Ca' Foscari
University
of Venice

Master's Degree
in Management

Final Thesis

Ethical implications of Artificial Intelligence: the relationship between algorithms and kindness

Supervisor
Ch. Prof. Giovanni Vaia

Graduand
Valentina Buttol
Matriculation Number 888522

Academic Year
2022 / 2023

Table of Contents

Introduction

In recent years, digital transformation has accelerated considerably and has increasingly impacted the lives of all of us. Against this backdrop, artificial intelligence emerges as one of the driving technologies of this transformation – Impacting society, the world of work and thus each individual – and is considered by many to be a revolutionary technology.

Its disruptive impact is kicking off a series of technological advances that are affecting all areas and generally a multitude of aspects related to human life and our daily activities. The revolution that artificial intelligence brings with it also implies great concerns about the possible negative consequences on humans and our future.

Although AI represents an extraordinary technology with high potential, its development, implementation, and use should be regulated and also well thought out in order to avoid the risks involved. The present thesis aims to make a critical and constructive contribution regarding especially the responsible and sustainable use of artificial intelligence, while also trying to shed light on the relationship between this technology and kindness, thus going beyond mere technology and trying to focus on a fundamental human value such as precisely kindness. Indeed, it is essential that in an increasingly digitized world, technologies are able to reflect our values. In this context, kindness represents a human and ethical value, but also an element that can lead to a more ethical and humane development of artificial intelligence systems.

In the first chapter, this thesis briefly traces the history of artificial intelligence, going through its different seasons, from spring to winter, and then to more recent developments. Next, the different types of AI, both capability-based and feature-based, are examined, and the definition framework proposed by the OECD is consequently presented. Key concepts are also introduced, including Machine Learning and Deep Learning, as well as other technologies behind artificial intelligence systems.

The second chapter considers ethical aspects and issues related to AI. Ethical issues that are still being confronted today are explored, such as bias and discrimination, privacy, transparency and accountability, safety, inequality and unemployment, and finally the impact on human and social relations. The ethical principles that should serve as the basis for the development of responsible AI and the legal framework concerning, both of which lack a single framework, are also introduced. Finally, possible strategies and tools to promote ethical AI are explained.

The third chapter introduces the concept of kindness, an ethical and human value, its meaning is defined, and it is explored in its declination as a social value, particularly in the context of the digital age in which we live and in the context of leadership, in which it succeeds in defining a new virtuous model.

In the fourth and final chapter we get to the heart of the proposed topic, which is the relationship between kindness and artificial intelligence, a particular combination that is not yet given enough attention, but one with high potential. The importance of kindness in the design of artificial intelligence systems and the impact that kindness can have on these systems is discussed. Finally, the challenges that arise in the context of combining this technology and the value of kindness are addressed.

Chapter 1. Artificial Intelligence: history and definitions

Artificial intelligence is one of the most impactful and disruptive digital technologies of recent years. Its deployment, which especially in recent years has been increasing dramatically, is having an impact not only at the economic and corporate level, but also at the level of society and culture, gradually leading to a revolution in the way we work, live, and interact with the rest of the world. The ability of this technology to process vast amounts of data and learn from it is creating new opportunities for growth and innovation, but it is also introducing new risks and challenges that certainly require strategic vision and collaboration among the various stakeholders.

As Artificial Intelligence continues to reconfigure the economic landscape, it is critical for businesses and decision makers to understand its drivers, dynamics, and implications, as well as develop effective strategies to leverage on its benefits and mitigate its drawbacks, also in terms of society.

Although it is only in recent decades that AI has made significant progress, the idea of creating a machine that can think and learn just like humans is not new, but dates to at least the 17th century.

This chapter will review the historical evolution of AI, starting from the beginnings of this technology to the present day, and then attempt to give a definition and understand how it works.

1.1 Brief history of Artificial Intelligence – seasons of AI

The origins of AI date back to the 17th century, specifically when René Descartes, a French philosopher and mathematician, considered the idea of creating a machine that would mimic the workings of the human mind, but even so, it was still some time before this idea became a reality, a result that was achieved with the advent of the modern computer. Descartes, famous for the phrase "I think, therefore I am," reflected more than 400 years ago on the possible reasoning capacity of machines. According to the French philosopher, in fact, machines had limited potential compared to the human mind, which can adapt to any instruction. Incorporating the workings of a human mind into a machine, therefore, would have made it possible to overcome its limitations. This reasoning of his served, in later centuries, as the basis for the famous Turing test[1].

1.1.1. The Spring of AI

The stages in the history of Artificial Intelligence are often traced back to the four seasons, and the birth of this discipline thus represents its Spring being marked by many breakthroughs.

According to many, the actual birth of the concept of artificial intelligence can be traced back to English mathematician Alan Turing, some 300 years after Descartes' idea, around the 1950s. In 1950, Turing published a paper entitled "Computing machinery and intelligence," the incipit of which read "Can machines think?" and described how a machine could actually be called intelligent. This was done by distinctly comparing the communication of a human and a computer with an evaluator placed in front of a terminal. If the judge failed to distinguish computer and human being, the machine would pass the test[2].

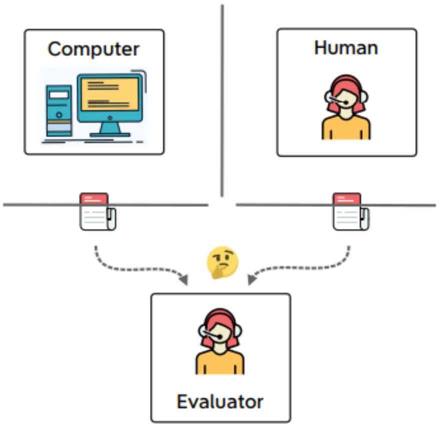

FIGURE 1.1 - Turing Test

SOURCE: A. Toosi-A. Bottino-B. Saboury-E. Siegel-A. Rahmim, 2021

[1] J. Gorman, A Brief History of AI, From French Philosophy to Self-Driving Cars, Dell, 2019 https://www.dell.com/en-us/perspectives/a-brief-history-of-ai-from-french-philosophy-to-self-driving-cars/

[2] M. T. Stecher, La storia dell'intelligenza artificiale, da Turing ad oggi, CyberLaws, 2018 https://www.cyberlaws.it/2018/la-storia-dellintelligenza-artificiale-da-turing-ad-oggi/

The methodology he described, is still known today as the "Turing Test" or "Imitation game" and is still used as a reference for detecting intelligence in an artificial system. In fact, Alan Turing developed a machine, called “Le Bombe”, which enabled the British government to crack the Enigma code, used by the German military to communicate during World War II, a result that made even the mathematician himself marvel[3].

Despite the significant scientific contribution Turing brought, some scholars trace the origin of this discipline back a few years earlier, namely to 1942, when Isaac Asimov, an American science fiction writer wrote the short story "Runaround." This story, despite not bringing any concrete contributions, inspired several scientists in the field of robotics in the following years, and was developed around the three laws of robotics. The first law states that a robot cannot injure or cause to be injured a human being, the second concerns the fact that a robot must obey orders that come to it from humans, unless those orders go against the first law, and finally the third determines that a robot must also protect its own existence, while still giving priority to the first two laws[4].

As for the term Artificial Intelligence, it was first used in the summer of 1956 by John McCarthy at the Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI) conference, which lasted approximately eight weeks and was attended by leading researchers in the field. This term was coined by Marvin Minsky, who defined Artificial Intelligence as "the science and engineering of making intelligent machines”[5]

1.1.2. AI Summer and Winter

Between 1940 and 1973 there was a period of increasing success in the emerging discipline of Artificial Intelligence. The Dartmouth Conference and World War II, accompanied by the fact that computers were becoming faster and more accessible, were in particular two events that stimulated increasing scientific research in this field, which led to remarkable achievements. One of the earliest achievements was the work of Newell and Simon, which was considered one of the first intelligent systems. It was called Logic

[3] M. Haenlein-A. Kaplan, A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence, in California Management Review, fasc. 61, 4, 2019, p. 5–14 https://doi.org/10.1177/0008125619864925

[4] M. Haenlein-A. Kaplan, cit.

[5] A. Toosi-A. Bottino-B. Saboury-E. Siegel-A. Rahmim, A brief history of AI: how to prevent another winter (a critical review), in PET Clinics, fasc. 16, 4, 2021, p. 449–469

Theorist (LT) and was an algorithm that was able to automatically prove mathematical theorems. Soon after, the same scientists realized the so-called General Problem Solver (GPS) which could manipulate objects within the room representation[6], inspired by the reasoning and problem-solving that characterizes human beings. The General Problem Solver is credited as the first project in the field of human reasoning regarding Artificial Intelligence. These years also saw the creation of the LISP programming language, which made it easier to write AI programs.[7]

A further example of success as far as AI is concerned is the computer program ELIZA. This project was created between 1964 and 1966 by Joseph Weizenbaum and can be called the first chatbot in history, thus developing a human-machine interaction. ELIZA, in fact, was to simulate the behavior of a therapist, creating the illusion of a conversation between human beings and thus trying to pass the Turing test[8].

Despite the many breakthroughs in the preceding years and the substantial funding that followed, in the decade between 1970 and 1980, we can identify a phase of disillusionment and arising of problems, comparable to the winter season[9].

As early as 1973, the first criticisms were made of AI research spending, and furthermore, the British mathematician Lighthill, through a report, questioned the rosy development expectations budgeted by scholars, leading to the discontinuation of most of the investments by the British government earmarked for various universities. The British government was not alone in cutting investment; in fact, the U.S. government also followed the British example[10]. This was also compounded by problems and difficulties related to the limited computing power and handling of large amounts of data [11].

After a period of stagnation, some progress was to be seen in 1980, but fell back into a second Winter in the early 1990s.

A pivotal moment in the history of AI, and one that opened a new season marked by development, was when the Deep Blue supercomputer created by IBM defeated world chess champion Garry Kasparov. Deep Blue thus became the first artificial intelligence

[6] M. T. Stecher, cit.
[7] A. Toosi-A. Bottino-B. Saboury-E. Siegel-A. Rahmim, cit.
[8] J. Gorman, cit.
[9] M. Haenlein-A. Kaplan, cit.
[10] M. Haenlein-A. Kaplan, cit.
[11] M. T. Stecher, cit.

system to beat a human. This, however, came about through advances based purely on computing power, and not through artificial "thinking." In 1997 the first speech recognition software implemented on Windows was also developed, and at the same time Kismet, a robot capable of demonstrating and recognizing human emotions, was created, developed by Cynthia Breazeal[12].

1.1.3. AI developments in recent years

A fundamental problem in the development of artificial intelligence is the limit imposed by the storage capacity of computers, it follows that the development of this discipline should go hand in hand with the current computational power, that is, the speed and storage capacity of computers. In 1965 Gordon Moore, stated that we can expect the speed and capability of our computers to increase every two years due to the fact that every two years the number of transistors, i.e., the active components of microchips, should double[13]. Today, in fact, we no longer have the limitations that held us back in past decades, and we can say that computational capacity is meeting, and sometimes exceeding, our needs, leading to exponential growth[14]. Nowadays, moreover, we can also rely on a massive collection of data; in fact, we live in the era of so-called big data, thanks to which artificial intelligence continues to grow steadily and thanks to which better algorithms have been developed[15]. Technological advances have also led to famous successes such as IBM-developed Watson, which in 2011 won against the champions of the game show Jeopardy!; or Google X, which in 2012 was able to recognize cats in a video; or AlphaGo, Google's artificial intelligence that in 2016 beat the world champion of the game Go. This all came about partly by leaving computers to discover correlations and classifications on their own, after providing them with a large amount of data[16].

Regarding the development of Artificial Intelligence, in terms of performance, the last two decades can be summarized in the following graph. On the x-axis we find the reference years, while on the y-axis the human performance is taken as reference 0. This means that

[12] A. Rockwell, The History of Artificial Intelligence, Science in the News, 2017 https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

[13] A. Rockwell, cit.

[14] A. Rockwell, cit.

[15] A. Toosi-A. Bottino-B. Saboury-E. Siegel-A. Rahmim, cit.

[16] Council of Europe, History of Artificial Intelligence - Artificial Intelligence - www.coe.int, Artificial Intelligence, s.d. https://www.coe.int/en/web/artificial-intelligence/history-of-ai

when performance exceeds zero, the Artificial Intelligence was able to score higher on a test than the humans who performed the same test. In addition, the various lines represent the five domains: handwriting recognition, speech recognition, image recognition, reading comprehension, and language understanding[17].

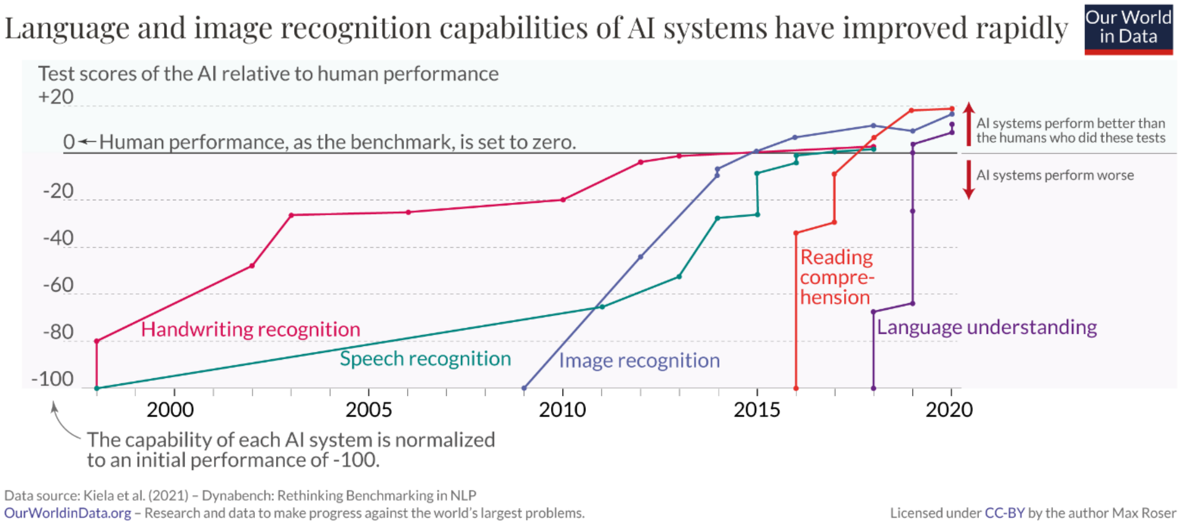

FIGURE 1.2 - Growth in AI systems capabilities

SOURCE: L. Tremolada, 2022ù

The trend of all domains is clear, especially in recent years, artificial intelligence systems have evolved significantly and manage to beat humans in tests of all domains. The results that can be seen from this graph, however, are related to specific tests, this in fact does not mean that AI has outperformed humans or that it is a sentient system, we need to be aware of this before drawing rash conclusions[18].

The enormous capabilities of this technology are leading to unprecedented innovations and developments in all fields, but its ability to process huge amounts of data and its autonomy can lead to discrimination and decisions that may turn out to be wrong and risky. It is precisely for this reason that in recent years more attention has been paid to the

[17] L. Tremolada, Blog | La storia dell'intelligenza artificiale in due grafici, Info Data, 2022 https://www.infodata.ilsole24ore.com/2022/12/24/la-storia-dellintelligenza-artificiale-in-due-grafici/

[18] L. Tremolada, cit.

ethical challenges that Artificial Intelligence brings, which should also be addressed through appropriate regulation.

1.2. The different types of AI

Artificial intelligence, defined by the European Commission Communication in 2018 as "systems that display intelligent behavior by analyzing their environment and taking action-with some degree of autonomy-to achieve specific goals [19]", can be categorized in several ways. The two most famous and popular classifications are the classification based on AI capabilities, and the categorization based on the capabilities offered by machines.

1.2.1. AI Types based on capabilities

The first proposed classification distinguishes types of AI based on their capabilities. Using this categorization, three stages are defined, starting from very limited machine capability to machines that can far exceed human intellect and are endowed with self-awareness.

The first of the three stages of artificial intelligence is called Narrow AI, Weak AI, or more briefly ANI. This stage includes machine learning and includes technologies such as the virtual assistants Siri, Alexa and Cortana[20]. It specializes in a specific area and solves certain problems, so it has a limited capability that is also not transferable to other systems. Other examples that fall into this category may be self-driving cars, voice and image recognition, and Google itself[21]. This category includes all the major artificial intelligence-related technologies we have in use today.

The second stage is called general AI, strong AI or AGI. It is referred to as the stage of machine intelligence because machines are characterized by an intellect on an equal level with humans, also possessing the ability to surpass it. At this stage, machines are able to extrapolate concepts from limited experience and are able to transfer this knowledge of

[19] European Parliament. Directorate General for Parliamentary Research Services., Artificial intelligence: how does it work, why does it matter, and what we can do about it?, Publications Office, LU, 2020

[20] O. Strelkova, Three types of artificial intelligence, 2017

[21] D. Petersson, 4 Main Types of Artificial Intelligence: Explained, Enterprise AI, 2023. https://www.techtarget.com/searchenterpriseai/tip/4-main-types-of-AI-explained

theirs to other domains. Thus, this stage represents a machine that is considered rational and is able to solve problems posed independently and is able to learn on its own, and this is precisely what research is being done for in recent years[22].

The third and final stage of artificial intelligence is the so-called ASI, or super AI, whose main trait is machine consciousness. Precisely because of this trait, we can say that it is a stage that we have not yet reached with current technology, and it is a situation that many fear, as it would be represented by a machine far more intelligent than even the best human brains, thus exponentially surpassing them in every field and acquiring self-awareness. It represents the highest possible level of artificial intelligence and to this day remains a futuristic hypothesis, although according to many it poses a great risk, often called singularity.[23] The term singularity was coined in 1993 by Vernor Vinge, a science fiction author, and was intended to represent a point of no return as far as human history is concerned. This would pose a risk to humans in that the consequences of technological breakthroughs would no longer be predictable and controllable[24]. However, to date the singularity conceptualized by Vernor Vinge remains a distant hypothesis.

1.2.2. AI types based on functionalities

An alternative classification to the one proposed above is the classification of types of artificial intelligence based on a functionality criterion.

Proceeding in order of increasing potential, the first type of AI according to this subdivision are Reactive Machines. This type of artificial intelligence possesses a very limited capability; they do not rely on learning or memory but merely produce an output based on the input provided[25]. This category includes early AI algorithms, many machine learning models, and even IBM's famous Deep Blue that beat chess champion Kasparov in 1997[26].

[22] N. Joshi, 7 Types Of Artificial Intelligence, Forbes, 2019. https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/

[23] N. Joshi, cit.

[24] V. Vinge, TECHNOLOGICAL SINGULARITY, 1993. https://frc.ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html

[25] D. Petersson, cit.

[26] N. Joshi, cit.

The second type in this subdivision is Limited Memory Machines. Limited Memory Machines are more advanced machines than reactive ones and are based on the ability to learn. In fact, they learn from historical data sets and provide predictions that gradually become more and more accurate. Deep learning, in fact, gets smarter and smarter as it is given more data, which gives it a better information picture. Nowadays, many applications rely on just this specific functionality[27]. For example, image recognition is also based on this logic, the more the AI scans a large number of images, the more accurate the recognition will be, developing real learning that leads to increasing accuracy[28]. Other examples include virtual assistants and self-driving cars, vehicles that operate through a combination of computer vision and observational knowledge.

The third type of AI described by feature classification is Theory of Mind. In the Theory of Mind, a term derived from psychology, machines interact with the emotions and thoughts of humans and attribute mental states to the other entities they interface with, so for example, emotions and feelings[29]. To date, it is a stage of AI that is being worked on and is an important research goal, but we do not yet have machines that demonstrate comparable capabilities to humans, one above all emotional intelligence, although some may prove similar during a conversation[30]. Nevertheless, steps toward this stage of AI evolution have also been taken in the past. Just think of the robot developed by Cynthia Breazeal, called Kismet, which, as mentioned earlier, was able to recognize on people's faces their emotions and replicate them; or the robot developed by Hong Kong-based Hanson Robotic, the humanoid Sophia, which is able to recognize faces and respond with its own facial expressions[31].

The last type of AI described by this classification is self-aware AI. Self-aware AI represents the last stage that machines will be able to reach, the most advanced, in that it is described as a situation in which AI will possess an independent intelligence, considered on par with human beings and thus demonstrating the same emotions and

[27] N. Joshi, cit.

[28] N. Joshi, cit.

[29] D. Petersson, cit.

[30] B. Marr, Understanding the 4 Types of Artificial intelligence, Bernard Marr, 2021 https://bernardmarr.com/understanding-the-4-types-of-artificial-intelligence/

[31] B. Marr, cit.

needs[32]. It represents a still distant scenario as we do not currently possess the hardware or algorithms that would be able to enable it [33].

1.2.3. The OECD Framework

An alternative classification to the two previously analyzed is the one proposed by the OECD. The OECD, in fact, given the significant impact of AI on both the economy and society at large and the resulting complexity, has decided to create a framework for classifying AI systems, aimed at ensuring international standards and assessing their challenges. It is thus a framework that helps regulators, legislators and others assess not only the opportunities but also the risks presented by a given AI system[34]. The proposed distinction is based on the potential impact of this technology on society, individuals, and planet and aims to link the characteristics of AI systems to the OECD AI Principles, which will be referred to in the second chapter and which also take into account ethical challenges.

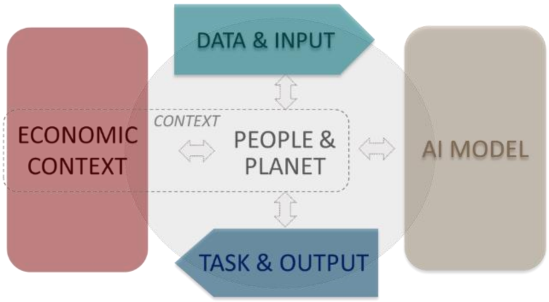

FIGURE 1.3 - OECD Framework for the classification of AI systems

SOURCE: OECD

The main dimensions of the framework, as shown in the figure, are.:

- People and Planet, a dimension that takes into account users, stakeholders, optionality, human rights, well-being and environment;
- Economic Context, which covers industry, business function and model, criticality and finally scale and maturity;

[32] D. Petersson, cit.
[33] B. Marr, cit
[34] OECD, OECD Framework for the Classification of AI Systems: a tool for effective AI policies, s.d. https://oecd.ai/en/p/classification

- Data and Input, formed by the subdimensions collection and rights and identifiability;
- AI model, a dimension that takes into account model characteristics, model-building and model inference;
- Task and Output, composed by tasks, action and application area[35].

