualtrics experimental survey with four random scenarios. The study used a between-subject 2x2 factorial design, with candidate selection as the dependent variable and applicant characteristics as the moderator.
The independent variables were AI racial bias (black candidate vs. white candidate) and work experience (no experience vs. 5 years). The research objective was to examine how these variables influence candidate selection. The questionnaire included a hireability scale to measure the likelihood of hiring the candidate. Convenience sampling was used to recruit respondents, and a total of 282 participants from Italy were included in the analysis. The sample had a gender distribution of 48.9% men, 50.0% women, and 1.1% unspecified. The age of the sample ranged from 21 to 68 years (Mage=38,1915). The occupation of the participants included workers (72.0%), students (22.3%), retirees (3.5%), and unemployed (2.1%).

3.1.2 Study 2

Study 2 involved conducting interviews with real recruiters to gain deeper insights into the topic and validate the results of the previous experiment. Ten recruiters from different companies, primarily major banks, financial firms, engineering, and IT consultancy firms, were interviewed. Some recruiters from the fashion or luxury sector were also included to explore potential differences. The interviews were conducted in person during the Luiss Career Day event and online using Teams. The average duration of the interviews was 20 minutes. Recruiters were assured of anonymity, and the questions focused on the candidate selection process, evaluation criteria, objectivity and impartiality, diversity and inclusion policies, and assessment of candidates from different backgrounds. After asking the prepared questions, the experiment's graphical stimuli were shown to the recruiters, and their reactions and decision-making processes were discussed. The results of the interviews will be analyzed in conjunction with the experiment's findings, with relevant insights, common thoughts, and key points summarized to enhance the validity and value of the overall research.

3.2 Results

3.2.1 Study 1

After having cleaned the dataset and conducted descriptive statistics analysis to describe the sample, it was analyzed the validity and reliability of the "Hireability Scale" and tested the relationship between race and candidate selection. A Factor Analysis was performed to confirm the validity of the scale, and the results showed satisfactory values for the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test, indicating the sample's adequacy and the presence of significant correlations. The communalities and total explained variance also supported the inclusion of all three items in the scale.

The reliability of the Hireability Scale was assessed using Cronbach's alpha, which showed good reliability with a coefficient of 0.870. The scale was deemed reliable and suitable for statistical analysis. Moving on to the hypothesis testing, an independent samples T-test was conducted to examine the difference in means between white and black candidates in terms of candidate selection. The Levene's test confirmed equal variances between the groups. However, the T-test did not reject the null hypothesis of equal means, indicating that the difference in candidate selection between white and black candidates was not statistically significant. The third phase of the analysis involved investigating the impact of the moderator "work experience" on candidate selection for white and black candidates. A two-way ANOVA was performed, and the results indicated a significant main effect for work experience but not for race. Additionally, an interaction effect between race and work experience was observed, indicating that the relationship between race and candidate selection was influenced by work experience. The interaction plot revealed that the difference in candidate selection between white and black candidates was more pronounced for candidates with no work experience, but the difference decreased when considering candidates with five years of work experience.

3.2.2 Study 2

The results of Study 2, which involved interviews with recruiters, shed light on the ethical concerns and biases in the recruitment process. Recruiters prioritized transparency, efficiency, and fit in their selection processes. They emphasized multi-step assessments, feedback, and thorough evaluation. Recruiters had different criteria for evaluating candidates, with a focus on educational background, soft skills, and work experience. Technical skills were prioritized in some cases. Key competencies sought in candidates included technical expertise, communication skills, problem-solving, and adaptability. Recruiters ensured objective and impartial evaluations through training, multiple interviews, feedback normalization, and involving different perspectives. They valued candidates based on skills, qualifications, and experience rather than protected characteristics. Efforts were made to include diverse candidates and reduce biases.

Recruiters considered alignment with job requirements, transferable skills, and relevant experience. Many companies had policies promoting diversity and inclusion, while others were working on developing them. The presence of individuals from different ethnicities in companies varied, influenced by language requirements and international operations. The number of applicants from diverse ethnicities also varied across recruiters, but physical appearance or race were not determining factors in the evaluation process. Overall, recruiters gave importance to individual qualities, soft skills, fit with the position and company culture.