This classification is certainly more complex and perhaps less tied to the mere technicalities that enabled the previous two classifications, but it can be considered more comprehensive and at the same time complex, as it looks not only at AI systems per se, but also at the impact they may have, highlighting the challenges associated with this disruptive technology[36].

1.3. Machine Learning and Deep Learning

When people talk about artificial intelligence, they often refer to the terms Machine Learning and Deep Learning. These two, in fact, are fields of study that fall under the broader discipline of AI, which generally represents "the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem solving”[37].

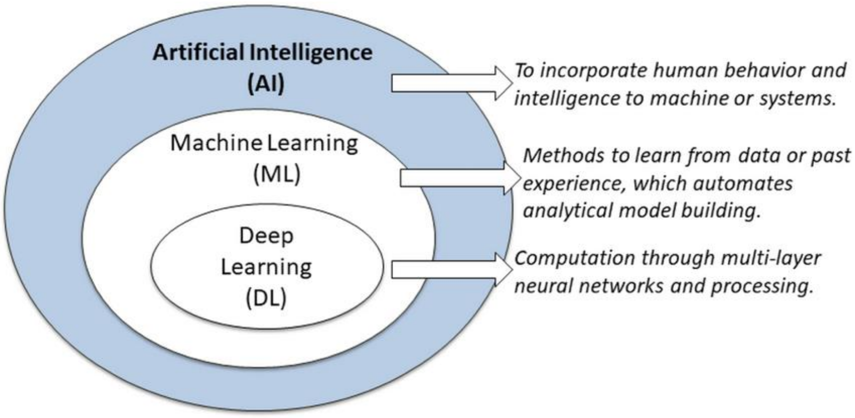
Artificial Intelligence
(AI)
To incorporate human behavior and intelligence to machine or systems.
Machine Learning
(ML)
Methods to learn from data or past experience, which automates analytical model building.
Deep
Learning
(DL)
Computation through multi-layer neural networks and processing.

FIGURE 1.4 – The relationship between AI, ML and DL

SOURCE: SN Computer Science

[35] OECD, cit.
[36] OECD, cit.
[37] M. Chui-V. Kamalnath-B. McCarthy, An executive’s guide to AI - McKinsey, 2020 http://ceros.mckinsey.com/quick-guide-to-ai-12

Machine Learning refers to the part of AI that uses mathematical models, subjected to a training process, which are aimed at obtaining a result within a specific task. Because of this very characteristic, Machine Learning is considered Weak AI. Machine Learning thus uses algorithms to make machines learn, which will obtain a result in a specific domain, and if the result is considered incorrect there is a need for human intervention[38]. The three types of analytics that a Machine Learning model can provide are descriptive, meaning they describe the situation, or predictive and prescriptive, which are the focus of this technology. With predictive data, it is possible to anticipate what will happen in the future, a valuable source of insight for companies, for example, while prescriptive analytics can provide recommendations on what to do to achieve established goals[39].

The training these models undergo can be distinguished into supervised learning, unsupervised learning, and reinforcement learning[40].

Unsupervised learning involves providing input data to models, which will then provide output. The input data are not labeled, and this is the major difference between supervised and unsupervised learning. Algorithms thus independently search for hidden patterns, which is why they are called "unsupervised". They are used primarily for three activities[41]:

- Clustering, a technique that involves grouping data without the presence of labels, pooling based on similarities or differences. It is usually used for image compression and market segmentation[42].
- Association, a technique that uses a different method from clustering and is usually used for suggestions that can be found on e-commerce or search engines [43].
- Dimensionality reduction is a technique that is used when the number of dimensions in a dataset is too large; it goes in fact to reduce the input data to allow for analysis and preserving the integrity of the data[44].

[38] E. Kavlakoglu, AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?, 2022 https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

[39] M. Chui-V. Kamalnath-B. McCarthy, cit.

[40] M. Chui-V. Kamalnath-B. McCarthy, cit.

[41] J. Delua, Supervised vs. Unsupervised Learning: What's the Difference?, 2022 https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning

[42] J. Delua, cit.

[43] J. Delua, cit.

[44] J. Delua, cit.

Instead, supervised learning is said to occur when, in addition to the input data, the model is provided with examples of the desired output, i.e., the data are labeled, for example. The algorithm thus fits the correct answer, and these models are considered more accurate than unsupervised models; however, they require more initial effort corresponding to labeling the data[45]. Supervised learning is also categorized into two types of data problems: classification and regression. Classification problems make use of an algorithm that can classify and divide data into specific categories; for example, they can be used to divide spam from incoming e-mail. Regression, on the other hand, uses an algorithm to distinguish relationships between dependent and independent variables[46].

The third type of algorithm learning is reinforcement learning. Reinforcement learning involves the algorithm learning to perform a task by looking for ways to maximize the payoff it receives for its actions, meaning it learns by trial and error. It is characterized by sparse training data and an ideal end state that is not clearly delineated, alternatively the only way to know and learn from the environment is by interacting with it. Its applications range from optimizing the driving of self-driving vehicles, to optimizing a trading strategy for a portfolio.[47]

A subcategory of Machine Learning is Deep Learning, a field that refers to the concept of artificial neuron. Deep Learning refers, in fact, to the study of neural networks, which are formed by artificial neurons. Artificial neurons, which together form a neural network, are placed in a particular layer of the network, and connect with other neurons in the layers that precede or succeed it. This communication takes place through signals that are represented by mathematical operations. The set of all the signals exchanged and computed represents the learning of the neural network, which gradually arrives at the result that was requested of it. As in the case of Machine Learning, Deep Learning learning can take place in a supervised manner if provided with an example of output, or unsupervised if only input is provided[48].

Deep Learning is so called because of the characteristic of being applied to artificial neural networks with a large number of layers, which are endowed with considerable depth. It is

[45] J. Delua, cit.

[46] J. Delua, cit.

[47] M. Chui-V. Kamalnath-B. McCarthy, cit.

[48] I. Sarker, AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems, in SN Computer Science, fasc. 3, 2022

precisely this depth that allows increasingly detailed representations to be obtained, leading to a very precise solution to the problem that was initially posed and without requiring human intervention. Due to the greater depth, the performance found, as highlighted in the figure, is greater than that of other ML algorithms, especially as the amount of data available increases[49]. Specifically, Deep Learning can be defined as an algorithm that possesses a neural network with more than three layers [50].

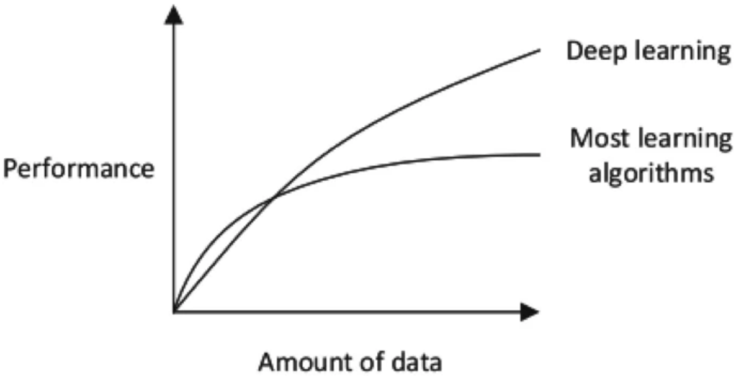

FIGURE 1.5 – Performance of ML and DL compared
SOURCE: SN Computer Science

Neural networks, or artificial neural networks (ANNs), are specifically composed of four basic components (namely input, weights, bias or threshold, and output) and seek to mimic through algorithms the functioning of a human brain. They are composed of three main types of layers: input layers, hidden layers, and output layers. The layer of neural networks is composed of the inputs, and once this layer is formed weights are assigned to indicate the importance of a given variable. The input is then processed through a function that will then produce an output, which will be passed to the next layers of the network. In fact, the hidden layers are responsible for processing information derived from the previous layers and attempt to define patterns and relationships. Finally, output layers are responsible for producing the final output[51].

[49] I. Sarker, cit.
[50] E. Kavlakoglu, cit.
[51] E. Kavlakoglu, cit.

The concept of a neural network is not of recent origin, but rather dates back to 1944, when it was proposed by two researchers at the University of Chicago-Warren McCullough and Walter Pitts. The neural networks that were described by the two researchers were not those used today; they had thresholds and weights but no mention of training and were not divided into layers[52]. Later, in 1957, the first trainable neural network, Perceptron, was demonstrated by psychologist Frank Rosenblatt, which was very close to today's ANNs. The only difference was that there was only one layer between the input and output layers[53].

Machine Learning, and Deep Learning in particular, are achieving good results in several areas, and are mainly used for speech and image recognition, but nevertheless require a large amount of training data and high computing power in order to train artificial neural networks, facts that can make them expensive and difficult to use. An additional challenge that arises with Deep Learning is also the overfitting of the model, which is considered "overlearning", as it represents the case where the algorithm learns so well about the data it is trained on that it leads to the inability to generalize new data[54].

1.4. Other technologies behind AI

Artificial intelligence represents a very broad and interdisciplinary field of study, consisting of multiple technologies that represent and compose it. The two most discussed are usually Machine Learning and Deep Learning, which have already been described above, but there are others that are equally important and impactful. The following areas are only a part of what lies behind artificial intelligence but represent the most famous technologies used to develop applications of AI in different fields and sectors.

Natural Language Processing (NLP) is a branch of artificial intelligence that aims to understand and process human language; in fact, it deals with its analysis and manipulation and should help overcoming the difference between computer

[52] L. Hardesty, Explained: Neural networks, MIT News | Massachusetts Institute of Technology, 2017 https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

[53] L. Hardesty, cit.

[54] R. Schmelzer, Are We Overly Infatuated With Deep Learning?, Forbes, 2019 https://www.forbes.com/sites/cognitiveworld/2019/12/26/are-we-overly-infatuated-with-deep-learning/

understanding and human communication[55]. It is considered an interdisciplinary research field as it combines computer science, linguistics and artificial intelligence and has evolved considerably in recent years due to the large amounts of data available and the proliferation of new techniques regarding Deep Learning. The goal of this technology is to understand language as humans do, humans who use their brains and five senses to process input; similarly, computers have their own programs and algorithms that are supposed to mimic this processing and understanding functionality. Computer understanding of language is based on the meaning of words and their use within a context, thus their appropriateness, but also on rules of sentence structuring and grammar[56]. NLP also allows computers, not only to read text, but also to listen to speech and then interpret it and measure the sentiment [57].

The importance of NLP lies in several factors and capabilities that this technology possesses. The first lies in its ability to analyze and process large amounts of textual data and then consequently extract insights from it that are useful, for example, for business decisions. Insights are typically derived from unstructured textual data, which must be "understood" by the machine[58]. The second capability lies in being able to give structure to a highly unstructured data source. Referring to human language, in fact, there are many different languages and dialects, but each of these is also characterized by different rules and idioms. In addition to this, there is the fact that within textual messages we can find both errors, whether they are of syntax or punctuation, and abbreviations, and these same nuances are also found in the spoken language that the machine has to process. The NLP's ability in this case comes from being able to resolve ambiguities with respect to the input data and being able to give it a structure[59]. To accomplish this, NLP is composed of several tasks, including: speech recognition, or text to speech; part-of-speech labeling, or grammatical labeling; word sense disambiguation, that is, selecting the correct meaning the word takes on in its context; named entity recognition; co-reference resolution; sentiment analysis, which seeks to extrapolate subjective qualities from text such as

[55] SAS, Cos'è il Natural Language Processing (NLP)? https://www.sas.com/it_it/insights/analytics/what-is-natural-language-processing-nlp.html

[56] M. Esposito, Natural language processing (NLP), che cos'è, i progressi e le sfide dell'AI, Agenda Digitale, 2019 https://www.agendadigitale.eu/cultura-digitale/linguaggio-naturale-e-intelligenza-artificiale-a-che-punto-siamo/

[57] SAS, Natural Language Processing (NLP): What it is and why it matters https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

[58] SAS, cit.

[59] SAS, cit

emotions; natural language generation, usually identified as the opposite of speech recognition, that is, inserting structured information into human language[60].

NLP also has a role in human-computer interaction, a role that in fact can be appreciated through the use of increasingly advanced and widespread chatbots and virtual assistants. Many of the activities we perform on a daily basis in fact rely on NLP, one of its main uses being virtual assistance, through the famous Siri and Alexa, for example, which perform everyday tasks for us, including planning events, playing music, and searching for information. In these cases, NLP allows the machine to interpret the voice commands provided and give the user an appropriate response. Another common application of NLP are chatbots or virtual assistants, which we can observe on many websites and through which assistance is provided. Other examples also consist of spam detection in email inboxes, machine translation such as through Google Translate, text summaries and sentiment analysis on social media[61].

Computer vision is another subfield of artificial intelligence, analogous to Natural Language Processing, which, however, instead of dealing with human language in the form of text and words, deals with understanding and interpreting images and videos, thus trying to mimic the human eye. Videos, images, and graphics in fact constitute a rich source of data that can be analyzed, and computer vision, again through algorithms, deals with recognizing patterns in them. Through this technology, machines are able to analyze both images and videos, but they have one disadvantage compared to humans. Computers, in fact, do not have vision that has been trained for years and years, a quality that distinguishes humans in this case[62]. Computer vision, however, when trained with a large amount of data, is able to develop systems that can inspect a very large amount of processes or products in a very short time, thus exceeding human capabilities. To make this happen, however, it is necessary to feed the computer with a large amount of images or video referring to the object it will have to analyze[63].

Until recently, the capabilities of this technology were limited, but now, thanks to advances in recent years in neural networks and deep learning, these capabilities have grown

[60] IBM, What is Natural Language Processing? | IBM. https://www.ibm.com/topics/natural-language-processing

[61] IBM, cit.

[62] IBM, What is Computer Vision? | IBM. https://www.ibm.com/topics/computer-vision

[63] IBM, cit.

exponentially, so much so that they sometimes surpass humans in some tasks.[64] This growth is also due to other factors, such as mobile technology with the integration of cameras that have enabled the generation of large volumes of photos and videos, more easily accessible computing power that is also affordable, new algorithms, and finally the high availability of the hardware that is needed for computer vision and its analysis[65].

Computer vision is used in all industries and is often used to improve user experience and increase security levels [66]. Among its most common uses there are:

- facial recognition, however, which often raises ethical and privacy concerns especially when affiliated with profiling and surveillance purposes;
- autonomous vehicle navigation, which requires understanding of the surrounding environment and processes images in real time;
- mixed or augmented reality, which is used not only by various applications found on smartphones, but is essential in many areas such as training;
- the use of this technology in health care, which is performing well in detecting disease in MRI or X-ray images and scans and helping doctors produce accurate diagnoses[67];
- the use of computer vision in sports, here computer vision is used to produce analysis and strategies;
- content organization, in which case both people and objects can be identified and organized within the photos.[68]

Moreover, the operation of this technology is specifically divided into several tasks with reference to an object within an image, namely: classification, identification, verification, detection, landmark detection, segmentation, and finally recognition[69].

Another branch of artificial intelligence is Generative AI, which has come into the limelight in recent months thanks in part to the success of ChatGPT. Generative AI is a technology

[64] I. Mihajlovic, Everything You Ever Wanted To Know About Computer Vision. Here's A Look Why It's So Awesome., Medium, 2021 https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e

[65] SAS, Computer Vision: What it is and why it matters. https://www.sas.com/en_us/insights/analytics/computer-vision.html

[66] SAS, cit.

[67] I. Mihajlovic, cit.

[68] Microsoft, What Is Computer Vision? | Microsoft Azure. https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-computer-vision/

[69] I. Mihajlovic, cit.

through which machines are able to generate sounds, images, data, text and other types of information autonomously through machine learning algorithms. Machine learning in Generative AI is done mainly through techniques such as generative adversarial networks, called GANs, and natural language models, which are used to generate information that appears to have been created by humans. Using GANs, machines are able to generate new text, images, and videos from existing input data. The operation of these algorithms is given by two neural networks, which respectively generate and evaluate data to continuously improve themselves and are called generator and discriminator[70].

Generative AI has multiple applications and is used in various fields. In fact, thanks to this technology, videos and images can be created for movies and entertainment. It can be used to generate product models and designs, but also to create generative art, so music, novels or artwork. Applications of this type have several uses, for example, chatbots can be created, artistic and musical creations can be created, they can generate codes for programming, and they can also be used to enhance personalized education[71]. The outputs generated can be numerous, the most common being:

- Images, new images can indeed be created from existing ones or through textual description. Some examples are DALL-E, MidJourney and Stable Diffusion.
- - Texts, the main example of which is given by ChatGPT. ChatGPT is the Generative AI model developed by OpenAI, which is based on the GPT architecture, or Generative Pretrained Transformer. It was launched in November 2022 and in just five days was able to attract more than one million users[72].
- - Audio, generative artificial intelligence can in fact create new music tracks and even perform dubbing.

This new type of Artificial Intelligence has varied capabilities that can be summarized into three types: idea and content generation, improving efficiency by performing repetitive tasks, and finally creating personalized experiences for users. These abilities translate into

[70] N. Routley, What is generative AI? An AI explains, World Economic Forum, 2023 https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/

[71] I. Vrtaric, History of Generative AI. Paper explained., Artificialis, 2023 https://medium.com/artificialis/history-of-generative-ai-paper-explained-6a0edda1b909

[72] Boston Consulting Group, Generative AI, BCG Global. https://www.bcg.com/x/artificial-intelligence/generative-ai

a multitude of tasks, such as marketing text generation, text summaries, document search, performing data entry, analyzing even large amounts of data, finding bugs in code, and many others. Indeed, we are currently only at the beginning of the adoption of Generative AI, and it is estimated that by 2025 this technology will reach about 30 percent of the AI market, thus representing a slice of it estimated to be around $60 billion.[73]

In recent months this technology has been in the spotlight, a spotlight that, however, has also brought with it many concerns about ethical and even privacy issues. Indeed, being able to generate content artificially brings with it several problems, including the dangerous creation of deepfakes but also the creation in general of false and misleading content. It must therefore be used with caution to limit its possible negative consequences.

Many have greeted the deployment of this technology with fear, this is given by the fact that having a conversation with ChatGPT gives one the feeling of dealing with a sentient machine, but these models certainly do not possess a consciousness like humans, and their conversations and responses that may seem realistic but at the same time creative, are the result of models trained on a huge amount of data[74].

A further cause for concern for some is the fear that technologies such as this will take jobs away from different categories of people, leading them to unemployment. This is a valid observation in this case, but it does not take into account the benefits that such a technology can have on the economy and society, or even the jobs that will be created by its use. In general, artificial intelligence will certainly help workers by automating repetitive tasks and analyzing huge amounts of data and then leading to informed and fast decisions. The genuine value of human activities, which is unlikely to be replaced by machines, lies in the added value they bring[75].

[73] Boston Consulting Group, cit.

[74] J. Fruhlinger, What is generative AI? The evolution of artificial intelligence, InfoWorld, 2023 https://www.infoworld.com/article/3689973/what-is-generative-ai-the-evolution-of-artificial-intelligence.html

[75] N. Routley, cit.

Chapter 2. AI and Ethical implications

In recent years, the impact that artificial intelligence has had on our daily lives has been remarkable. Indeed, this technology is being used in a variety of sectors, including finance, healthcare, logistics and public services, where it aims to improve the effectiveness, efficiency and quality of both services and processes. AI has more generally also revolutionized the approach used for problem solving, allowing large amounts of data to be precisely analyzed and enabling processes and decisions to be automated. AI is also referred to as one of the main drivers of the fourth industrial revolution, an industrial revolution, however, that appears different from previous ones, characterized by great promise due to the new technologies developed, but also by great dangers. On the other hand, however, like the three that preceded it, it will permanently alter our future and our relationship with technology.

However, the use of AI is raising ethical and moral concerns that require special attention. Indeed, the issue related to the ethics of AI is a key research field in assessing the ethical consequences related to this technology, particularly the consequences derived from decisions involving AI. This field of research is therefore concerned with both analyzing such consequences and developing ethical guidelines and principles that succeed in ensuring their sustainable and responsible use.

In general, AI can be seen as a technology that exhibits exponential growth, and because of this trait of its some worry about a future scenario that could lead to the singularity. In this regard, many scholars have argued in favor of a government regarding this technology, and letters such as the 2015 one signed by the likes of Stephen Hawking and Elon Musk attest to this. This document has been referred to as an open letter for the future of AI, titled "Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter," outlines research priorities regarding this technology, particularly with regard to preventing potential negative scenarios impacting society so as to ensure that this technology is beneficial to society as a whole[76]. Also in 2015 was the "Open Letter on Autonomous Weapons," a letter signed by more than a thousand experts in artificial intelligence and robotics, which raised concerns about the possibility of the use of

[76] S. Russell-D. Dewey-M. Tegmark, Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter, Future of Life Institute, 2015. https://futureoflife.org/open-letter/ai-open-letter/

autonomous weapons-that is, weapons that can decide to attack without the need for human intervention-and called for an international ban on them[77]. Ethical concerns are thus relevant and pervasive even within the scientific community, and they continue to grow. In fact, it was recently written, notably in March 2023, that the letter "Pause Giant AI Experiments: An Open Letter”[78]. This open letter, like previous ones, expresses concerns regarding artificial intelligence, particularly in this case about its uncontrolled development. The appeal made through this letter calls for a pause of at least six months regarding AI systems more powerful than GPT-4, a request that, according to the signatories, is based on the fact that a kind of out-of-control race is taking place regarding the development of this technology, a race that is leading to the development of systems that even the creators themselves cannot understand and control.

In this context, this chapter focuses on ethical issues concerning AI, examining the ethical principles that can lead to responsible use of this technology so as to succeed in achieving trustworthy AI, the legal framework, and the main issues and challenges concerning this field. Possible solutions and tools to best address these ethical issues are also explored, thus providing an up-to-date overview of the opportunities but also the challenges related to AI ethics.

2.1. Current ethical challenges

The technological revolution brought by artificial intelligence is raising quite a few ethical issues, as witnessed by the numerous open letters signed by thousands of scholars and professionals in the field, issues that need to be addressed so that a sustainable future can be ensured. If there are those who call for a 6-month pause regarding the training of intelligent systems more advanced than GPT-4, there are even those who think that this measure and the related letter, underestimate the real dangers of this technology. This is the case of Eliezer Yudkowsky, an artificial intelligence and decision theorist, who believes that the only way to stop an irreparable and catastrophic decline is to shut down all large CPU clusters, without exception, and to impose limits regarding the computing power that

[77] S. Gibbs, Musk, Wozniak and Hawking urge ban on warfare AI and autonomous weapons, in The Guardian, 2015

[78] Future of Life Institute, Pause Giant AI Experiments: An Open Letter, Future of Life Institute, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

can be used to train machines, aiming for a reduction in this power over time[79]. Although Yudkowsky's opinion sounds as extreme and apocalyptic, the U.S. theorist states that we are not yet ready for responsible use of this technology, which is partly true. Indeed, a unified government or uniform rules would be needed especially with regard to the responsible use of artificial intelligence. While scientific developments are indeed making great strides in recent years, legislation in this area must adapt quickly.

Ethical controversies and incidents regarding AI are a significant and current challenge. In this regard, an independent, open, and public dataset regarding incidents and controversies related to artificial intelligence to algorithms was established in 2019: the AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC) Repository. It appears from the AIAAIC's database that the number of disputes increased by as much as 26 times in 2021 compared to 2012, as seen in Figure.[80] This growing number is the result of both the increase of these technologies in the world but also of an increased awareness of the ethical issues faced when discussing AI[81].

[79] E. Yudkowsky, The Open Letter on AI Doesn't Go Far Enough, Time, 2023 https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/

[80] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, AI Index Report 2023 – Artificial Intelligence Index, Institute for Human-Centered AI, Stanford University, 2023

[81] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault

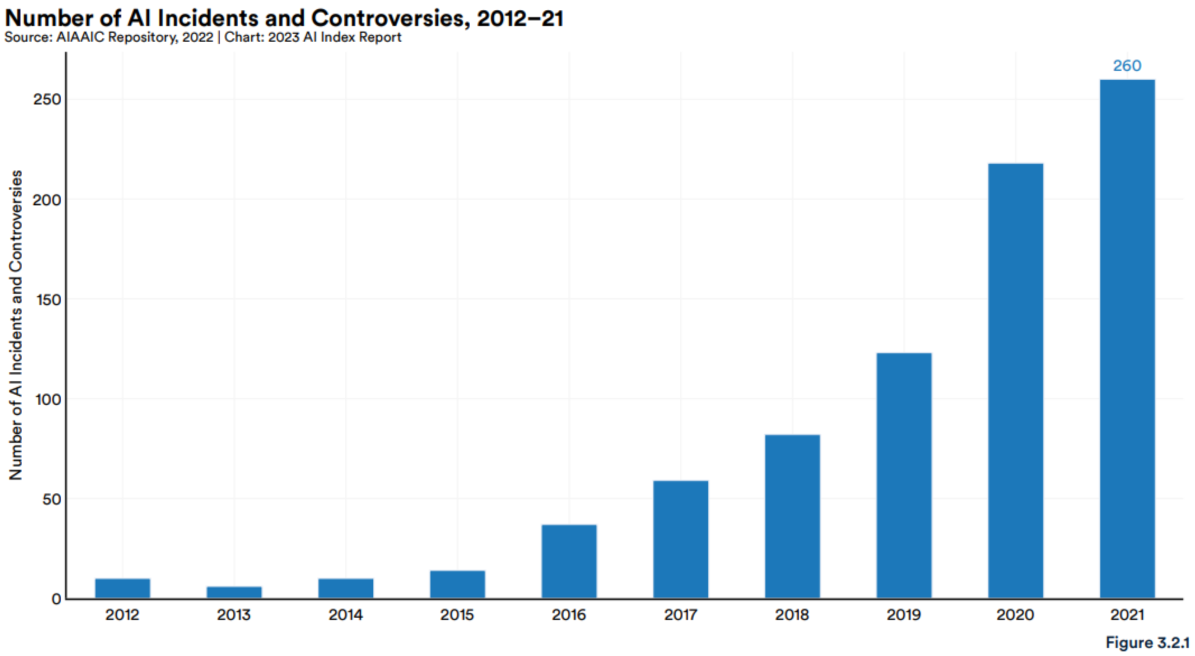

FIGURE 2.1 – Number of AI Incidents and Controversies, 2012-2021

SOURCE: 2023 AI Index Report

The first step in responsible governance of this technology is surely awareness of the various ethical challenges it faces, as it can lead to unreliable results and consequences contrary to the ethical and moral standards that should be proper to a healthy society. These challenges are numerous and complex and refer to the use and adoption of this technology. The following will detail the main ones.

2.1.1. Bias and discrimination

Artificial intelligence, based on the data processed and used for learning, can be a dangerous vehicle for the spread of bias and discrimination present in society. The quality of training data is central here to ensure that decisions made are unbiased and based on representative data that can guarantee fair choices. Data containing bias and prejudice can lead to discriminatory and unfair decisions in a variety of areas, such as health care, education and employment, there is therefore a need to have training data free of any bias.

One of the most famous cases in this context is the ImageNet face database, which contained a majority of white faces. By training artificial intelligence algorithms on this database for face recognition, it will then turn out that the algorithm will not perform as well on nonwhite faces as on white faces, representing the minority. Also within this dataset, it is clear that the representation of images does not correspond to the real slice of the corresponding world population, and the 14 million images labeled in ImageNet are predominantly from a few states, suffice it to say that the United States represents only 4 percent of the world's people, a far cry from the 45 percent represented[82]. In addition to this, for example, China and India, which account for 36 percent of the world's population, within ImageNet have a representation that corresponds to 3 percent of the total data [83]. This creates an inherent bias in the algorithm that can have a great discriminatory impact.

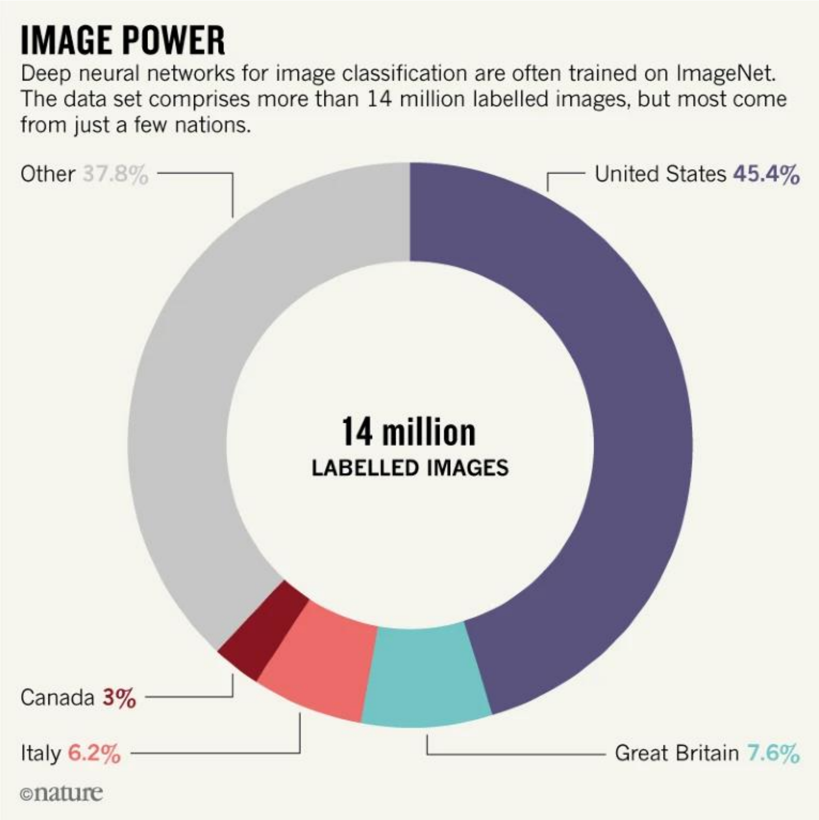

FIGURE 2.2 – ImageNet representation of nationalities

SOURCE: Nature

[82] J. Zou-L. Schiebinger, AI can be sexist and racist — it's time to make it fair, in Nature, fasc. 559, 7714, 2018, p. 324–326

[83] J. Zou-L. Schiebinger, cit.

One of the academics to shed light on this type of bias, both gender and racial, is Joy Buolamwini, a researcher at the Massachusetts Institute of Technology and founder of the Algorithmic Justice League. Buolamwini discovered, within artificial intelligence systems belonging to companies such as IBM, Microsoft and Amazon, the presence of widespread biases, both racial and gender[84]. Her famous study focused on facial recognition biases, biases of which she herself was a victim. In fact, she discovered this flaw in the algorithms precisely by finding that the facial recognition software she was using for the Aspire Mirror project did not recognize her face, which it did if she put on a white mask[85]. Boulamwini thus realized that for the algorithm, a black woman's face could not be analyzed and therefore did not exist; this happens because the machine was trained on a dataset containing many more white and male faces than in other categories. After this realization, his project changes and becomes Gender Shades, which focuses on the accuracy of products that work on the basis of artificial intelligence and are supposed to classify the gender of people[86]. The results of this project confirmed what Boulamwini had already experienced on his own skin, the systems analyzed in the study, in fact, finding that the best results were achieved in lighter-skinned and male subjects, while the worst results were found for black women[87].

Just as a result of her discovery, the researcher founded the Algorithmic Justic League, an organization that aims to bring about a responsible, fair and inclusive technological world[88].

The consequences of these biases inherent in algorithms and derived from the datasets on which they are trained can be truly pervasive. Such is the case with Amazon and its algorithm created in 2014 to analyze the various resumes of applicants[89]. Only a year later it was discovered that the AI in question discarded all women and resumes that contained related terms such as "female." Again, the fault lies in the dataset taken as a training model.

[84] MIT Media Lab, Person Overview ‹ Joy Buolamwini, MIT Media Lab https://www.media.mit.edu/people/joyab/overview/

[85] F. Gaetani, Coded Bias: la responsabilità dell'algoritmo., 2023 https://www.lisia.it/post/coded-bias-la-responsabilita-dellalgoritmo

[86] J. Buolamwini-T. Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, 2018, p. 77–91

[87] J. Buolamwini-T. Gebru, cit.

[88] MIT Media Lab, cit.

[89] L. Saettone, «Coded Bias», così abbiamo delegato il nostro razzismo agli algoritmi, Agenda Digitale, 2021 https://www.agendadigitale.eu/cultura-digitale/coded-bias-cosi-abbiamo-delegato-il-nostro-razzismo-agli-algoritmi/

This, in fact, contained old resumes analyzed by Amazon, a company that at the time had 60 percent of male employees in the technology sector, hence the resulting generalization that women were not suitable for those roles[90].

Another case of gender discrimination is found within the most popular image datasets such as the one mentioned above, which not only have national representation rates that do not match reality, but also promote gender bias. Indeed, in these images it is common to find men intent on hunting and women doing kitchen work, a typical representation of gender stereotypes.[91]

Gender inclusiveness is also the main subject of Caroline Criado-Perez's book: Invisible Women, which describes how everyday objects that we take for granted are not very inclusive of the female gender. The author highlights how these discriminations can be even better hidden in algorithms, as already in objects it is difficult to notice, and with software it is much harder to realize the extent of exclusion from certain processes.[92]

Another important case of discrimination due to algorithm bias is the one described by the investigative journalism organization ProPublica. ProPublica showed that COMPAS, software based on machine learning technology and used for the purpose of assessing the likelihood of recidivism of convicted and accused persons, proved to be unreliable and biased against people of color, thus leading to racial disparity in the assessment[93]. The errors made by the software were basically of two types: black defendants were more likely to be mis-categorized as likely future criminals, at a rate that was almost twice as high as non-black people, while white defendants were mis-categorized as low-risk more often than black defendants[94].

Similar racial discrimination occurred in America, where a system had been created to predict the identification of patients who would need extra medical support [95]. Again, a major bias toward people of color emerged, as the dataset contained old, unrepresentative

[90] L. Saettone, cit.
[91] European Parliament. Directorate General for Parliamentary Research Services., The ethics of artificial intelligence: issues and initiatives., Publications Office, LU, 2020
[92] D. Huyskes, AI ed Etica: L'importanza di usare dati rappresentativi, 2021 https://www.sas.com/it_it/news/leading-art-innovation/innovation-sparks/ai-etica-importanza-dati-rappresentativi.html
[93] J. Angwin-J. Larson-S. Mattu-L. Kirchner, Machine Bias, ProPublica, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[94] J. Angwin-J. Larson-S. Mattu-L. Kirchner, cit.
[95] L. Saettone, cit.

health expenditures. This is unfortunately not the only case in which the biases contained in artificial intelligence systems have a negative impact in health care settings, which often err in identifying diseases in patients of color[96]. This is the case with the deep learning system used to identify, from images, skin cancer. In this case, the dataset included only less than 5 percent of photographs depicting people with dark skin; therefore, the resulting analyses vary depending on skin color and are biased[97].

These cases are unfortunately not isolated but are now recurrent news. Being aware of these biases is the first step to be able to avoid and eliminate them during the artificial intelligence training process. This however is not easy, if in fact the algorithms are biased it is also because they learn from data coming from us humans, the biases in fact are inherent within each of us and it comes difficult to separate in this context technology and society[98]. In these cases, the algorithms are guilty of coding them.

2.1.2. Privacy

One of the strengths of artificial intelligence is certainly the fact that it can process and analyze huge amounts of data, data that is often personal. It can also happen that this personal data is collected without the consent of the individual owner, thus giving rise to concerns regarding privacy and data protection. It is also important to remember that there are many companies that collect and then resell the collected data to other companies, it should be an imperative, given also the monetization, that this data was collected in full respect of privacy. On a daily basis, in fact, we leave a huge trail of data behind us, which is often how we pay for services that seem free to us, giving rise to a kind of surveillance economy, if you can call it that. In a context such as the one outlined, losing control of our personal data can pose a danger[99]. Moreover, historian and philosopher Yuval Noah Harari says that in the not-too-distant future algorithms will come to know us

[96] L. Saettone, cit.

[97] J. Zou-L. Schiebinger, cit.

[98] F. Gaetani, cit.

[99] V. C. Müller, Ethics of Artificial Intelligence and Robotics, in E. N. Zalta (a cura di), The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, 2021Summer 2021

better than we know ourselves, thereby being able to have disproportionate control over people and their data[100].

Personal data can also fall into the wrong hands, or be misused. This is precisely why privacy is a fundamental right that must be guaranteed and respected during the adoption and use of artificial intelligence systems.

Artificial intelligence systems operating through machine learning have great potential, but they are also capable of turning seemingly innocuous data into sensitive and personal information, such as political preferences, personality, and emotional state. This often happens in the context of social media and is definitely a danger to be aware of[101].

One of the most famous cases that raised concerns regarding privacy juxtaposed with artificial intelligence was that of the interactive Hello Barbie doll, produced by Mattel in collaboration with ToyTalk. Hello Barbie was equipped with voice recognition and an AI system that allowed interaction between children and doll. However, the data collected from voice recognition raised widespread concerns as Hello Barbie recorded conversations with children that were then collected in ToyTalk's servers[102]. The concerns of parents and privacy activists touched on several points: the storage of the data collected from children could violate their privacy, the security of the data as there were widespread fears that the data could be hacked, the misuse of the information collected, and finally the informed consent given the use of the product by children who would surely not understand the implications of being recorded.

A case that involves many people and their privacy on a daily basis is undoubtedly that of voice assistants, such as Alexa, Siri, and Amazon Echo. Due to the recording and storage of conversations with users, privacy concerns have been raised. Indeed, these artificial intelligence systems always seem to be "listening," and in some cases recordings have been accidentally shared with unauthorized parties or used for marketing purposes.[103]

[100] M. Zur, Homo Deus: After God and Man, Algorithms Will Make the Decisions, Yuval Noah Harari, 2019 https://www.ynharari.com/homo-deus-after-god-and-man-algorithms-will-make-the-decisions/

[101] European Parliament. Directorate General for Parliamentary Research Services., cit.

[102] B. Marr, Barbie Wants To Chat With Your Child -- But Is Big Data Listening In?, Forbes, 2015 https://www.forbes.com/sites/bernardmarr/2015/12/17/barbie-wants-to-chat-with-your-child-but-is-big-data-listening-in/

[103] European Parliament. Directorate General for Parliamentary Research Services., cit.

Cambridge Analytica was perhaps one of the most important scandals of recent years. In 2018 it was revealed that Cambridge Analytica, a British political consulting firm, had collected the data of millions of users illegally, that is, without their consent, and then used it to influence elections and campaigns around the world[104]. This scandal was considered a turning point in terms of how and for what purposes user data is collected and used, marking a breaking point in terms of trust between consumers and large Tech companies such as Google and Facebook[105].

Google is another big tech company that has become embroiled in a privacy scandal. In 2010, the U.S. IT company, through vehicles used to map the Street View service, unintentionally collected data from unsecured wi-fi networks, including passwords, e-mails, credit card numbers and bank access codes[106]. Indeed, Google claimed that this incident was not voluntary, but that was not enough to calm privacy concerns given the amount of data collected and especially its sensitivity. Some users took legal action that resulted in a promise to delete all private data collected there and a $13 million settlement[107].

Another company that became party to a scandal concerning privacy violations was Clearview AI. The notorious company was exposed thanks to a 2020 New York Times article, which made it known that Clearview AI had collected a database of more than three billion images of people taken from various websites, including Facebook and YouTube, and then used them as a database for a facial recognition system[108]. Facial recognition software has also been used by law enforcement to identify criminals and suspects, as well as by other organizations, raising no small amount of concern. The affair ended in a £7.5 million fine by the U.K. data controller but has opened a global debate on

[104] C. Stroud, Cambridge Analytica: The Turning Point In The Crisis About Big Data, Forbes, 2018 https://www.forbes.com/sites/courtstroud/2018/04/30/cambridge-analytica-the-turning-point-in-the-crisis-about-big-data/

[105] C. Stroud, cit.

[106] C. Gerino, Google raccoglieva dati via WiFi: multa da 13 milioni di dollari per Street View, la Repubblica, 2019 https://www.repubblica.it/tecnologia/sicurezza/2019/07/24/news/google_raccoglieva_dati_via_wi-fi_multa_da_13_milioni_di_dollari_per_street_view-231905698/

[107] C. Gerino, cit.

[108] A. Hern, TechScape: Clearview AI was fined £7.5m for brazenly harvesting your data – does it care?, in The Guardian, 2022

the regulation and ethics of facial recognition, which needs a balance between privacy and security needs.[109]

Technologies such as the facial recognition technology just mentioned have been on the rise in recent years and can lead to benefits on safety and community but raise a privacy issue. It is crucial in this context to understand how to achieve the goals beneficial to cities while acting with respect for privacy and personal data[110]. Although in Western countries surveillance and predictive policing are considered sensitive issues especially with regard to privacy, in Asia these technologies are in wide use and acceptance among the population[111].

Among the Asian countries adopting artificial intelligence systems for citizen surveillance, China undoubtedly stands out. Indeed, the Chinese government has implemented an extremely extensive surveillance system through the use of facial recognition. China has always been at the forefront of the development and use of new technologies in recent years, a fact that has alarmed some governments as it makes the Asian state emerge as a world leader in these aspects. [112] However, the recurrent and massive use to this type of artificial intelligence systems risks leading to major privacy and human rights violations. In particular, in China these technologies are also used to monitor ethnic minorities such as the Uighurs, a use that is certainly not for the good[113]. In the book Surveillance State, written by Wall Street Journal journalists Josh Chin and Liza Lin, it is argued that the Chinese government has created a social contract with Chinese citizens, who agree to give up their data in order to receive governance that succeeds in making their lives easier and safer, or so it promises. It is also clear from this fact that the concept of privacy may differ depending on the culture in which one finds oneself; in fact, in China they have redefined this concept in a collectivistic key, and not an individualistic one as Westerners are inclined to understand it[114]. In addition, according to the AI Index Report 2023, the Chinese are among the people who have the most confidence in artificial intelligence

[109] A. Hern, cit.

[110] J. Barroca, Surveillance and Predictive Policing Through AI | Deloitte https://www.deloitte.com/global/en/Industries/government-public/perspectives/urban-future-with-a-purpose/surveillance-and-predictive-policing-through-ai.html

[111] J. Barroca, cit.

[112] Z. Yang, The Chinese surveillance state proves that the idea of privacy is more “malleable” than you’d expect, MIT Technology Review, 2022 https://www.technologyreview.com/2022/10/10/1060982/china-pandemic-cameras-surveillance-state-book/

[113] Z. Yang, cit.

[114] Z. Yang, cit.

technology; in fact, 78% of Chinese respondents agree with the statement "Products and services using AI have more benefits than drawbacks," which is the highest proportion among all nations surveyed[115].