3.3 Discussion

The studies conducted provided valuable insights into the hiring decision-making process. The first study, quantitative and experimental, found no significant difference in preference between white and black candidates among ordinary people. However, candidates with five years of work experience were preferred over those with no experience, regardless of race. Moreover, with five years of work experience, no difference was observed between the white and black candidates, indicating that work experience might be a more critical factor in the hiring decision-making process than race. The second study, qualitative and involving real recruiters, confirmed these findings and emphasized the focus on qualifications and experience rather than race. Diversity and inclusion were mentioned as important, but not all companies have policies for it. Geographic factors and language proficiency also affected diversity in the candidate pool. Some recruiters observed less demand for candidates of different ethnicities due to industry dynamics and cultural norms. Implicit biases and the limitations of the recruitment process were highlighted. Overall, organizations should prioritize work experience over race, promote diversity and inclusion, and minimize biases and prejudices.

3.3.1 Managerial implications

This work contributes to the research already existing in the literature under various aspects and can be useful for managerial purposes. It focuses on the problem of bias in AI systems, particularly in the context of recruitment, and aims to understand how AI can be integrated into personnel offices while ensuring fairness and avoiding discrimination. Previous research has identified biases within AI algorithms that reinforce stereotypes and discriminate against candidates. This study further explores these biases and emphasizes the negative impact they can have, such as issues related to racism. It also emphasizes the importance of regulating AI and provides recommendations for organizations to create unbiased AI and promote greater accountability. The study suggests guidelines for managers and companies to combine ethics and AI, offering real suggestions for developing fair and inclusive AI solutions. This includes regular auditing and testing of algorithms, training hiring managers to recognize and mitigate implicit bias, and promoting transparency and accountability in recruitment processes.

3.3.2 Limitations and future research

This study has several limitations and suggests areas for future research. One limitation is that it focuses only on two races (black and white), potentially overlooking the impact of AI racial bias on other ethnicities. This may limit the generalizability of the findings. Another limitation is that the study assumes AI racial bias as the sole influencing factor in candidate selection outcomes, neglecting other factors like candidate qualifications. Thus, the study may overestimate the impact of AI racial bias. Another limitation is the lack of a clear definition of “AI racial bias”, which may hinder accurate measurement of its impact on candidate selection. Additionally, the data collection was exclusively done in Italy, which may not be representative of the entire country. The use of a convenience sample may introduce selection bias, necessitating future research to employ a stratified sampling technique for improved generalizability.

Future research should explore the impact of AI racial bias on other stages of the recruitment process and investigate other types of bias, such as gender or age bias. Examining different types of AI systems and their potential for racial bias is also important. Research should also focus on potential solutions to mitigate AI racial bias, such as algorithmic fairness or diversity training for recruiters.

Lastly, it could be valuable to explore the impact of AI racial bias beyond candidate selection, extending to performance evaluations or promotions for a comprehensive understanding of AI bias in the workplace.

CONCLUSION

In conclusion, this thesis examines the dark side of AI in the context of recruitment, focusing on ethical issues and biases associated with its use. Through a comprehensive analysis of AI’s history, definition, and potential, as well as its benefits, opportunities, threats, and risks, the research highlights the complexities and challenges of deploying AI systems. The study reveals the potential for biases and discrimination to be perpetuated through AI-driven recruitment processes, emphasizing the need for scrutiny and oversight. Case studies demonstrate how biases in training data can be amplified by AI, leading to discriminatory outcomes.

The literature review establishes the significance of ethical considerations in AI applications and the need for examination and proactive action. The benefits of AI must be balanced with addressing biases, preserving privacy, and preventing discrimination. The methodology used in this thesis provides valuable insights into the prevalence and implications of AI bias in recruitment. The results highlight the importance of evaluating and auditing AI algorithms to prevent bias and discrimination. The managerial implications stress responsible AI deployment, transparency, accountability, and fairness in the recruitment process.

The research acknowledges the limitations of the study and the evolving nature of AI, calling for further research and collaboration to keep up with advancements and develop robust frameworks and guidelines for responsible AI use.

Overall, the thesis emphasizes the importance of integrating ethics into the development and deployment of AI technologies, particularly in recruitment. It urges organizations to prioritize fairness, transparency, and accountability in AI-driven recruitment practices. Stakeholders must work together to leverage AI in a way that respects individual rights, minimizes biases, and upholds fairness and equality, ensuring technology serves as a force for good in society.