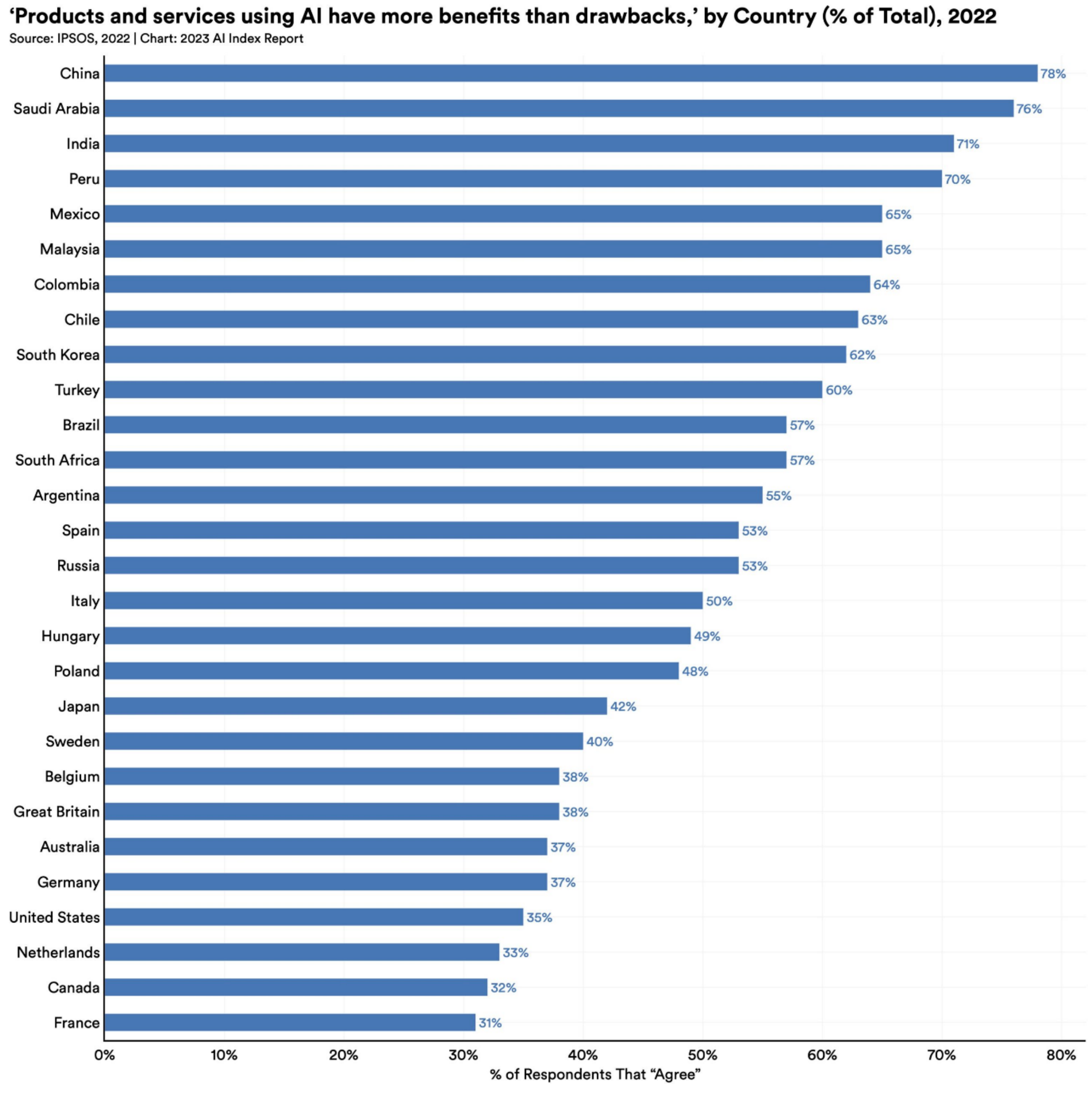

FIGURE 2.3 – Chinese citizens are among those who feel the most positively about AI products and services.

SOURCE: 2023 AI Index Report

[115] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, AI Index Report 2023 – Artificial Intelligence Index, Institute for Human-Centered AI, Stanford University, 2023

The Chinese government, as far as privacy is concerned, thus takes the same side as its citizens, and the privacy battles are against private companies. At this juncture, China has the Personal Information Protection Law and the Data Security Law, which provide for severe penalties for private companies that violate them[116]. China's momentum toward surveillance of its citizens began during the covid-19 pandemic, during which a number of tracking and surveillance measures operated through artificial intelligence systems were implemented. Among the various measures implemented are: previously mentioned facial recognition, often merged with thermal recognition to monitor people's body temperature; drones; applications for tracking people's contacts and movements; and social media monitoring for data analysis. The use of artificial intelligence has certainly helped contend with the spread of covid-19 in China, but the pervasive use made by the Chinese government has raised several criticisms based on that government's abuse of power and erosion of citizens' privacy leading to serious ethical concerns. Chinese ambition also seems not to stop there, and it turns out to be greater than previously expected; just think of the fact that the police through facial recognition systems are collecting citizens' voiceprints, or the fact that China's police are creating one of the largest DNA databases in the world, DNA that is being collected even from people who have no ties to the criminal world[117]. The goal of the Chinese authorities would seem to be to want to create comprehensive profiles, accessible to the government, for every citizen. Similar to this, in the city of Rongcheng, citizens go about their daily lives through a social scoring system, which rewards good deeds and punishes dishonest ones, a situation that appears more utopian than real[118]. The level of pervasiveness of these technologies in the Chinese context is indeed high, and although the tools used may lead to greater compliance with regulations, this scenario raises big questions regarding privacy but also human rights and freedom of the individual, underscoring the need for a global conversation regarding the balance between control and freedom and privacy especially in the context of advanced technologies such as artificial intelligence.

The potential of artificial intelligence, especially related to generative intelligence, is countless. But if text and images can be created, it is also possible to create deepfakes i.e.,

[116] Z. Yang, cit.

[117] I. Qian-M. Xiao-P. Mozur-A. Cardia, Four Takeaways From a Times Investigation Into China's Expanding Surveillance State, in The New York Times, 2022

[118] D. Davies, Facial Recognition And Beyond: Journalist Ventures Inside China's «Surveillance State», in NPR, 2021

videos that replace through AI one person's face with another, which can also damage a person's reputation and image, as well as their basic rights. Deepfakes bring both concerns related to the dissemination of false information and, of course, privacy of personal data. One of the biggest problems with deepfakes is that they are also extremely realistic fakes, which makes it very difficult to distinguish between real and fake content. Some of the most problematic cases regarding this phenomenon are the dissemination of fake pornographic videos and highly manipulated political speeches, which not only compromise privacy but also cause harm to both individuals and society.

Artificial intelligence can unfortunately also be used for manipulative purposes, threatening in some cases the privacy of individuals and in others the quality of information. Misinformation has been especially noticeable in social media on political election occasions. Some artificial intelligence-enabled technologies, for example, have been misused to manipulate voters; this is the case with bots, i.e., autonomous and fictitious accounts, which have been used to spread fake news, mostly for propaganda purposes[119]. Some well-known cases are that of the bots created by Russia to interfere on the Brexit-related referendum and that of the pro-Trump bots that tried to manipulate online spaces used by pro-Clinton Democratic supporters[120].

An alarming 2017 report highlighted that this problem has infiltrated even within a variety of governments. In fact, according to this Oxford University study, at least 28 countries, including states considered authoritarian but also democracies, have been "discovered" to employ so-called "cyber troops" aimed at manipulating public opinion within social networks[121]. The ways in which these cyber troops act to create disinformation and manipulation are many: they may target and heavily criticize users who criticize the government, or journalists and political dissidents; they spread false information aimed at propaganda; they magnify the number of interactions by creating artificial popularity; and much more[122].

[119] European Parliament. Directorate General for Parliamentary Research Services., cit.

[120] European Parliament. Directorate General for Parliamentary Research Services., cit.

[121] S. Bradshaw-P. N. Howard, DemTech | Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation, 2017 https://demtech.oii.ox.ac.uk/research/posts/troops-trolls-and-troublemakers-a-global-inventory-of-organized-social-media-manipulation/

[122] European Parliament. Directorate General for Parliamentary Research Services., cit.

The spread of misinformation (a consequence that can be either intentional or accidental) can, as just described, bring with it dangers to a just and healthy democracy, which is precisely why it would be crucial to have regular and uniform controls.

2.1.3. Transparency and responsibility

Many artificial intelligence algorithms, especially in more advanced systems, are considered black boxes, which makes it difficult if not impossible to understand the internal decision-making process of the machine. In this case, the liability derived from incorrect or harmful decisions derived from these systems also becomes opaque. The ideal scenario for ensuring ethical and responsible adoption of AI should involve either development of transparent AI processes and models to ensure that decision makers are able to understand and explain how a particular choice was arrived at.

Transparency in the context of artificial intelligence is about the ability to understand how algorithms work and thus to understand how they led to a particular decision, which is also a key piece in a trusting perspective toward this technology. Transparency, however, presents a challenge, especially with regard to more complex machine learning algorithms. When their complexity reaches a certain threshold, usually when we get to talk about deep neural networks, they are considered real black boxes. As the term itself says, it will be very difficult if not impossible to come to understand how a black box was able to lead to a certain decision, its operation being far from clear. In fact, in 2018, the AI Now Institute in New York proposed a stop to the use of these black box algorithms with regard to public agencies dealing especially with justice, health, education, and welfare because the decisions made by these systems cannot be explained[123].

This is the case with the infamous EVAAS software, which stands for Education Value Added Assessment System, used to evaluate teacher effectiveness in some states in the United States[124]. The evaluation was based on the students' academic progress and compared it accordingly with that of other students deemed similar, finally assigning a score to professors, a score that was then used to decide on the fate of teachers who could also be fired or promoted. Controversy ensued because it was unclear how the algorithm

[123] European Parliament. Directorate General for Parliamentary Research Services., cit.

[124] European Parliament. Directorate General for Parliamentary Research Services., cit.

worked, the software in fact being from SAS Institute, a company that treated the algorithm as a trade secret[125]. So it was that teachers filed a lawsuit against the school district claiming that the use of the software would violate their rights, particularly of equality. In 2017, a federal judge ruled in favor of the group of teachers, ruling that their right to due process had been violated being that it was not possible to challenge the scores assigned by EVAAS properly not knowing how it works. The teachers thus obtained a stop to the school district's use of the software. While artificial intelligence certainly has the potential to be able to improve several aspects in education, it is critical to address apprehensions related to the transparency and fairness of software such as AI-based EVAAS.

These situations, which negatively affect transparency, can lead to bias and discrimination, such as the examples described above. It is precisely this opacity that makes fertile ground for bias and prejudice. Lack of transparency will also hinder accountability, as it will be difficult to understand at what stage of the decision the error or problem occurred.

Assimilated with accountability, we can also identify uncertainty of ownership. If a deepfake is created, who owns it? It could be the creator, or the person represented, or even the company that owns the tool that generated the deepfake. This also applies to music and art generated in this way, it is important to determine who owns the rights to it.

It is equally important to establish responsibility, e.g., if an AI-equipped self-driving car causes an accident, it is necessary to establish who is responsible. It could be the developer of the artificial intelligence system, or the company that sold it, or even the car company. It is indeed a very important ethical challenge to determine liability especially for damages caused by a product featuring artificial intelligence. This also includes liability in the cases of racial and gender discrimination explained above, the unfair decision must have a responsible party who can answer for the damage done.

[125] I. Sample-I. S. E. di Scienze-I. Sample, Computer says no: why making AIs fair, accountable and transparent is crucial, in The Guardian, 2017

2.1.4. Security

The adoption and development of artificial intelligence technologies, present among many challenges, those related to security, security that can be declined in different contexts.

Assimilated to the privacy discourse is the issue of data security, which is one of the most critical aspects in the era of big data. Indeed, increasing the amount of data collected, which to date is extremely high, also increases the risk of personal data falling into the wrong hands, such as those of hackers, leading to situations such as identity theft, abuse and data breaches. One disastrous example was the data theft of Equifax, a U.S. credit reporting company, which announced on September 7, 2017, that the personal and sensitive information of more than 140 million consumers had been stolen in a major data breach.[126] Indeed, a major security problem is cyber attacks, which exploit vulnerabilities and flaws in infrastructures and then compromise them and steal their data. Thanks to artificial intelligence, cyber criminals can improve and even automate their attacks, a widespread example being "phishing" messages that seem increasingly personalized and convincing. There is a clear need to protect artificial intelligence systems and computing facilities in general to prevent such cases from continuing to occur.

One threat in this regard are so-called adversarial attacks. Adversarial attacks are attacks carried out by malicious actors whose purpose is to manipulate input data to produce erroneous outputs with negative consequences, thereby deceiving artificial intelligence systems and insinuating themselves into the algorithm training process[127]. The effects of these attacks could lead to very dangerous situations, just think of an attack that leads to misinterpreted signals in an autonomous driving vehicle or the consequences that such an attack would cause to military defense systems.

There have also been many concerns regarding autonomous weapons, such as in the previously mentioned 2015 Open Letter on Autonomous Weapons calling for an international ban on them[128]. Autonomous weapons are weapons that are able to act without the need for direct human intervention, a fact that brings safety and

[126] S. Srinivasan, Data Breach at Equifax - Case - Faculty & Research - Harvard Business School, 2017 https://www.hbs.edu/faculty/Pages/item.aspx?num=53509

[127] W. Knight, How malevolent machine learning could derail AI, MIT Technology Review, 2019 https://www.technologyreview.com/2019/03/25/1216/emtech-digital-dawn-song-adversarial-machine-learning/

[128] S. Gibbs, cit.

accountability concerns. It is a lethal force that delegates responsibility to an artificial intelligence system, thus raising issues not only of safety but also of accountability and responsibility, just think of the eventuality in which an autonomous weapon makes unintended civilian casualties or causes international incidents that can undermine relations between different countries. This new type of weaponry also risks leading to a potential arms race; indeed, an international legislative framework in this regard would be appropriate to limit the risk of armed conflict. Such legislations should certainly take into account the adoption of ethical principles to ensure that the introduction of this technology into the military is done in full compliance with ethical principles and accountability.

2.1.5. Inequalities and Unemployment

Artificial intelligence certainly brings many benefits to the world of work as well, automating repetitive tasks and being able to analyze large amounts of data in very little time. However, automation can lead to a reduction in employment in some cases, which can potentially lead to the exacerbation of social inequalities as the jobs that artificial intelligence replaces are mainly low-skill jobs.

It was only a few days ago that IBM will suspend hiring for jobs that can be replaced by artificial intelligence, which make up about 7,800 jobs within the company. These jobs are mostly non-customer-facing roles such as back office and human resources, and it is expected that they could be replaced by AI in about five years[129].

Globally about 300 million jobs could be impacted by the new generative AI technology according to Goldman Sachs, a number that corresponds to 18 percent of jobs worldwide and is expected to weigh more in already advanced economies[130]. On the other hand, the advent of these new developments in AI has the potential to increase global GDP by 7 percent each year in 10 years, and it is estimated that the impact on workers will be more marked by integration than complete replacement[131].

[129] Reuters, IBM to pause hiring in plan to replace 7,800 jobs with AI, Bloomberg reports, in Reuters, 2023

[130] M. Toh, 300 million jobs could be affected by latest wave of AI, says Goldman Sachs | CNN Business, CNN, 2023 https://www.cnn.com/2023/03/29/tech/chatgpt-ai-automation-jobs-impact-intl-hnk/index.html

[131] M. Toh, cit.

Although such news sounds alarming one must consider the opportunities for growth that this technology can bring and the new jobs that will be created, especially in engineering, which will require new skills. Jobs that are exempt from these dangers are mostly those related to human creativity and interpretation, such as politicians, psychiatrists, and investigators[132]. From this scenario and the continued development of new artificial intelligence technologies, new industries will also emerge, which in turn will create new job opportunities, such as AI ethics supervision, AI-driven education, or AI-enhanced healthcare[133]. Generative AI, however, has the potential to impact not only, as previously, low-skilled jobs, but also a range of so-called white-collar roles [134]. In fact, as the 2023 Future Of Jobs report compiled by the World Economic Forum makes clear, right now the jobs with the fastest decline are related to white-collar, clerical and secretarial mostly, while those with the most growth are those related to AI and engineering[135].

Indeed, some fear that this new wave of artificial intelligence may reduce the value of some skills by de-skilling typical middle-class jobs. This could lead to the transfer of workers to jobs where the chance of earning a living is lower[136].

In this scenario, it therefore becomes a priority to balance the benefits brought by automation with the need to ensure equitable distribution. Outplacement and training policies could be proposed in this regard, so as to stem job losses and help workers potentially impacted by automation to develop new skills by adapting to the new work environment.

One possible solution to these problems was proposed by Juliet Schor, an economist at Boston College, who proposed reducing employees' hours to three or four days a week, rather than laying them off, thus avoiding the net displacement of work[137]. Schor also proposes the introduction of some kind of universal income system so as to curb the danger of social inequality. The issue of inequality and income inequality also worries scholars such as Lawrence Katz, a labor economist at Harvard, who believes in the

[132] D. Lo, Will you lose your job to AI and tech like ChatGPT?, Fast Company, 2023 https://www.fastcompany.com/90881876/ai-chatgpt-take-jobs
[133] D. Lo, cit.
[134] S. serra-S. Greenhouse, US experts warn AI likely to kill off jobs – and widen wealth inequality, in The Guardian, 2023
[135] World Economic Forum, The Future of Jobs Report 2023, 2023
[136] S. serra-S. Greenhouse, cit.
[137] S. serra-S. Greenhouse, cit.

possibility that the share of labor income will go down gradually if many activities are automated[138].

Inequalities can also be exacerbated by the concentration of power relative to large corporations. Indeed, AI can lead to a concentration of power by increasing the potential risk of abuse of power, which can lead to both social and economic inequalities. This is the case with large companies such as Google, Meta and Amazon, which seem unstoppable when compared with their competitors. The reality of the Internet and new technologies is shaped by these few companies, which are also dominating the development of artificial intelligence. The resulting technological power allows these large technology companies to have enormous influence regarding areas of society that are related to the shaping of people's opinions[139]. Specifically, the power of big tech can be summarized in three macro areas: financial, as they can afford large investments; public discourse, as they control the infrastructure useful for public discourse; and personal data collection, collection that takes place for profiling and profit-making purposes[140].

This argument, however, does not only apply at the company level, but can also be dangerous when juxtaposed at the level of a country's government, such as China, cited above for its ambitious strategies of surveillance and use of artificial intelligence.

Balancing power then becomes an additional ethical challenge concerning the sphere of AI and technologies in general.

To remedy this problem, one proposal is to democratize artificial intelligence, thus placing this principle at the basis of AI ethics [141]. In fact, there are some discussions related to the democratization of AI, with many people arguing that it would be appropriate to make this technology accessible to a wider range of people, regardless of their technical, cultural, or socioeconomic background, so as to spread AI-related skills and knowledge. This process would prevent people from being "controlled" by a few experts and has several advantages[142]. In this way, people would be the ones to design and implement it,

[138] S. serra-S. Greenhouse, cit.

[139] European Parliament. Directorate General for Parliamentary Research Services., cit

[140] European Parliament. Directorate General for Parliamentary Research Services., cit

[141] L. Eliot, Democratization Of AI Is Said To Be Essential For AI Ethics But The Devil Is In The Details, Including The Case Of AI-Based Self-Driving Cars, Forbes, 2022 https://www.forbes.com/sites/lanceeliot/2022/03/24/democratization-of-ai-is-said-to-be-essential-for-ai-ethics-but-the-devil-is-in-the-details-including-the-case-of-ai-based-self-driving-cars/

[142] L. Eliot, cit.

potentially leading to more inclusive solutions, wider education regarding digital skills would also be promoted, and digital inequalities could be combated. Citizen involvement would then be very useful in terms of ethics in general related to artificial intelligence, but also in terms of transparency and accountability related to it, leading to greater inclusion, innovation and even sustainability in general.

2.1.6. Impact on Human Relations

The impact of artificial intelligence is evident and significant on society; in fact, it is transforming not only how we work and communicate, but also how we interact. With regard to this technology, both positive and negative effects on human relationships and society can be observed.

Interactions with artificial intelligence systems, such as virtual assistance and chatbots, can negatively affect human relationships and psychological well-being. In this context, concerns can arise regarding people's sense of isolation and loss of authenticity in relationships. AI has indeed revolutionized the way we communicate and interact with others, but interacting with systems such as Siri and Alexa, which certainly make it easier for us to manage everyday tasks, can reduce the time we spend on authentic human relationships. Overuse of these virtual assistants could result in addiction, leading to a reduction in the development of meaningful human relationships.

Nowadays we also find this technology within popular dating platforms, which use algorithms to suggest potential partners based on shared interests and preferences, thus revolutionizing the way many people seek romantic relationships. This certainly enables connection between otherwise unknown people but moves away from authentic social interactions. The use of these platforms can also lead to a superficiality of relationships, emphasizing physical attraction and quantity of compatible profiles rather than on the quality and authenticity of interactions. Still regarding the influence of artificial intelligence on human-to-human relationships, another real risk is also the decrease in habitual daily relationships, which can lead to the isolation of people but also to a reduced ability to cooperate with others[143].

[143] European Parliament. Directorate General for Parliamentary Research Services., cit

If these can be considered negative externalities of artificial intelligence, mention should also be made of the development of systems that want to be useful in this sense and provide companionship, with the goal of reducing the sense of isolation. This is the case with social robots, including Jibo, which by understanding speech and expressions tries to form relationships; Tico, which aims to increase children's motivation at school; and Kismet, a robot that can both understand and exhibit emotions during an interaction[144]. These applications can also provide support and advocacy for people with disabilities or the elderly, as well as tutoring for students, companionship, and even increased client engagement [145]. However, it must still be remembered that they are social robots and not humans; in fact, they possess neither emotions nor empathy, and there is always the risk that by making inordinate use of them, genuine human relationships will be diminished. Usually, moreover, people have a tendency to treat robots and machines in general as if they were real people[146]. Concern is also raised in these cases regarding the lack of authenticity in relationships, which could lead to reduced social "skills" but also reduced empathy.

Another criticism that can be made of social robots is that, once the human's trust has been won, there is a risk that the person will be manipulated, a scenario that can also be realized through a hacker attack[147]. It is also possible that the company designs robots that can increase their reliability, and in this scenario they could also be used to interrogate people or to sell products and services[148].

2.2. Ethical principles for responsible AI

Artificial intelligence ethics is a field of study that has emerged to address concerns related to the impact of artificial intelligence, and it is constantly evolving. Artificial intelligence ethics is considered a set of values, principles, and techniques that use widely

[144] I. Wigmore-D. Shao, What is social robot? | Definition from TechTarget, Enterprise AI, 2022 https://www.techtarget.com/searchenterpriseai/definition/social-robot

[145] I. Wigmore-D. Shao, cit.

[146] G. Lughi, Robotica sociale: gli impatti de «La società dei robot» tra spazi ibridi ed etica, Agenda Digitale, 2022 https://www.agendadigitale.eu/cultura-digitale/robotica-sociale-gli-impatti-de-la-societa-dei-robot-tra-spazi-ibridi-ed-etica/

[147] European Parliament. Directorate General for Parliamentary Research Services., cit

[148] European Parliament. Directorate General for Parliamentary Research Services., cit

accepted standards of right and wrong to guide moral conduct in the development and use of artificial intelligence technologies.[149].

Recent years have seen an increasing number of guidelines and treatises regarding ethics in AI, this has led to a multitude of guidelines and principles that can be confusing being in large numbers, but it also stands to signify the importance and relevance of the topic in the current context.

Among the various guidelines that have been developed, we can certainly mention the famous 2017 Asilomar Principles, designed by the nonprofit organization Future Of Life Institute, whose goal is to reduce the risks associated with transformative technologies[150]. The Asilomar Principles consist of 23 principles that serve as the basis for the development of safe and ethical artificial intelligence. The main goal of this work is to ensure that AI is used for beneficial purposes and prevent any harm that might result. Principles outlined in this paper include safety, beneficence, non-maleficence, autonomy, accountability, transparency, respect for freedom and privacy, and many more[151].

Another landmark regarding ethical principles related to AI is certainly the 2018 Montreal Declaration for a Responsible Development of Artificial Intelligence[152]. The Montreal Declaration provides ten ethical principles useful for the responsible development of artificial intelligence to ensure that these technologies benefit people and respect human rights. The 10 principles described are: well-being, respect for autonomy, protection of privacy and intimacy, solidarity, democratic participation, equity, diversity inclusion, caution, responsibility, and sustainable development[153].

In 2019, guidelines regarding ethics were published by the European Commission to succeed in achieving reliable AI. These guidelines are based on seven principles, which are: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; accountability[154]. The framework outlined by the European

[149] E. Kazim-A. S. Koshiyama, A high-level overview of AI ethics, in Patterns, fasc. 2, 9, 2021.

[150] Future of Life Institute, AI Principles, 2017 https://futureoflife.org/open-letter/ai-principles/

[151] Future of Life Institute, cit.

[152] Université de Montréal, The Declaration - Montreal Responsible AI, 2018. https://www.montrealdeclaration-responsibleai.com/the-declaration

[153] Université de Montréal, cit.

[154] European Commission, Ethics guidelines for trustworthy AI | Shaping Europe's digital future, 2019 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Commission aims for reliable AI, i.e., lawful, ethical and robust, and recommends an approach that relies on risk analysis to assess its ethical impacts.

Also in 2019, the Organization for economic cooperation and development (OECD) also drew up its own guidelines regarding this technology with the aim of promoting the development of ethical and responsible AI.[155] Here the principles are divided into five macro-areas: beneficence and human development; justice and nondiscrimination; transparency and explainability; robustness and security; and accountability. In addition to the principles, the OECD has also developed recommendations addressed to policymakers, which we will see below.

Those just described are only four examples of guidelines developed in artificial intelligence ethics, but there are many more. Indeed, in recent years several entities have been busy developing their own ethical principles, and the list includes governments, nonprofit or for-profit organizations, universities, and other types of organizations. This multitude of principles results in a very broad and fragmented picture, which can lead to confusion given the lack of a unified framework.

An analysis based on 20 documents presenting the guidelines of key stakeholders shows there are both convergences and overlapping principles among the different frameworks, which point to a common direction, and differences[156]. The analysis carried out by Giovanni Vaia, Noor UL-Ain, Elisa Gritti, and Marco Bisogno is the result of an effort that would like to lead to an unambiguous reference regarding ethical lines regarding artificial intelligence, and five ethical approaches regarding AI are in fact proposed, which turn out to be the common themes found among the 20 papers analyzed. The five approaches are:

- AI-Tech approach: this first approach focuses on the purely technological aspects of artificial intelligence and covers 15 papers out of the 20 analyzed. One of the key principles of this approach is explicability; to ensure that people can trust AI, it is necessary for them to understand its decisions; therefore, algorithms should be able to allow people to understand how they work. Another important principle in this approach, and related to the previous one, is transparency. Indeed, it is crucial that different stakeholders have access to

[155] OECD, The OECD Artificial Intelligence (AI) Principles, 2019 https://oecd.ai/en/ai-principles

[156] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, Ethical Systems for Artificial Intelligence, in Intellectual Capital, Smart Technologies and Digitalization, Springer, 2021

the algorithms but also to the data used. Privacy is another key principle in this approach; given the multitude of data that this technology can process, it is essential that this data be collected in a way that respects privacy and the protection of personal data. The goal then of the AI-Tech approach is to ensure that AI is interpretable, understandable, and ethically responsible[157].

- AI-Security Approach: the second approach covers 12 out of 20 papers and focuses on the security and protection aspect of AI systems. Indeed, in order to preserve the integrity and proper functioning of artificial intelligence, it is essential to protect systems from external attacks but also to prevent and avoid deliberate incidents that can create damage[158].
- AI-Human Centric approach: this approach can be found in 15 of the 20 papers and considers artificial intelligence as if it were a kind of extension of the human being. Respect for autonomy is a key concept here, as technology should always respect the different personal choices of individuals, thus not interfering with their freedom. Another key principle is fairness; artificial intelligence should treat all people equally, without discrimination. This principle is also important in the vision that this technology succeeds in working for the good of all individuals. Additional principles of this approach are human-centric orientation as the name suggests, and human control. The goal of AI should always be aimed at improving people's lives, and people themselves should have control over artificial intelligence; this would ensure that AI serves humans and poses no threat to the human race[159].
- AI-Ethical Approach: this approach is included in 16 of the 20 papers analyzed and holds that artificial intelligence is an entity similar to humans in terms of shared values and responsibilities. The principles included here are human rights and values, accountability, protection of privacy and intimacy, judicial transparency, and the right to equality and nondiscrimination. Indeed, it is critical that AI respect and adhere to the values shared by the international community and respect human rights, such as the right to privacy, the right to equality, and freedom of expression. Regarding accountability, it is crucial that

[157] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.
[158] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.
[159] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.

any negative consequences derived from AI are addressed and that there are accountable parties for decisions made through algorithms. In addition, in the justice system, artificial intelligence should provide transparency so that there can be justice and fairness.[160]

- AI-Benefit approach: this approach is found in 8 of the 20 papers analyzed and focuses on the benefits of artificial intelligence on society and humanity in general. Central principles of this approach are the common good and the benefit of humanity, but also empowerment. Indeed, AI should be able to empower communities and individuals so that they gain greater capabilities. Two other key principles are sustainable development and cooperative culture, this technology should indeed promote collaboration, both between individuals but also between organizations and governments, so as to ensure that benefits are shared equitably. The goal of this point is to ensure that this technology is not only developed, but also used in a way that reaps maximum benefit both socially and economically.[161]

These five approaches succeed in providing a comprehensive view of the potential impacts of artificial intelligence and its role in society, and from this analysis ethical principles were then proposed for each approach[162]. The proposed principles are as follows:

- AI-Tech approach: explainability transparency, privacy, availability, reliability, quality, flexibility, accuracy;
- AI-Human centric approach: individual's freedom, respect for human autonomy, fairness, effective human-machine interaction, education;
- AI-security approach: harm prevention, safety, security, control, risks identification and management;
- AI-Ethical approach: human values and human rights, accountability, responsibility, equality, well-being;
- AI-Benefit approach: common goof and benefit of humanity, sustainable development, inclusive growth, human empowerment, cooperative culture[163].

[160] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.
[161] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.
[162] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.
[163] G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, cit.

The framework outlined unifies the multitude of ethical guidelines that have developed in recent years and provides a solid basis for the development of standards and regulations to ensure that artificial intelligence is used in a sustainable and safe way that benefits everyone. Indeed, a single reference in this context would be needed so that governments also have a clear and unambiguous basis on which to legislate.

Another framework that aims to unify the multitude of principles proposed is the one Prof. Luciano Floridi describes. Floridi proposes five ethical principles, of which four are considered traditional principles of bioethics, while one is defined as a new enabling principle for artificial intelligence[164].

The first principle described by Floridi is beneficence, which is aimed at promoting well-being, preserving dignity, and sustaining the planet. In fact, as described above, technology should be used for the benefit of humanity, and this fact is echoed by the majority of the guidelines[165]. The second principle is the principle of non-maleficence, which takes into consideration the aspects of privacy, security and "caution of capacity." Although it seems logically similar to the first, it actually has different contents. Here, in fact, the various negative consequences that could result from the misuse of artificial intelligence are highlighted; however, it remains ambiguous whether this principle is aimed at the people who develop this technology or at AI itself, but this is where the next principle comes into play: autonomy[166]. Autonomy is the third principle proposed by Floridi and aims to balance artificial autonomy, i.e., that of the machine, with human autonomy, i.e., the decision-making power we reserve for ourselves. In this context, it is clear that human autonomy is promoted while machine autonomy is limited and also made reversible, reversibility that guarantees the power to "decide to decide again"[167]. The fourth principle proposed is that of justice, which implies promoting prosperity, preserving solidarity and avoiding inequity. Justice, in the various guidelines takes on different connotations but is often linked to fairness thus absence of discrimination[168]. The four principles mentioned are all considered traditional principles of bioethics, while the next is considered a new enabling principle: explicitness. As mentioned earlier and as recurs in the majority of the guidelines, it appears necessary to both understand and account for what are the decision-

[164] L. Floridi, The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities, 2022
[165] L. Floridi, cit.
[166] L. Floridi, cit.
[167] L. Floridi, cit.
[168] L. Floridi, cit.

making processes of this technology. Thus in this principle we find concepts such as transparency, accountability and precisely explicability[169].

Despite the importance of these guidelines and frameworks described, it is important to remember that these principles are only a starting point if the ethical challenges of artificial intelligence are to be addressed. Indeed, as the technology develops and evolves, its influence on people and society will also increase, likely creating new ethical challenges that will require updates to the guidelines and new thinking. In order to promptly address these challenges, an open but also timely dialogue between the various stakeholders involved in both the development and use of AI is also crucial.

2.3. Legal framework

Technology has changed substantially in recent decades and has seen considerable development especially in artificial intelligence, but regulation is often slow to respond to change, a slowness that can be exploited particularly by first movers in the field. The current legal framework regarding AI and ethics is still evolving, but many countries are already developing their regulatory references in this regard.

According to the Artificial Intelligence Index Report 2023, the growing popularity of this technology has led many governments and intergovernmental organizations and to define and develop management strategies regarding AI. The report recorded a nearly 6.5-fold increase since 2016 in mentions of AI in legislative proceedings in 81 countries, a sign that policymakers' interest in this discipline is growing[170]. On the other hand, as for the analysis of a pool of 127 countries, if there were only one law passed concerning "artificial intelligence" in 2016, there are 37 in 2022 alone, as shown in the following chart[171].

[169] L. Floridi, cit.

[170] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, cit.

[171] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, cit.

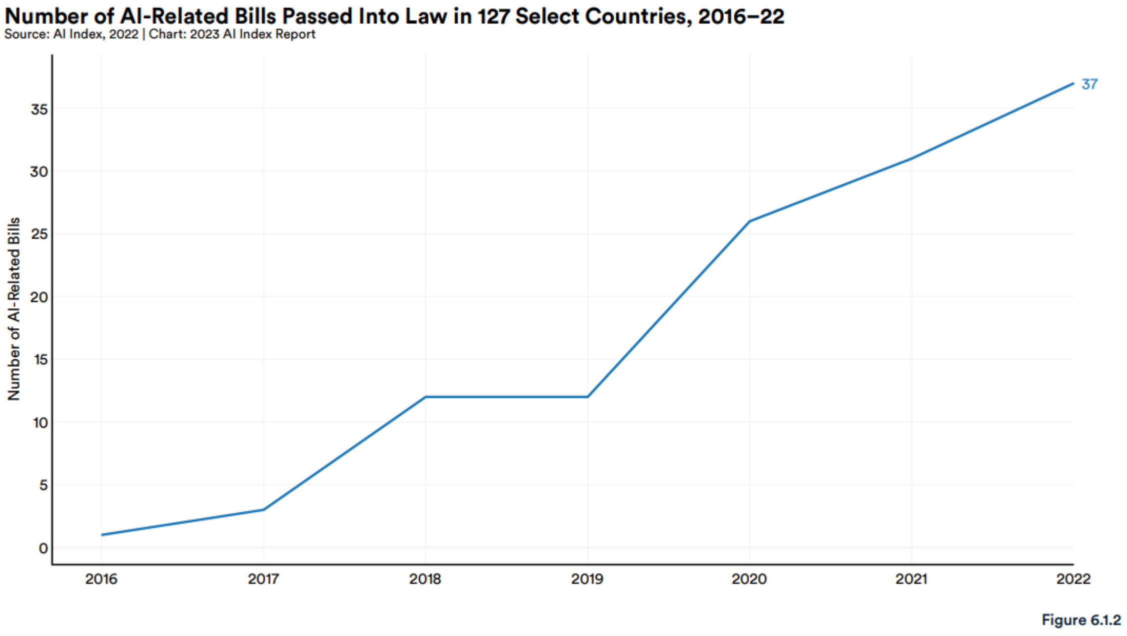

FIGURE 2.4 – Number of AI related bills passed in 127 selected countries 2016-2022.
SOURCE: 2023 AI Index Report

If one also goes to investigate artificial intelligence-related lawsuits, there are about seven times as many in the United States than in 2016.

Despite this, the regulatory framework appears, similarly to that of ethical principles, to be very fragmented, and more harmonization at the international level would be needed. Arriving at a uniform and up-to-date framework of rules is a challenge to date; in fact, it would be ideal to arrive at a holistic approach to ethical issues related to artificial intelligence, involving a large number of stakeholders so as to ensure an effective but also balanced outcome.

Looking at the European landscape, the European Parliament as early as 2017 took the first steps toward regulation inherent in artificial intelligence through the Report with Recommendations to the Commission Concerning Civil Law Rules on Robotics; this report calls on the Commission to develop legislation on intelligent robots and proposes two possible codes of ethics of conduct, one for ethics committees and the other for the engineers involved[172]. In 2019, in addition to the Declaration of Cooperation on Artificial

[172] Stefanelli & Stefanelli Studio Legale, Raccolta fonti normative sull'AI, Stefanelli & Stefanelli Studio Legale https://www.studiolegalestefanelli.it/it/raccolta-fonti-normative-intelligenza-artificiale

Intelligence being signed by member states, the "White Paper on Artificial Intelligence at the Service of the Citizen" is developed and published by the European Commission. However, the White Paper represents a fundamental step regarding a single regulatory framework for the development and use of AI in the European Union. This document is based on three pillars: the goal of creating a system of excellence, the goal of creating an ecosystem of trust (i.e., having ethically aligned AI), and a human-centric approach (which requires that fundamental rights be respected)[173]. In addition to the creation of ecosystems of excellence to promote research and development in Europe and the establishment of an ecosystem of trust that leads back to the various ethical principles described above, the white paper recognizes the importance of international collaboration for the proper promotion of global standards on AI and for addressing its challenges. A high-level expert group (AI HLEG) was also appointed by the European Commission to succeed in providing advice regarding a strategy on AI. In fact, the AI HLEG has drafted four documents[174]:

- Ethical Guidelines for Trustworthy AI: proposes the human-centered approach and develops seven principles for trustworthy AI. The seven principles are: human agency and oversight; technical robustness and security; privacy and data governance; transparency; diversity, nondiscrimination and equity; social and environmental well-being; and accountability.[175].
- Policy and investment recommendations for reliable AI: a collection of 33 recommendations to guide the development of reliable AI, based on the seven principles of the previous paper [176].
- The final Assessment List for Trustworthy AI (ALTAI): a document that reassimilates previous documents into a checklist, so as to make them easier to use for developers and companies that want to implement its principles[177].
- Sectorial Considerations on the Policy and Investment Recommendations: document that develops the considerations made in the previous documents

[173] L. Mischitelli, La strategia europea sull'intelligenza artificiale: stato dell'arte e scenari futuri, Agenda Digitale, 2020 https://www.agendadigitale.eu/cultura-digitale/la-strategia-europea-sullintelligenza-artificiale-stato-dellarte-e-scenari-futuri/

[174] European Commission, High-level expert group on artificial intelligence, 2023 https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[175] European Commission, cit.

[176] European Commission, cit.

[177] European Commission, cit.

particularly in three specific areas, namely public sector, health care and manufacturing, and Internet of Things[178].

Then, in 2021, the AI package was proposed by the European Commission, which includes a communication on promoting a European approach to AI, a review of the coordinated plan on AI, and a proposal for a regulation establishing harmonized rules on AI with related impact assessment[179]. The Commission also proposed three interlinked legal initiatives to succeed in achieving a reliable AI: a European legal framework for AI that can address fundamental rights and security risks, a framework for AI liability, a review of sectoral security legislation[180].

The proposal for a unified framework is now becoming a reality; in fact, the green light was recently given in Brussels to the AI Act proposal brought by the Commission. This is the world's first proposal for a unified legal framework, which aims to ensure that the democratic values of the European Union and legislation are respected by systems marketed within the EU[181]. The Artificial Intelligence Act thus aims to regulate AI within the borders of the EU and establishes a regulatory framework that classifies artificial intelligence systems according to the risk and potential for harm they may pose by imposing specific legal requirements and definition of responsibilities for each of the categories.[182] An additional level of high risk has also been introduced, which includes possible harm to health, security, and fundamental rights. This regulation provides stricter obligations for the category to which ChatGPT also belongs and extends the ban on biometric identification software, which previously was prohibited only when used in real time, now it will in fact be possible to use them ex post in cases of serious crimes and through the authorization of a judge[183]. It also establishes a ban in the areas of law enforcement, border management, labor, and education on the use of emotion recognition software and there is a ban on applications deemed to be of unacceptable risk such as

[178] European Commission, cit.

[179] European Commission, A European approach to artificial intelligence | Shaping Europe's digital future, 2023 https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

[180] European Commission, cit.

[181] R. Carrozzo, Diritto dell'AI: il ruolo dell'etica, AI4Business, 2022 https://www.ai4business.it/intelligenza-artificiale/diritto-dellai-il-ruolo-delletica/

[182] F. Meta, Artificial Intelligence Act, accordo politico al Parlamento Ue sulle nuove norme, CorCom, 2023 https://www.corrierecomunicazioni.it/digital-economy/artificial-intelligence-act-accordo-politico-al-parlamento-ue-sulle-nuove-norme/

[183] F. Meta, cit.

manipulative techniques and social scoring[184]. The proposal also extended the prohibition of predictive monitoring to administrative crimes, in addition to criminal offenses, and protections for sensitive data are increased by providing stricter controls[185]. The AI Act not only sets limits and prohibitions (which do not cover national security and military purposes that are excluded from the scope) but also provides measures to support innovation[186]. In fact, the principle is introduced that spaces designed for regulatory testing of artificial intelligence, which create a controlled environment, can also be used to test new AI technologies under real conditions. A further provision stipulates that new systems can be tested in real conditions without controls, but with pre-established conditions and guarantees, and there are also provisions for easing administrative obligations for smaller companies[187]. This new and comprehensive legal framework therefore addresses the risks and sets the limits with regard to AI, seeking to position Europe as a leader in this field globally, with the goal on the one hand to take advantage of the opportunities offered by artificial intelligence and on the other hand to limit its risks and address its challenges.

With regard to European legislation concerning artificial intelligence, mention must also be made of the GDPR (General Data Protection Regulation). The GDPR regulates respect for the privacy and personal data of European citizens. As far as the relationship with AI is concerned it is closely related to the scope of personal data, as artificial intelligence processes huge amounts of personal data and it is important to ensure that AI complies with the protection standards set by this legislation. Although the GDPR does not make specific reference to artificial intelligence, it does regulate the processing of personal data in any technology that makes use of it, and therefore provisions relevant to AI systems can also be identified. These provisions are specifically the principles of data minimization, purpose limitation, and retention limitation[188]. Despite this, the regulation only deals

[184] L. Bertuzzi, AI Act moves ahead in EU Parliament with key committee vote, www.euractiv.com, 2023 https://www.euractiv.com/section/artificial-intelligence/news/ai-act-moves-ahead-in-eu-parliament-with-key-committee-vote/

[185] F. Meta, cit.

[186] Consiglio dell'UE, Normativa sull'intelligenza artificiale: il Consiglio chiede di promuovere un'IA sicura che rispetti i diritti fondamentali, 2022 https://www.consilium.europa.eu/it/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/

[187] Consiglio dell'UE, cit.

[188] M. Borzacchelli, Intelligenza artificiale e GDPR sono compatibili, ma servono norme più chiare: lo studio, Agenda Digitale, 2020 https://www.agendadigitale.eu/sicurezza/intelligenza-artificiale-e-gdpr-sono-compatibili-ma-servono-norme-piu-chiare-lo-studio/

with the personal data of citizens, not the aggregated and anonymized data that are often used for algorithm training. Problems can arise from this aspect, as it is sometimes possible to reconstruct personal data in a model thus violating privacy[189]. Another blind spot in the GDPR is the fact that there are no rights or obligations related to algorithmic systems in the period after their creation, but only before their use, thus posing a risk to citizens[190]. The new wave of generative artificial intelligence, symbolically represented by ChatGPT, has already clashed with the privacy-related regulations of the GDPR. In fact, at the end of March, in Italy, the Privacy Guarantor blocked ChatGPT on grounds related to unlawful collection of personal data and absence of systems to verify the age of minors.[191]

2.4. Strategies and tools to promote ethical AI

In the current landscape, artificial intelligence is taking on an increasing and widespread role so it is crucial to be able to ensure that its use and development takes place in full compliance with ethical principles and fundamental rights, thereby benefiting society. Indeed, being able to create reliable AI systems in this regard has become a key area of research.

In order to promote ethical and reliable AI, it is first necessary to establish an ethical framework that establishes well-defined principles. Ideally, this framework should be unambiguous, so as not to create differences, and should be the result of an open and collaborative dialogue among all key stakeholders; this is to ensure respect for human rights and other previously mentioned principles such as equity, accountability and transparency. In this regard, it would be useful to promote broad participation and inclusion regarding the development of an ethics AI in order to involve the various stakeholders to ensure that everyone's concerns are respected. The importance of genuine international cooperation, both among governments but also among organizations and research communities, is also emphasized, so as to facilitate knowledge exchange, promote innovative solutions, and achieve a coordinated approach.

[189] European Parliament. Directorate General for Parliamentary Research Services., cit.

[190] European Parliament. Directorate General for Parliamentary Research Services., cit.

[191] C. Goujard-G. Volpicelli, ChatGPT is entering a world of regulatory pain in Europe, POLITICO, 2023 https://www.politico.eu/article/chatgpt-world-regulatory-pain-eu-privacy-data-protection-gdpr/

Consequent upon the establishment of ethical principles that are considered standard, it is necessary to adapt the various regulations, which can ensure compliance with the ethical principles thus determined. These regulations should also be updated to the latest research developments in AI systems so as to be able to promote responsible and ethical AI that extends throughout the market without exception. Constructive and concrete policies need to be formulated in this area that are able to address the challenges that artificial intelligence poses to us. It would also be important to develop independent audit systems, linked to the aforementioned regulations, to ensure that the various artificial intelligence applications comply with ethical principles by enabling the assessment of compliance with existing regulations and principles. In this regard, it would also be useful to hold companies that develop and use AI systems accountable through ad hoc regulations that can target corporate social responsibility (CSR).

Regarding regulation, IBM has proposed a governance approach based on the three pillars of accountability, transparency and fairness, and security. Regarding accountability, IBM proposes greater differentiation in order to better mitigate any damages[192]. There are five proposals from IBM for implementation within companies and three pieces of advice addressed to policy makers. According to the company, within organizations, it is critical to designate an artificial intelligence ethics manager to oversee risk assessment and mitigation strategies, so as to not only mitigate risks but also increase people's trust in AI[193]. The second advice is to adopt different rules for different risks, thus employing a risk-based classification [194]. IBM believes that a key principle for AI governance is precisely a risk-based approach, in that the harms and obligations involved should be proportionate. High risk will therefore require high standards, and that is precisely what the European AI Act provides for.[195] The third imperative is related to transparency, as it generates public trust. Therefore, companies should disclose the operation and purpose of an AI system, without having to reveal trade secrets[196]. IBM also advises explaining its artificial intelligence, maintaining audit trails surrounding both inputs and data used for training, and making essential information known to consumers regarding, for example,

[192] R. Haghemann-J.-M. Leclerc, Precision regulation for artificial intelligenze, IBM, 2020 https://www.ibm.com/policy/ai-precision-regulation/
[193] R. Haghemann-J.-M. Leclerc, cit.
[194] R. Haghemann-J.-M. Leclerc, cit.
[195] C. Montgomery-F. Rossi-J. New, A Policymaker's Guide to Generative AI, IBM Newsroom, 2023 https://newsroom.ibm.com/Whitepaper-A-Policymakers-Guide-to-Foundation-Models
[196] R. Haghemann-J.-M. Leclerc, cit.

levels of regularity and privacy measures. The final piece of advice for companies is to put AI to the test with regard to bias[197]. To get to the point of ensuring this, IBM suggests governments recognize existing and efficient co-regulatory regulations to unite the various parties and promote a joint effort, support funding and development of AI testbeds in controlled environments, and finally provide incentives for those who would voluntarily adhere to recognized standards[198].

There may also be technological tools that help to have ethical AI. With regard to so-called "black box" algorithms, for example, systems are being developed that can understand backwards the AI's decisions and then explain how a particular decision was arrived at by exposing the inner workings, an example being the Layerwise Relevance Propagation program developed by a machine learning professor, Klaus-Robert Müller with his team[199]. An alternative solution to opening the black box is one developed by Sandra Wacher, Brent Mittelstadt, and Chris Russell, who are working on trying to understand what would be required to make the decision made by the algorithm change[200].

However, responsible use of systems must be preceded by responsible design. According to many, responsible AI is based on human-centered design, thus focusing first on the human experience and then aligning with business goals. The AI For Good Foundation proposes five steps proper to lean design that can be useful for the early stages of an AI system life cycle in a human-centered design context.[201] The first step is empathy, it is necessary to empathize with users but also with those who will be indirectly affected by this application, perhaps conducting interviews and having an open dialogue. The second step suggests conducting inclusive research to best understand the environment in which AI will operate, in this context it will be important to understand the current environment but also to anticipate the impact of AI[202]. The third step focuses on human interaction and presence; in fact, a feedback loop between humans and the AI system is often required. The fourth step is based on prototyping and iteration. When one wants to develop responsible AI systems, it is indeed good not only to prototype, but also to plan for the

[197] R. Haghemann-J.-M. Leclerc, cit.
[198] R. Haghemann-J.-M. Leclerc, cit.
[199] I. Sample, cit.
[200] I. Sample, cit.
[201] C. Lamoutte, Responsible AI Begins with Human-Centered Design - AI for Good Foundation, 2022 https://ai4good.org/blog/responsible-ai/
[202] C. Lamoutte, cit.

possible errors that the machine might make. The fifth and final principle proposed is to promote multidisciplinary teams; indeed, a diverse set of skills, knowledge but also perspectives is needed to succeed in developing a responsible AI solution[203].

Even big players in the field like Google are using a human-centered design approach. In fact, Google envisions several recommendations to get to the point of having an accountable AI system.[204] In terms of general recommendations, the company recommends using the human-centric design approach; identifying many different metrics to evaluate training and monitoring; examining raw data whenever possible; understanding the limitations of the model you are going to develop; focusing on testing; and continuing to monitor/update the AI system even after its deployment[205]. Regarding specific recommendations concerning fairness, Google recommends that you design your model using concrete fairness and inclusion goals; use inclusive and representative datasets; check the system for the presence of bias; and analyze its performance[206]. Related to the interpretability aspect, it is recommended to consider this principle as a fundamental component of the user experience; plan options to pursue it; design the model to be interpretable; choose metrics that reflect the end goal; communicate explanations of the model to users; and conduct various tests[207]. Google also offers advice related to privacy and security. Regarding confidentiality, it recommends collecting and managing data responsibly perhaps by using a public data source, or encrypting or even anonymizing and aggregating the data; and adequately safeguarding the privacy of machine learning models[208]. Related to security, the company recommends identifying potential threats to the system; developing an approach plan to minimize threats; and finally continuing to learn, continually updating to keep up with the times[209].

The Organization for Economic Co-operation and Development (OECD) has also created recommendations, particularly for policy makers, and has also identified five value-based principles, again taking a human-centric approach. The OECD's recommendations addressed to policy makers are: invest in AI research and development, promote a digital

[203] C. Lamoutte, cit
[204] Google AI, Responsible AI practices, Google AI, https://ai.google/responsibilities/responsible-ai-practices/
[205] Google AI, cit.
[206] Google AI, cit.
[207] Google AI, cit.
[208] Google AI, cit.
[209] Google AI, cit.

ecosystem for AI, provide an enabling policy environment, develop human capacity and prepare for labor market transition, and finally cooperate internationally to get to responsible AI[210].

As can be observed, a widely recommended technique at the level of institutions and organizations is that of human-centered design. To implement it successfully, attention must be paid to the needs, circumstances, skills and capabilities of the people who will later become the users of the AI system.[211]

A list of methodologies for achieving reliable AI is also provided to us by the Independent Group of High-Level Experts on Artificial Intelligence established by the European Commission. First of all, that expert group indicates how it is crucial to ensure throughout the lifecycle of an artificial intelligence system that the requirements for reliable AI are met, namely: human intervention and oversight; technical robustness and security; confidentiality and data governance; transparency; diversity, nondiscrimination and equity; environmental and social well-being; accountability[212]. This is in order to arrive at the goal of reliable AI, which must conform to the principles of legality, ethicality and robustness, technical methods and non-technical methods are provided that succeed in ensuring the implementation of the requirements just listed. The technical methods are as follows[213]:

- Architectures for reliable AI. The requirements mentioned should be translated into the system architecture's own procedures; in this regard, it may be useful to draw up a list of rules that the system must always comply with, with associated monitoring. If the system is running on machine learning, it is necessary to inculcate the requirements in the three phases of the cycle i.e., perception (in which all elements of the environment necessary to adhere to the principles must be recognized), planning (in which plans aligned with the

[210] OECD, cit.

[211] D. Leslie, Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, Zenodo, 2019

[212] C. and T. (European C. Directorate-General for Communications Networks-Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Orientamenti etici per un'IA affidabile, Publications Office of the European Union, LU, 2019

[213] European C. Directorate-General for Communications Networks, cit.

requirements should be considered), and action (actions may be limited to only those that take into account and implement the principles)[214]

- Ethics and rule of law right from the design (X-by-design): it is necessary to integrate the values defined from the design of AI and it is therefore the responsibility of the company to implement them from the beginning[215].
- Explanatory methods: in order to achieve a reliable system, it must also be explicable so that the mechanisms involved in its operation can be understood. This point is definitely still a challenge especially for black box algorithms but it is necessary if you want to develop a technology that is reliable[216].
- Testing and validation: traditional testing and validation methods are not effective given the nondeterministic nature of AI systems. Errors emerge only as the AI faces realistic data, so constant monitoring is needed both during training and after deployment. Testing is needed as early as possible to ensure proper operation throughout the life cycle. To ensure a comprehensive evaluation, it is advisable to involve a diverse group of people and develop metrics that have different perspectives[217].
- Service quality indicators: indicators can be defined to measure the quality of the system, also taking into account safety [218].

Regarding non-technical methods, which still play a relevant role and need to be evaluated continuously, the following are identified:

- Current regulations
- Codes of conduct: guidelines may also be included indicators within codes of conduct or internal policy [219]
- Normalization: standards, such as design standards, can serve as a quality management system for users and organizations, enabling ethical behavior to be promoted and recognized. In the future, it may prove useful and interesting

[214] European C. Directorate-General for Communications Networks, cit.
[215] European C. Directorate-General for Communications Networks, cit.
[216] European C. Directorate-General for Communications Networks, cit.
[217] European C. Directorate-General for Communications Networks, cit.
[218] European C. Directorate-General for Communications Networks, cit.
[219] European C. Directorate-General for Communications Networks, cit.

to introduce as a label "Trustworthy AI" (if the system meets specific standards)[220].

- Certification: certifications are useful to make the characteristics of the system understandable even to non-experts, such as whether it is transparent or not. Accompanying the certification, however, is also the need for an accountability system[221].
- Accountability through governance frameworks: it is useful to provide accountability frameworks within organizations, through appointments of officers and committees [222].
- Education and outreach to promote an ethical mindset: these practices are key to educating the public about the direct and indirect effects of artificial intelligence but also to engage multiple parties who can work together to influence its development. Appropriate ethics training by experts in this field also contributes to public education[223].
- Stakeholder participation and social dialogue: there is a need for open dialogue and broad involvement of all stakeholders as well as the public [224].
- Diversity and inclusive design teams: an inclusive team also helps to reflect diversity and thus the different needs of the audience, thus contributing to objectivity [225].

It is also critical to understand that for each stage of developing an AI system, different strategies that complement each other are useful. Regarding the first stage of design, frameworks i.e., structures of concepts that can serve as a skeleton and guide for ethical concepts can be useful; regarding the testing stage, on the other hand, metrics and audit systems (formal examinations of ethical aspects) are useful. If reference is made instead to system information, declarations (describing algorithms and systems providing useful information for evaluating ethical aspects), labels and licenses can be adopted.[226]

[220] European C. Directorate-General for Communications Networks, cit.
[221] European C. Directorate-General for Communications Networks, cit.
[222] European C. Directorate-General for Communications Networks, cit.
[223] European C. Directorate-General for Communications Networks, cit.
[224] European C. Directorate-General for Communications Networks, cit.
[225] European C. Directorate-General for Communications Networks, cit.
[226] E. Prem, From ethical AI frameworks to tools: a review of approaches, in AI and Ethics, 2023

In conclusion, the development of ethical and responsible artificial intelligence systems represents a task that requires commitment on many fronts and from many sides. Starting from algorithm design and development to data management, every aspect must be oriented toward ethical rigor. The strategies and tools listed are varied and diverse, but they all lead to the same goal and complement each other. Thus, there is no one-size-fits-all solution that succeeds in ensuring the development of AI systems that are ethical and sensitive to human and moral values, but there is a need for sustained efforts on the part of all actors involved. It is also important to remember that AI-related technologies are constantly developing, which implies that strategies and tools must also be continuously updated.

Chapter 3. The ethical value of kindness

3.1. Definition of kindness

Kindness is defined by Cambridge vocabulary as "the quality of being generous, helpful and attentive to others, or an act that demonstrates this quality" and is "synonymous with politeness, friendliness, courtesy or amiability" according to the Treccani dictionary. It is a concept we all know the value of and has roots as far back as the age of the ancient philosophers. The Greek philosopher Aristotle, in fact, defined the concept of kindness as "Welcoming someone in need, in exchange for nothing, neither for the benefit of the helper himself, but for that of the person helped."[227]. Later Marcus Aurelius, an emperor and a member of the philosophical current of Stoicism, defined kindness as "the greatest delight of mankind," and Friedrich Wilhelm Nietzsche, a German philosopher, described kindness and love as "the most healing herbs and agents in human relations." Kindness was also considered one of the so-called chivalrous virtues.

Kindness can also be considered one of the virtues of virtue ethics, since it is an excellent character trait that can guide ethical decisions. People who are considered kind, in fact, act that way because they believe it is the right way to behave, without expecting anything in return. Virtue ethics (known also as natural ethics) constitutes one of the three main approaches of what is considered normative ethics, and is a moral theory that focuses on virtues, that is, qualities of character, rather than rules or outcomes of actions[228]. This ethical position is rooted in pre-Enlightenment Aristotelianism. For the purposes of this theory, it therefore turns out to be important what kind of person one is, virtues in fact reflect a way of being and are not merely external behaviors[229]. This development of character is also associated with perfectionism, in that a person over time should develop and hold more and more to that which would be the Perfect self (in religious terms thus becoming divine)[230]. Moreover, according to ancient philosophers, including Aristotle, happiness is the cardinal principle of ethics, and to achieve it is possible to transform the

[227] L. Honeycutt, Book II - Chapter 7 : Aristotle's Rhetoric, 2004 https://web.archive.org/web/20041213221951/http://www.public.iastate.edu/~honeyl/Rhetoric/rhet2-7.html

[228] R. Hursthouse-G. Pettigrove, Virtue Ethics, in E. N. Zalta, U. Nodelman (a cura di), The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, 2022

[229] R. Hursthouse-G. Pettigrove, cit.

[230] E. Kazim-A. S. Koshiyama, cit.

use of our intellectual and reasoning faculties into true virtues, precisely including the manifestation of kindness[231].

The origin of the word gentile originates from Latin, where "gentilis" indicated belonging to a "gens" (i.e., lineage), a membership that determined social class and the "gentiles" corresponded to the nobles. This connection to the highest strata of society remained even in the Middle Ages, today's concept of gentility was in fact still linked to aristocrats, called "gentiles," and later in the 13th century it spread to social classes that imitated aristocrats, such as merchants. The word "gentile" is still linked to elevated behavior, of the rules of behavior in society, and it can be said that in the past it constituted an imitative factor of what were considered the higher classes, hence the saying "a person of class".[232]

The concept of kindness may not be easy to define since it takes on different meanings and forms for each of us. It is also considered a concept related to wisdom, which has an ethical dimension and is included in what a person can set as an ethical ideal, at the roots of which there must be a deep moral sense as Johann Wolfgang Goethe, German writer and philosopher, wrote[233]. The meaning of kindness therefore has been preserved from the ancients to the contemporary age, symbolizing a form of moral self-determination that is based on sympathy. There is an enhancement of the individual entity, but also a search for intersubjectivity (i.e. valuing the other person's point of view).[234]

This concept, in Western history, is closely linked to Christianity, which in fact holds sacred expressions of generosity such as charity, love and selflessness. In fact, kindness is one of the main themes of the Bible[235]. The Christian concept of charity played the role of cultural glue for centuries, challenged only by the 16th century with the development of individualism [236]. For Christians, in fact, kindness represents the fifth element that

[231] Le virtù etiche in Aristotele | Platone 2.0 – La rinascita della filosofia come palestra di vita. https://www.platon.it/storia/il-bene-e-la-felicita/perche-e-cosi-importante-la-virtu/le-virtu-etiche-in-aristotele/

[232] A. Iasimone, Gentilezza: significato, potere e storia a cura del prof. Canettieri, Scuola di Comunicazione Gentile, 2021 https://www.comunicazionegentile.it/gentilezza-significato-potere-storia-paolo-canettieri/

[233] G. Brunetti, L'etica della gentilezza e la sua funzione terapeutica, Riflessioni.it, 2021 https://www.riflessioni.it/finestre-anima/etica-della-gentilezza-e-sua-funzione-terapeutica.htm

[234] R. Roni, La gentilezza verso l'altro parte dalla vita interiore personale, Blasting News, 2022 https://it.blastingnews.com/opinioni/2022/09/la-gentilezza-verso-l-altro-parte-dalla-vita-interiore-personale-003565158.html

[235] G. Brunetti, cit.

[236] A. Phillips-B. Taylor, Sulla gentilezza, Internazionale, 2018 https://www.internazionale.it/notizie/adam-phillips/2018/05/31/gentilezza

Scripture lists as the fruit of the spirit (Galatians 5:22) and they desire to be kind since God has been merciful and kind to them[237].

In modern philosophy, this concept is always present but sometimes in the form of different words, which, however, may refer to the same meaning. This is the case with Spinoza, who in his 1677 work "Ethics" defines sympathy as "love toward one who has done good to another," demonstrating how "we feel well disposed toward one who has done good to a thing similar to us," and further states that "if someone has done something that he imagines arouses gladness in others, he will be affected by a gladness accompanied by the idea of himself as the cause; that is, he will regard himself with gladness."[238]. What Spinoza describes is overlaid on a concept of kindness and is put into practice with the concept of politeness, called humanitas (meaning benevolence, politeness and civility in Latin). In eighteenth-century philosophy, philosopher Adam Smith in his 1759 work "Theory of Moral Sentiments" lists politeness as one of the social passions. According to Smith, social passions are made pleasant by so-called "sympathy doubled," that is, generosity, compassion, benevolence, esteem and friendship[239]. For the philosopher, the word sympathy expressed a concept that could be likened to kindness, as it meant the ability to identify with the other person to understand his or her feelings, and also to empathy. Smith also uses the word kindness directly, saying that nature created humans with the purpose of this mutual kindness, so that each person is the object of kindness to those people toward whom he or she has been kind[240]. The German philosopher Nietzsche, mentioned earlier, considers kindness as a noble form of altruism, which, however, departs from the simple concept of compassion, and identifies in the attitudes of courtesy (typical of noble men), the exact opposite of resentment[241].

If for Marcus Aurelius, emperor and philosopher, kindness symbolizes a guiding principle of life and a winning weapon against the ignorant and malevolent, can the same be said nowadays? Marcus Aurelius was also convinced that kindness was a rewarding behavior and an unconditional feeling, to be exhibited even in the face of one's enemies, that one

[237] D. D. Delzell- collaboratore di C. Post, The difference between natural kindness and Christian kindness, The Christian Post, 2022 https://www.christianpost.com/voices/the-difference-between-natural-kindness-and-christian-kindness.html

[238] R. Roni, cit.

[239] R. Roni, cit.

[240] R. Roni, cit.

[241] R. Roni, cit.

should seek neither reward nor fame. Many say that today we live in a competitive and individualistic world, in which kindness is now a rare gift and may even be viewed with suspicion.[242] Indeed, there are those who say that people are now driven by selfishness and that kindness has become a kind of forbidden pleasure, risky since it is closely linked to sensitivity, but it is still a gift considered fulfilling.[243] Nowadays there is a strong need for humanity and therefore also for kindness. This need is also coming to the fore, for example, in the medical field, where some argue that in order to heal, it is necessary to arm oneself with empathy and communication as well, bringing back an ethics of medical education that stands in contrast to the trend of the "dehumanization" of this subject[244].

The concept of kindness nowadays is related to other feelings and words that can capture its different nuances such as: empathy, that is, treating people as they would like to be treated; compassion; humanity; altruism, solidarity; benevolence; listening; openness, that is, removing all barriers both mental and physical as well as prejudices; caring[245]. Looking back to the past, these terms were specifically recognized as caritas, or love of others, and philanthropy, or love of humanity [246].

Kindness is thus considered a social, moral and human value, a quality therefore inherent within us and which can be composed of several levels, at a basic level it is related to consideration of others, which is related to the struggle for survival and well-being, common traits of human beings[247].

Kindness also is part of the so-called soft skills, the skills that do not have to do with technical knowledge but rather with being able to relate, being creative, being able to solve complex situations, and so on. According to research conducted by IBM, these skills are desperately sought after by all CEOs and are not as easy to learn as technical skills. With the development of new technologies, many "no-routine" occupations have also been created, growing 25 times more than routine jobs from 1976 to 2014, which require intense human skills that technologies and automation cannot replace. Soft skills thus

[242] S. Primiceri, Marco Aurelio e il culto della gentilezza, L'AltraPagina.it, 2017 https://www.laltrapagina.it/mag/marco-aurelio-e-il-culto-della-gentilezza/

[243] A. Phillips-B. Taylor, cit.

[244] G. Brunetti, cit.

[245] Accademia della Gentilezza, Parole e Comportamenti gentili, Accademia della Gentilezza, 2022 https://www.accademiadellagentilezza.it/words/

[246] A. Phillips-B. Taylor, cit.

[247] M. Karlin-B. Ozawa-De Silva, The Science, Theory and Practice of Kindness: A Brief Overview, UNESCO MGIEP https://mgiep.unesco.org/article/the-science-theory-and-practice-of-kindness-a-brief-overview

become crucial in today's scenario, which is complex, fast-paced and constantly changing. Both ethics and kindness are part of these skills, which actually seem more like virtues, skills that are defined as experiential (i.e., shaped through practice) and difficult to measure, but which should not be put on the back burner for this reason and indeed are necessary in today's context. [248]

3.2. Kindness as a social value

Human beings are social animals by nature, who also need the kindness of others to survive; we are not born and grow up for the first few years self-sufficient; in fact, we remain interdependent beings. This characteristic also influences our biology, such as our nervous system[249].

A study conducted by the United Nations, "The World Happiness Report," shows that rather than economic factors, it is trust and social support that give happiness and satisfaction in life. These two key factors are manifestations of kindness, which can therefore be inferred to be directly related to people's happiness, and indeed it emerges that the kindest societies are also the happiest societies[250].

Kindness not only makes other people happier, but it also makes us happy ourselves both because of the consequences it has on certain brain areas but also because of the social effects it can spill over, as indeed was proven by a 2017 study from the University of Zurich and Northwestern University in Chicago[251]. According to other studies, it has also been observed that people who practice kindness, gratitude and are optimistic possess longer telomeres (sections of DNA that shorten with age) than those who do not. These practices, therefore, succeed in counteracting oxidative stress and inflammatory processes in our bodies, and it can be said that in a sense, kindness makes us live healthier and longer, protecting our DNA[252]. The data also disprove the Darwinian paradigm that the fittest are

[248] R. Zazza, Blog | Metti etica e gentilezza nel curriculum e il lavoro del futuro è tuo, Alley Oop | Il Sole 24 Ore, 2022 https://alleyoop.ilsole24ore.com/2022/11/18/metti-etica-gentilezza-nel-curriculum-lavoro-del-futuro/

[249] M. Karlin-B. Ozawa-De Silva, cit.

[250] M. Karlin-B. Ozawa-De Silva, cit.

[251] S. Q. Park-T. Kahnt-A. Dogan-S. Strang-E. Fehr-P. N. Tobler, A neural link between generosity and happiness, in Nature Communications, fasc. 8, 1, 2017, p. 15964

[252] A. Zampa, Biologia della gentilezza: 6 scelte per benessere, salute e longevità, LifeGate, 2020 https://www.lifegate.it/biologia-della-gentilezza-libro-interviste-lumera-de-vivo

the strongest to survive; in fact, the fittest have been shown to be those who are kindest. Being kind but also happy and optimistic thus represents the best evolutionary strategy for survival, contrary to what many people think these values in a competitive society should be interpreted as a symbol of weakness.

Kindness represents a fundamental prosocial value, that is, promoting social welfare, which can be manifested through behaviors aimed at benefiting other people. Indeed, it also represents a key element in not only building but also maintaining good social relationships. Kindness as prosocial behavior is motivated by the value of benevolence, often associated with empathy and compassion[253]. Connections with other values are possible, and these connections can be useful in understanding how our behavior in social situations is affected but also how to promote kindness in daily living[254]. Indeed, it is possible to promote a culture of kindness in the community by going out and identifying the values that motivate prosocial behavior (such as benevolence), promoting them, and creating social norms that encourage them[255].

Thus, kindness proves to be crucial both biologically and socially, we should in fact invest in new ways or tools that can cultivate kindness [256]. In recent years in this regard, a number of special initiatives and programs have sprung up, including the "Accademia della Gentilezza", a nonprofit organization that aims to spread kindness throughout the country system and key sectors.

3.2.1. Kindness in the digital age

Although kindness has represented a moral value since ancient times, its presence is also essential in the digital age in which we live. By now, online interactions have become part of our daily lives, and kindness is also declined or not declined in this context. If in the real world politeness is related to "good behavior," its declination in the online world goes further, also implying empathy, respect and understanding toward the other in what is considered a digital environment. This includes responding courteously on social media,

[253] R. Sanderson-J. Mcquilkin, Many Kinds of Kindness: The Relationship Between Values and Prosocial Behaviour, in Values and Behavior: Taking a Cross Cultural Perspective, 2017, p. 75–96

[254] R. Sanderson-J. Mcquilkin, cit.

[255] R. Sanderson-J. Mcquilkin, cit.

[256] M. Karlin-B. Ozawa-De Silva, cit.

promoting constructive dialogue and avoiding offensive language. Compared to real interaction, a digital interaction is more difficult to intrepret, and in addition, sometimes people hide behind anonymity, which can give rise to negative behavior. The important thing would be to encourage dialogue and understanding, trying to create a positive digital environment that can prevent problems such as cyberbullying.

Social media have indeed transformed the way we interact online and represent for most of us a form of expression and sharing of ideas but also a place for community building. These places, however, can be vehicles for negativity and conflict. In this sense, kindness within social media can certainly create a constructive and positive environment by encouraging dialogue and promoting understanding. Expressions of kindness in this context can take many forms, such as appreciations, sharing useful information, or offers of support. These small gestures can have a positive impact on both the individual but also with regard to the community. Among the different ways in which online kindness can be shown is by treating others as one would like to be treated, that is, according to the so-called golden rule of good manners[257]. You can then practice random acts of kindness online, which can take the form of a message, words of encouragement, a compliment or a positive review. It is also important in a digital environment to adopt a positive tone by always taking a moment before posting, always thinking that beyond the screen there is always a person, and it is also appropriate to avoid online drama[258].

Promoting kindness in a digital age is an important challenge that can have a great positive impact. Several initiatives have sprung up in this regard. One of these is the legaltech platform "Chiodiapaga," which fights online hate by helping people who are victims of it and helping them get just compensation[259]. In fact, online hate crimes are numerous, such as stalking, cyberbullying, revenge porn, defamation, and hate speech. The startup thus settles into a digital world in which out of 215,000 Twitter posts analyzed by research, about 70 percent turned out to be hate speech, aiming to spread a culture that aspires to safer and more aware use of digital platforms[260]. Another startup that aims to spread

[257] M. Fox, Be kind online, 2021 https://www.bupa.com.au/healthlink/mental-health-wellbeing/mental-health/be-kind-online
[258] M. Fox, cit.
[259] C. Zaccarelli, Per la Giornata mondiale della gentilezza, ecco 8 startup che ne hanno fatto la loro filosofia, LifeGate, 2022 https://www.lifegate.it/startup-giornata-mondiale-della-gentilezza
[260] C. Zaccarelli, cit.

kindness is Hi!Founders, which through a digital platform provides an online space for innovators to exchange advice and help, without receiving anything in return[261].

In 2018, a study was carried out on cyber-gentility, which showed that it is a widespread and distinct phenomenon, capable of succeeding in the future over bullying [262]. The authors state that this should not be surprising, as the digital world is a reflection of human nature and is a product of human endeavors; in fact, since kindness is widespread in the real world, it should be so online as well. A previous study was able to distinguish three main aspects of human kindness, distinguishing between benign tolerance, empathic responsiveness, and social proaction, and it was shown that all of these components can also be identified in online behaviors[263]. The prevailing behaviors in cyber space turn out to be those of benevolent tolerance, by far greater than acts of kindness considered proactive. Indeed, benevolent tolerance is easier to exercise and requires little moral involvement[264]. It also appears that those who perform acts of cyber kindness are really motivated by purposes directed toward others, with personal benefit playing virtually no role, and there is a positive correlation between offline and online acts of kindness[265].

In recent years, in the digital world thanks to the rise of social media, a new figure has emerged, namely the influencer. Influencers, by leveraging their large following and popularity, have gained wide influence over the public, with the ability to shape public opinions in fact. Their figure obviously does not stop at promoting products and services, but their opinions also go a long way in influencing their following. One of the areas in which their influence could be considered particularly positive is certainly in promotions of kind behavior in the digital world. A 2023 study explores precisely this potential positive impact on the spread of kindness and introduces the concept of "kindness contagion"[266]. The study states that in order for kindness contagion to occur, underlying relationship building is essential, and from this also comes greater beneficial effects.

[261] C. Zaccarelli, cit.
[262] L. Rowland-D. Klisanin, Cyber-kindness: Spreading kindness in cyberspace, Media Psychology Review, 2018 https://mprcenter.org/review/cyber-kindness-spreading-kindness-in-cyberspace/
[263] L. Rowland-D. Klisanin, cit.
[264] L. Rowland-D. Klisanin, cit.
[265] L. Rowland-D. Klisanin, cit.
[266] T. Bradley-K. C. Anderson-A. Hass, The Virtuous Cycle: Social Media Influencers' Potential for Kindness Contagion, in Journal of Macromarketing, fasc. 43, 2, 2023, p. 110–118

However, contagion is null and void if the kind act is not visible to others[267]. As far as social media is concerned, it can be said that the relationships established, for example, between followers and influencers are parasocial relationships; there is no real interaction present, but only a one-sided relationship that, however, allows the follower to feel, in quotes, a friend of the influencer and thus to be easily influenced by him[268]. However, this power of influencers is contingent on authenticity, authenticity that manifests itself in being true to oneself and building trust. Influencers, in fact, can leverage the relationship of trust with their followers, and their impact could easily be used to spread kindness online. The contagion of kindness, according to this study, should be based on four main elements that are part of a cycle: authenticity of the influencer, minimization of parasocial distance, minimization of the level of construction, and internalized kindness[269]. Although, therefore, the figure of influencers is sometimes juxtaposed with the negative aspects of social media that can lead to anxiety and depression, they could be a key figure in promoting the contagion of kindness online, thus creating a significant positive impact[270].

As in the case of the influencers just mentioned, digital platforms can help to spread acts of kindness. Such is the case with a platform, called the MIT Community Challenge, which is designed to study and encourage acts of kindness[271]. This project came about as a result of concerns about well-being and mental health on the MIT campus, thus giving rise to a web platform that succeeds in facilitating donations, which are found to benefit personal well-being, and succeeds in providing insights into how prosocial behavior can be facilitated. It has been found that 77 percent of MIT undergraduate students think that the university environment negatively affects their mental well-being, and at the same time it is known that kindness, particularly the act of giving, has positive effects on individual well-being[272]. Combining the two was the development of a platform aimed at the student community whose goals are to encourage acts of kindness and at the same time collect data on relevant variables (nature of the act, nature of the recipient, and frequency of reciprocity). The platform poses challenges to users in the form of acts of kindness to be

[267] T. Bradley-K. C. Anderson-A. Hass, cit.
[268] T. Bradley-K. C. Anderson-A. Hass, cit.
[269] T. Bradley-K. C. Anderson-A. Hass, cit.
[270] T. Bradley-K. C. Anderson-A. Hass, cit.
[271] I. Ananthabhotla-A. Rieger-D. Greenberg-R. Picard, MIT Community Challenge: Designing a Platform to Promote Kindness and Prosocial Behavior, in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2017, p. 2352–2358
[272] I. Ananthabhotla-A. Rieger-D. Greenberg-R. Picard, cit.

performed for another person, and in turn encourages the recipient to repay him or her by engaging in a challenge himself or herself[273].

In conclusion, kindness promoted online is a key aspect of digital communication, and the promotion of kind and positive attitudes can help counter negative and increasingly common phenomena such as trolling, which is the posting of provocative or misleading comments to elicit negative reactions, and cyberbullying, by helping to create a positive and supportive climate.

3.2.1. The example of kind leadership

Leadership represents a concept that has undergone several evolutions throughout history, and has been represented over time by various models, starting as far back as the ancient Greeks who considered the classic hero of mythology to be the perfect leader. Aristotle, speaking of leaders and people at the top, spoke of values and virtues, thus commanding with an education aimed at moral and ethical[274]. This was followed in the 500s by the leadership proposed by Macchiavelli, which was unscrupulous and according to which "the end justifies the means." From the more traditional theories that focused on authority and firmness, we moved to a more inclusive model of leadership: kind leadership. Gentle leadership is characterized by respect and empathy and is defined as "the mix of skills that characterizes a leader capable of guiding a group of people toward achievable goals, helping them make the best decisions."[275]. In this context, therefore, kindness is not seen as a sign of weakness, but rather as a tool that can build a healthy work environment and achieve successful results. A key feature of this type of leadership is also that it puts the person back at the center, emphasizing values such as listening, accompanying and welcoming[276].

During the Coronavirus pandemic, our lives have been all but suspended and major changes have taken place in our daily lives and routines. It is precisely in this context that it becomes essential to make small gestures that help others as well as ourselves. The

[273] I. Ananthabhotla-A. Rieger-D. Greenberg-R. Picard, cit.
[274] G. L. Teppati, Leadership gentile, Manageritalia, 2022 https://www.manageritalia.it/it/management/leadership-gentile
[275] F. Giffoni, Leadership gentile: caratteristiche e peculiarità, 2019 https://qsfera.it/blog/leadership-gentile-caratteristiche-e-peculiarita
[276] G. L. Teppati, cit.

problem of being able to express kind gestures despite social distancing also emerged during the pandemic, and furthermore, these problems were also reflected on organizations. Leadership models and more generally the organizational and operational modes of companies were challenged during this period, there was more focus on smart working and consequently the autonomy that comes with it[277]. In this scenario, the concept of gentle leadership, a model that places attention on trust and respect for talent, putting special emphasis on interpersonal relationships, is increasingly gaining ground. All this while always pursuing the goals set. Such a leadership model, by putting people at the center, succeeds in creating a positive and peaceful work climate, where employees can feel autonomous and empowered in their tasks[278]. The kind leader is thus able to see and value the talents of his or her employees, accompanying them on their career path and ensuring that they succeed in fulfilling their potential [279]. The leader thus assumes a role that can be likened to a facilitator, a figure that requires characteristics such as sensitivity, the ability to engage, innovate, and motivate [280].

One study analyzed and identified the competencies deemed most important regarding the emerging model of gentle leadership, confirming that it is a different and more evolved style than the democratic, transformational, and charismatic styles[281]. The three skills that stand out over the others are: teamwork, that is, encouraging teamwork and participation by allowing people to voice their opinions; motivation, in fact encouraging initiative and innovation by all employees; risk-taking, without fear of failure since every mistake is a lesson[282].

To date, gentle leadership is a much-discussed topic as we are realizing that in the current scenario it is the model that really succeeds in bringing about value creation, and its emergence has been accelerated by the advent of the pandemic, which, like all crises, has also been a driver of change[283]. It was not only the pandemic that accelerated the advent of this new leadership model; in fact, one must also mention sustainable digital

[277] G. Stratta, Leadership gentile: il nostro nuovo "way to be", 2021
https://www.enel.com/it/azienda/storie/articles/2021/05/leadership-gentile-nuovo-modo

[278] G. Stratta, cit.

[279] G. Stratta, cit.

[280] N. Spagnuolo, Leadership gentile: moda o cambiamento reale? - Il Sole 24 ORE, 2022
https://www.ilsole24ore.com/art/leadership-gentile-moda-o-cambiamento-reale-AE5RzwiB?refresh_ce

[281] A. Yela Aránega-C. Gonzalo Montesinos-M. T. del Val Núñez, Towards an entrepreneurial leadership based on kindness in a digital age, in Journal of Business Research, fasc. 159, 2023, p. 113747

[282] A. Yela Aránega-C. Gonzalo Montesinos-M. T. del Val Núñez, cit.

[283] N. Spagnuolo, cit.

transformation, which is pushing companies to find new ways to create value[284]. In today's highly uncertain environment, greater flexibility is also needed in the business environment, and creativity must be used to develop new organizational models[285]. Thus, there is a shift from a paradigm of an employee as a productive factor to an employee who serves a driving function for the company [286].

Kindness within leadership models, and especially with regard to human resource management, can generate an important impact as it can unlock the high potential that resides in employees; this definitely leads to better performance and creates a competitive advantage[287]. This way of doing things also creates high trust, kindness being a key element for trustworthiness as well as increased employee commitment. To this end, strong emotional intelligence on the part of leaders is essential[288].

Another foundational element of gentle leadership is undoubtedly trust, a resource that is in short supply within companies but can lead to better performance, better decisions, and lower absenteeism rates. Therefore, it would be ideal for a company to orient its processes toward trust so as to create a positive climate while maximizing results[289].

Kindness and ethics in the work environment are key elements, as they can lead to the establishment of healthy and transparent relationships with our stakeholders and colleagues[290]. Conversely, a lack of attention to these principles can backfire, leading to harm to the person or to the company that does not consider them. The role of kindness should not be juxtaposed with limited cleverness, but rather with something positive, which reconnects us to our humanity and can lead to success. The benefits of kindness are not only measured on a human level in the work context, but the presence of this work also turns into profitability[291]. Where considered, kindness brings economic benefits

[284] A. Yela Aránega-C. Gonzalo Montesinos-M. T. del Val Núñez, cit.

[285] N. Spagnuolo, cit.

[286] N. Spagnuolo, cit.

[287] C. Caldwell, Understanding Kindness – A Moral Duty of Human Resource Leaders, in The Journal of Values-Based Leadership, fasc. 10, 2, 2017

[288] C. Caldwell, cit.

[289] G. Stratta-P. Cervini, Nuovi modelli di gestione: l'evoluzione verso la leadership gentile, 2022 https://www.hbritalia.it/giugno-2022/2022/05/31/news/nuovi-modelli-di-gestione-levoluzione-verso-la-leadership-gentile-come-le-aziende-possono-coltivare-una-fiducia-vera-il-caso-di-enel-15294/

[290] D. Rampado, ETICA E GENTILEZZA NEL LAVORO « Donatella Rampado, Donatella Rampado, 2019 https://www.donatellarampado.com/etica-e-gentilezza-nel-lavoro/

[291] G. Dotto Pagnossin, L'impronta gentile. Attraversare la vita in punta di piedi. Vale anche per il relatore pubblico, 2022 https://www.ferpi.it/news/limpronta-gentile-attraversare-la-vita-in-punta-di-piedi-vale-anche-per-il-relatore-pubblico

within the company, and is closely linked to costs. In a 2019 study by Cisco, it was estimated that the time an employee wasted thinking about an unpleasant colleague and the resulting stress had an $8 million burden on company accounts[292]. In further research also conducted in 2019 by Inail and Bocconi University, business performance and the character of leaders were compared. It emerged that companies with kind leaders had had higher profitability[293]. Further U.S. research conducted by Jane Dutton and Monica Worline of the Ross School of Business analyzed forty companies in the financial sector over two years, and examining their performance, it was found that when compassion was part of corporate values, financial performance was higher, and in addition, customer but also employee retention was also raised.[294]

A more relational leadership style promotes inclusion, leads to treating others with respect, sharing information transparently, providing constructive and truthful feedback, listening to others and valuing their opinions[295]. Gentle leadership therefore puts people at the center, people who are critical to the success and growth of companies. While Enel in Italy has pioneered kind leadership in business, some examples of adoptions of this model can also be seen in the United Kingdom. Indeed, the retail company John Lewis has made fairness and kindness to workers a cornerstone, and the Nationwide Building Society promotes kindness by also publishing advertisements about it[296]. In Enel, gentle leadership is represented with a triangle, which includes sustainable results, well-being (people care) and motivation at the vertices. The model adopted also is considered people empowerment and is based on soft skills, which will then be combined with the hard skills already present in the company[297]. It is not enough, however, to create a values manifesto that includes the usual key words and remains abstract and unmeasurable; in fact, a more effective approach has been found at Enel, which consists of identifying precise behaviors that leaders commit to doing or not doing. From this, the Trust Behavior Index was created, which assigns a score to each manager that is based on questions asked anonymously of colleagues[298]. The index summarizes the three basic components of trust:

[292] L. Salonia, cit.
[293] L. Salonia, cit.
[294] G. Haskins, The value of kindness in corporate leadership, Chartered Governance Institute UK & Ireland, 2018 https://www.cgi.org.uk/knowledge/governance-and-compliance/features/kindness-corporate-leadership
[295] G. Haskins, cit.
[296] G. Haskins, cit.
[297] G. Stratta-P. Cervini, cit.
[298] G. Stratta-P. Cervini, cit.

credibility, i.e., competence and reliability, intellectual honesty, and commitment to goal achievement; vulnerability, i.e., active listening, sharing, asking for feedback, and admitting mistakes; and finally, orientation to "us" instead of "me"[299]. After scoring, open dialogue and retrospective sessions are conducted in a constructive pathway logic in favor of people's growth and enhancement (in fact, there is no evaluative intent)[300].

On the subject of kindness in the workplace, a study conducted by Infojobs (a job search platform), aimed at understanding what kindness means in the workplace and how it has changed during the pandemic, shows that in the world of work, 64.3 percent say there is always room for kindness and 25.4 percent specify that it depends on the context and role[301]. In contrast, 10.2 percent of the nearly 2,000 respondents believe that the work environment does not leave room for kind behavior, considering it too competitive[302]. Although, however, kindness is appreciated across the board, obstacles mentioned by respondents include hectic pace and stress for 43%, competitiveness for 27%, and routine (2%)[303].

InfoJobs research also finds that the concept of gentle leadership is well integrated within the vocabulary of Italians, and in this regard the research has compiled a ranking of the characteristics that a gentle leader should possess. In first place we find team spirit, thus knowing how to put the focus on "we" and not on "I," in second place is knowing how to lead one's team toward goals without imposing methods and ideas, in third place respondents indicate rewarding results and not blaming but investigating failures, and finally in fourth place is knowing how to listen and gratify[304]. Often the key traits of kind leaders include characteristics such as: empathy, the basis of emotional intelligence; altruism, with a focus on people and society; respect and humility; and finally fairness, so being able to value people regardless of hierarchy.[305]

[299] G. Stratta-P. Cervini, cit.
[300] G. Stratta-P. Cervini, cit.
[301] L. Salonia, Gentilezza ai tempi di Covid: i piccoli gesti che ci fanno bene - Foto iO Donna, iO Donna, 2020 https://www.iodonna.it/benessere/salute-e-psicologia/gallery/gentilezza-ai-tempi-di-covid/
[302] L. Salonia, cit.
[303] Ministero del Lavoro e delle Politiche Sociali, L'importanza della gentilezza nel curriculum vitae, 2020 https://www.cliclavoro.gov.it/page/limportanza_della_gentilezza_nel_curriculum_vitae?contentId=BLG10599
[304] InfoJobs, Un leader forte è un leader gentile, 2020 https://lavoroedintorni.infojobs.it/2020/11/13/un-leader-forte-e-un-leader-gentile
[305] G. Haskins, cit.

These, then, represent the key characteristics of a kind leader according to the respondents, but not everyone claims to have a kind leader leading their team. If 41% say they have a kind leader 41.5% say their boss does not consider kindness to be an important element, and even 17.5% have a boss who rewards a strict and rigid climate, thus having no room for kindness[306]. Despite this, 93 percent of respondents say that gentle leadership can make a difference, contributing to a peaceful climate and thus leading to the achievement of better and more innovative results[307].

Among the expressions of kindness in the work environment, support during difficulties or redistribution of workload, and sharing of successes and failures were recorded the most[308].

It has also been recorded that many managers struggle, however, to balance kindness and honesty, and indeed for some these two concepts may seem to conflict with each other, but this is not so [309]. While kindness contributes to a positive work climate, honesty is critical to ensuring transparency and clarity; when issues arise regarding, for example, difficult decisions or employee performance, it can be difficult to reconcile these two values. Often to balance the two, compromises are made and only part of the truth is told. These compromises were analyzed by Annabelle Roberts of the University of Chicago and Taya Cohen of Carnegie Mellon University, and it was concluded that all such strategies that try to balance honesty and kindness fail on both sides.[310] Therefore, it can be concluded that it is wrong to think we can be nicer by avoiding difficult conversations, and we should accept that the two values are not at odds with each other. In fact, researchers say that we need to find strategies to maximize both and combine them, but without compromising[311].

[306] InfoJobs, cit.
[307] InfoJobs, cit.
[308] InfoJobs, cit.
[309] B. Beasley, Be honest or kind? Do you really have to choose? // News // Notre Dame Deloitte Center for Ethical Leadership // University of Notre Dame https://ethicalleadership.nd.edu/news/be-honest-or-be-kind-do-you-really-have-to-choose/
[310] B. Beasley, cit.
[311] B. Beasley, cit.

Chapter 4. The relationship between kindness and Artificial Intelligence

Nowadays, technologies such as artificial intelligence have pervaded many aspects of our daily lives, and the consequences of this digital revolution also permeate aspects such as people's social and ethical spheres, as well as increasing productivity and efficiency. In this context, it is therefore important to analyze how human principles, such as kindness, can influence the design and implementation of artificial intelligence.

This concluding chapter focuses on this intersection of kindness and artificial intelligence, which can give rise to a positive impact on people, their lives, and society, thus helping the creation of a more equitable and sustainable future. The goal of the chapter is to provide insight into how AI and values such as kindness can coexist, including through conscious design, thus making it possible to create AI systems that can solve complex tasks while operating with kindness and with respect for human values. Such an approach is necessary to ensure that AI is guided by values that make humanity worth celebrating. Indeed, the technological systems we adopt should not only be the result and reflection of our technical and engineering capabilities, but also of our values and how we want to shape the future.

4.1. AI and ethical decision making

Artificial intelligence is revolutionizing many aspects of our daily lives as well as most areas of work, from education to medicine, from finance to the arts. With the expansion of this technology, however, ethical issues of no small importance are emerging, issues that deal with aspects such as security, privacy, accountability, transparency, and fairness, as explained in the second chapter.

The potential of artificial intelligence is truly enormous, but its ethical aspects need to be addressed, so it is crucial that all stakeholders, from developers to policy makers, address them to ensure that it is used ethically and does not have deleterious effects on people and society.

The revolution that artificial intelligence brings comes through the decision-making processes in which it is implemented, a fact that presents many opportunities but also many challenges.

Regarding ethical risk factors in artificial intelligence-based decision-making mechanisms, it was found that the main sources of risk lie in incomplete data, management errors, and technological uncertainty[312]. To ensure that these risks do not become obvious and are spread, the solution lies in the elements of risk governance [313]. If we go to look at the totality of ethical risk factors related to artificial intelligence decision making, they can be structured in the model shown below, which also includes the individual risks mentioned above. The risk factors are divided into two categories: technology risk identification (algorithm, data and technology risk) and management risk identification[314].

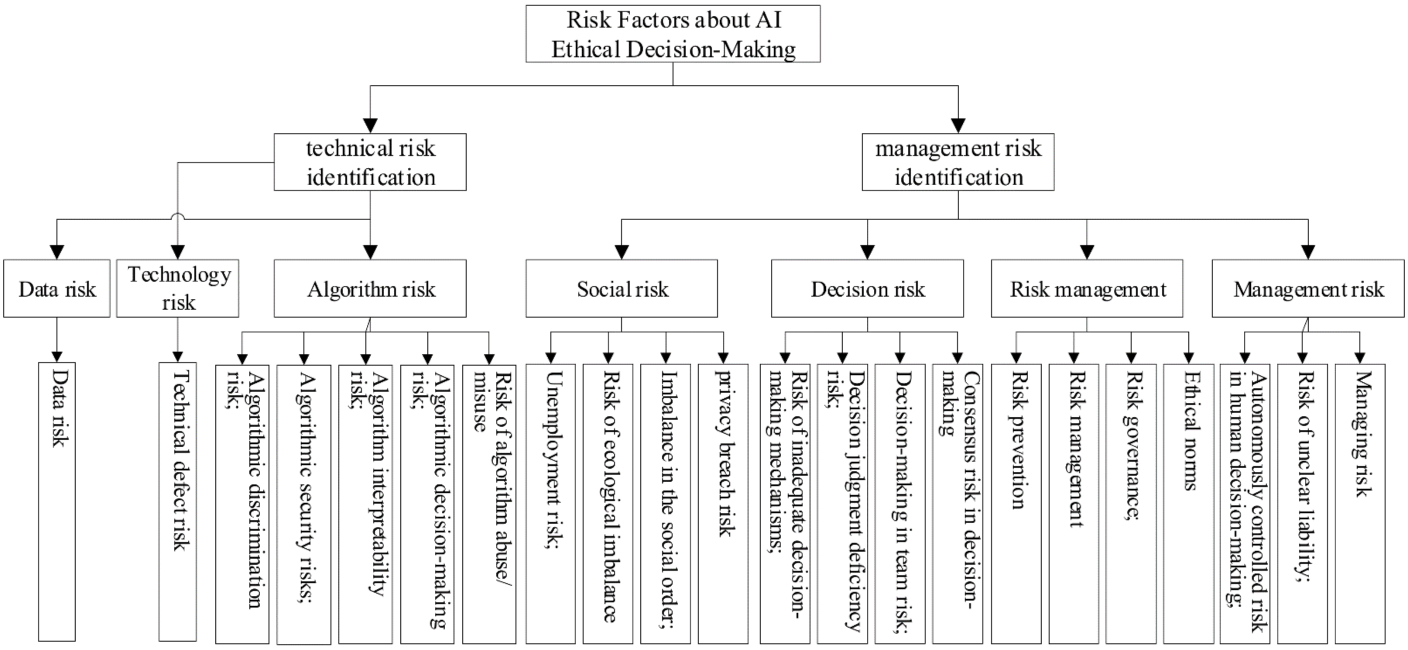
Risk Factors about AI Ethical Decision-Making
technical risk identification
management risk identification
Data risk
Technology risk
Algorithm risk
Social risk
Decision risk
Risk management
Management risk
Data risk
Technical defect risk
Algorithmic discrimination risk;
Algorithmic security risks;
Algorithm interpretability risk;
Algorithmic decision-making risk;
Risk of algorithm abuse/ misuse
Unemployment risk;
Risk of ecological imbalance
Imbalance in the social order;
privacy breach risk
Risk of inadequate decision-making mechanisms;
Decision judgment deficiency risk;
Decision-making in team risk;
Consensus risk in decision-making
Risk prevention
Risk management
Risk governance;
Ethical norms
Autonomously controlled risk in human decision-making;
Risk of unclear liability;
Managing risk

FIGURE 4.1 – Structural model of the dimensions of ethical risk factors regarding AI decision making. SOURCE: Guan, H.; Dong, L.; Zhao, A. Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making.

[312] H. Guan-L. Dong-A. Zhao, Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making, in Behavioral Sciences, fasc. 12, 9, 2022, p. 343

[313] H. Guan-L. Dong-A. Zhao, cit.

[314] H. Guan-L. Dong-A. Zhao, cit

It is a complex picture, however, because in addition to the multitude of identifiable ethical risk factors, there are complex relationships and influences among them [315].

On the ethical side with regard to artificial intelligence, there is still much to be done, starting even with just creating unified legislation, but there are those who argue that the ethical bar with regard to technology, is higher than the one set for us humans ourselves, thus reaching a paradox and questioning the relationship between AI and ethics[316]. Regarding, for example, the decision-making process behind professional hiring, concerns have arisen regarding the biases of AI applications used in this context. Thus, a total absence of bias is expected from these systems, biases that are, however, generally accepted in human recruiters, as they are precisely human[317]. Indeed, there is always the possibility of unconscious bias despite the fact that a person may strive to be impartial. The same phenomenon can also be observed in criminal justice and self-driving cars[318]. Thus, high and higher ethical standards are expected to be adhered to than those we follow in our daily lives, but this is necessary because artificial intelligence systems do not have morals or common sense; therefore, they need ethical guidelines to guide them to deal with decisions appropriately[319]. Another difference with us humans is the fact that it is much faster; this can lead to consequences of greater scale and thus possibly greater harm. Or, regarding bias, AI may amplify it unknowingly, not having the ability to question the data it is trained on[320]. In conclusion, high ethical standards are critical to ensure responsible and reliable use and implementation of this powerful technology. Thus, the criticality lies not in the fact that humans have lower standards, but in the fact that they have more limited capabilities and that everyone has their own limitations and flaws, a trait that makes us fallible but also human.

Through design it is possible to try to implement the ethical standards of the people who design it, striving to minimize discrimination.

In this context, the difficulty with using artificial intelligence in complex decision-making processes is the fact that this technology, unlike humans, does not possess moral values,

[315] H. Guan-L. Dong-A. Zhao, cit
[316] F. Moioli, Does AI have higher ethical standards than humans? | LinkedIn, 2023 https://www.linkedin.com/pulse/does-ai-have-higher-ethical-standard-than-humans-fabio-moioli/
[317] F. Moioli, cit.
[318] F. Moioli, cit.
[319] F. Moioli, cit.
[320] F. Moioli, cit.

emotions, or tacit knowledge such as customs and traditions, elements that are difficult to fully structure in these systems[321]. In addition to this, if one wants to include ethical frameworks within artificial intelligence systems, one must make a choice about which principles to include; alternatively, one would have to describe real situations and the resulting moral rule, but such a description would risk never being complete. This risk can be mitigated if a choice can be made about the decisions made by artificial intelligence systems by looking at them with a critical eye. This oversight is called the precautionary principle[322].

There are also recommendations regarding the ethical risks of AI decision-making processes. To be successful in governing these risks, it is necessary to develop management, research, and utilization standards, promoting agile governance, and putting in place preventive measures[323].

It has also been found that a promising method for decreasing the gap between ethical principles and practice in AI ethics is ethics-based auditing, such as through third-party auditors [324].

It remains crucial to ensure quality standards for training data, while at the same time ensuring their reliability and security, so that we have secure, transparent, and nondiscriminatory algorithms.

4.2. The importance of kindness in designing AI systems

Thus, it emerges as a solution to implement The ethical principles already during the design phase in artificial intelligence systems, so as to avoid the negative consequences described above and through the examples in the second chapter[325].

[321] H. Guan-L. Dong-A. Zhao, cit.

[322] D. Marino, Intelligenza artificiale, usi o decisioni non etiche: il dibattito, Agenda Digitale, 2021 https://www.agendadigitale.eu/cultura-digitale/se-lintelligenza-artificiale-non-e-etica-le-ricadute-sui-nostri-diritti/

[323] H. Guan-L. Dong-A. Zhao, cit.

[324] J. Mökander-L. Floridi, Ethics-Based Auditing to Develop Trustworthy AI, in Minds and Machines, fasc. 31, 2, 2021, p. 323–327

[325] H. Yu-Z. Shen-C. Miao-C. Leung-V. R. Lesser-Q. Yang, Building Ethics into Artificial Intelligence, arXiv, 2018

Implementation during the design phase would be crucial because once an artificial intelligence system is released and then consequently adopted on a large scale, it can become very difficult to try to change its direction and correct any ethical problems. Activating as early as during the design phase proves to be a more effective preventive measure rather than retroactively fixing the harms. Moreover, a priori implementation of ethical principles could also lead to greater trust in artificial intelligence, making people more likely to accept decisions made through AI and use its systems more comfortably. This would entail ethical reasoning capabilities implemented with algorithmic integration, through tools that describe a situation and ethical principles that would empower people to independently formulate a decision, justifying why a certain decision is ethically acceptable or not[326].

Wanting to incorporate a value such as kindness already in the design and development of artificial intelligence systems, one must keep in mind some challenges that would hinder its implementation. The main one is the fact that kindness, being a human value involving empathy and wanting to do good, is much more difficult to implement. It is possible to develop AI systems trained to behave in ways deemed kind and respectful (e.g., through polite responses), but AI is not able to feel emotions or understand them as humans do.

An example of training that demonstrates artificial kind and more generally prosocial behaviors (i.e., those behaviors enacted in such a way that positive effects on other people ensue) is that of Google Mini, a voice-activated wireless speaker created by Google[327]. When approached by a person in a conversational context deemed oppositional, Google Mini adopted a tone expressing artificial pro-sociality, being programmed to be pro-social. Using audio signal recognition software for automatic emotion recognition, it also showed that emphasis was placed on words where the human interlocutor's attention should fall, e.g. in the phrase "sorry, I sometimes make mistakes," the word "sorry" was highlighted[328]. Emphasis was also placed on all words considered "attenuators," that is, those terms used to lower the conversational tone. These facets emerging from the behavior of the Google Mini device allow it to lead to kinder and more polite communication, certainly not

[326] D. Marino, cit.
[327] C. Papapicco, Google Mini: Italian Example of Artificial Prosociality, in Online Journal of Communication and Media Technologies, fasc. 10, 3, 2020, p. e202015
[328] C. Papapicco, cit.

allowing to characterize all the nuances of this human value, but still trying to imitate it[329]. This case could also serve as a springboard for studies regarding bullying, as it embodies how useful it would be to behave when confronted with a provocative conversation [330].

An algorithmic model, called VirtuosA, has also been described, in which artificial intelligence systems are able to learn ethical behavior based on an adapted philosophical framework inspired by Christian philosopher Dallas Willard, coupled with broad systems thinking[331]. In this case, theology but also neuroscience has been integrated into the algorithm, so as to provide it with the broadest possible understanding of human action and especially of our moral values, thereby leading to an attempt to anatomize the moral sense[332]. The application and operation of the model described are divided into three phases: mentoring, exploratory, and virtue development[333]. Specifically, in the case described above, the robot learns to be kind through following the example of a mentor, who exhibits kind behaviors toward a third robot, thus initiating a virtuous habit[334]. Previously, in fact, some of the robot's harmful habits stemmed from the process of maximizing personal achievement, a fact, however, that could go against the interest of the other robot [335].

Through the study that led to the design of the VirtuosA algorithm, the conclusion can thus be drawn that it is possible to somehow implement in artificial intelligence systems patterns of moral principles proper to human beings; thereby guiding the algorithms to learning that we might call ethical, using state-of-the-art computing practices. This therefore becomes a solution to the negative ethical impacts that artificial intelligence can bring.

[329] C. Papapicco, cit.
[330] C. Papapicco, cit.
[331] N. Crook-J. Corneli, The Anatomy of moral agency: A theological and neuroscience inspired model of virtue ethics, in Cognitive Computation and Systems, fasc. 3, 2, 2021, p. 109–122
[332] N. Crook-J. Corneli, cit.
[333] N. Crook-J. Corneli, cit.
[334] N. Crook-J. Corneli, cit.
[335] N. Crook-J. Corneli, cit.

4.3. The impact of kindness on AI

When implemented in artificial intelligence systems, as in the case of Google Mini, there is thus the possibility that kindness will be disseminated, especially in a digital context, in a way that promotes behavior that is both polite and in some ways empathetic, helping to build an increasingly positive environment and also improving the user experience.

The development and application of a model like VirtuosA could have several benefits on human well-being and organizations. For example, the spread of kindness by an artificial intelligence could have the same effects that kind leadership, mentioned in Chapter Three, has on organizations and employees. Indeed, kind leadership leads to an increase in employee satisfaction, creativity, productivity, and sometimes even loyalty.

If similar diffusion were to occur worldwide through the creation of kind AI systems, the impacts would be far more widespread and not limited only to the enterprise. Such a scenario would first and foremost lead to an improvement in existing interactions between humans and machines, thereby increasing the use of and trust in this technology. If implemented, kindness could also bring emotional support for those in need, especially in key areas such as healthcare, where both patients and staff might find emotional support useful.

Gentle artificial intelligence could also help promote the ethical use of technology and digital spaces, preventing phenomena such as cyberbullying, abuse, and discrimination, fostering the development of a more equitable society.

A study was also recently conducted regarding the responses given by an artificial intelligence chatbot on a social media forum in the medical field. Comparing the chatbot's responses with those given by physicians, the former were preferred by patients, not only for quality but also for empathy[336].

On the same type of solution, back in 2019 Humana Pharmacy developed an artificial intelligence algorithm that instructed the company's employees to be kinder and show

[336] J. W. Ayers-A. Poliak-M. Dredze-E. C. Leas-Z. Zhu-J. B. Kelley-… D. M. Smith, Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum, in JAMA Internal Medicine, 2023

more compassion toward customers[337]. Artificial intelligence provides feedback to operators and looks for patterns that may indicate communication that is apparently rude or lacks empathy [338]. Robots, therefore, can certainly help us to be kinder.

A similar example is the communication robot developed by Hitachi, a Japanese multinational corporation, which can express kindness thus attracting people[339]. EMIEW4, this is the name of the robot, is able to welcome and guide people in offices and facilities in general, providing a communication experience that aims to be as natural as possible.

Another conversational robot based on artificial intelligence systems is Moxie, created by the startup Embodied[340]. Moxie is able to teach kindness to children and promotes aspects of learning that have to do with the social and emotional sphere. The robot expresses itself in emotionally charged language to which it combines credible facial expressions, succeeding in developing social skills, among them empathy and emotion management, for those who use it[341]. This development takes place through conversations but also through specific missions to be performed by the user. Thanks to the technology behind Moxie, namely machine learning, the robot is also able to personalize the user experience by adapting to the personality of the person interfacing with it, creating a sense of trust[342]. While this robot is currently being used to develop children's skills, the company's top management believes that in the future this technology may be able to be used to improve everyday social interactions[343].

One platform that leverages artificial intelligence to improve everyday interactions and spread kindness is KindWorks.AI. This platform was created to encourage acts of kindness for billions of people in the workplace, leveraging tools such as Slack, Whatsapp, and

[337] N. Martin, Robots Are Actually Teaching Humans To Be More Compassionate, Forbes, 2019 https://www.forbes.com/sites/nicolemartin1/2019/08/27/robots-are-actually-teaching-humans-to-be-more-compassionate/

[338] N. Martin, cit.

[339] Hitachi, Robot di comunicazione EMIEW4: Esprimere gentilezza che attrae le persone : Ricerca e sviluppo : Hitachi https://www.hitachi.com/rd/research/design/product/emiew4/index.html

[340] R. Stevenson, Moxie the Conversational AI Robot Teaches Children Kindness, SAMA, 2021 https://www.sama.com/podcast/moxie-the-robot-teaches-children-kindness-conversational-ai-child-development/

[341] F. Ioli, Moxie, il nuovo robot coach per lo sviluppo socio-emotivo dei bambini, IngegneriaBiomedica.org, 2020 https://www.ingegneriabiomedica.org/news/robotica-intelligenza-artificiale/moxie-il-nuovo-robot-coach-per-lo-sviluppo-socio-emotivo-dei-bambini/

[342] F. Ioli, cit.

[343] F. Ioli, cit.

Teams in a way that facilitates adoption[344]. Content is tailored to the user, and conversations are based on behavioral science in order to have a positive impact. The pillar of the platform is the conversational agent Beni, who encourages acts of kindness, through behavioral nudges (i.e., getting people to perform actions by creating the right conditions to influence people but without imposing anything, also referred to as kind nudges), aimed at themselves, others, and society in general[345]. The benefits found through KindWorks.AI and its encouragement of kindness are many: it increases employee well-being, collaboration, job satisfaction, sense of purpose, and generates benefits to both the user and the recipient of acts of kindness, and ultimately to society[346]. It can be seen that all the characteristics and benefits that KindWorks.AI succeeds in bringing to a company coincide with the traits of good leadership such as the gentle one mentioned above, namely trust, openness, generosity, friendly relationships, and the creation of a fulfilling culture that also leads to greater productivity[347].

It is also interesting to go to the other side of the coin, namely, how humans interface with artificial intelligence systems. In fact, it has been shown that humans tend to show kindness when approaching artificial intelligence assistants, although there are no consequences in proving grumpy on the contrary[348]. Dennis Mortensen, the CEO of x.ai, the company that created two virtual assistants, Amy and Andrew, said that as much as 11 percent of communication with the two virtual assistants is to show gratitude toward their work, despite the fact that they are artificial intelligence systems without feelings[349]. Therefore, it comes logical to think that the people who interface with these assistants treat them as if they were human beings, which they clearly are not and do not pretend to be, in fact clearly showing that they are artificial intelligences through their names incorporating ".ai," not wanting to fool anyone[350]. Despite this, we humans therefore tend to show compassion toward a tool that helps us in our daily tasks. On the other hand, Amy

[344] KindWorks.AI, KindWorks.AI - igniting purpose, impact, and connection, 2021 https://kindworks.ai/
[345] KindWorks.AI, cit.
[346] KindWorks.AI, cit.
[347] joyce shen, Kindness & AI, Medium, 2022 https://medium.com/@joycejshen/kindness-ai-dd4272f88299
[348] D. R. Polgar, Thanks, Robot! Humans are Showing Kindness with Their AI Helpers., Big Think, 2017 https://bigthink.com/sex-relationships/thanks-robot-how-should-we-be-treating-our-ai-helpers/
[349] D. R. Polgar, cit.
[350] D. R. Polgar, cit.

and Andrew also incorporate elements into the conversations that can increase the empathy of the dialogue[351].

There is also a scientist who has intersected the two concepts of algorithms and compassion-Dr. Amit Ray. Indeed, Dr. Ray has succeeded in demonstrating how machine learning algorithms can be used to address needs related to the human sphere[352]. Examples include compassionate care, coping with terrorism, and helping those afflicted with physical and mental problems.

While to date this technology has been used to optimize our time and tasks, the implementation of values such as kindness and compassion would ensure human well-being while also increasing people's happiness. If implemented and harnessed in the right way, it could also enhance human kindness[353]. This last aspect could prove very useful especially in third sector companies and companies working in the social field, Ray in fact shows how compassionate artificial intelligence could be a tool for solving humanitarian crises following natural disasters, but also for solving health problems by relieving suffering[354]. The scientist also believes and states that the best solution to the problems related to this field of technology would be the incorporation of "Compassion by design" and "Safety by design," thus making sure that rules are put in place from the development of these systems and preventing artificial intelligence from becoming a threat to humanity, as feared by many scholars[355].

Although kindness is not the first value that comes to mind when discussing the ethics of artificial intelligence, it could be a key element regarding an ethical and responsible approach to the development of artificial intelligence systems[356]. By promoting kindness, equity could also be increased accordingly, leading to more benefits for all. In addition, kindness encourages ethical behavior, taking into consideration the consequences of

[351] D. R. Polgar, cit.

[352] N. Martin, cit.

[353] J. Hinton, AI is no match for human kindness, but it can certainly bring it to life, 2020 https://razor.co.uk/insights/ai-is-no-match-for-human-kindness-but-it-can-certainly-bring-it-to-life-

[354] J. Coleman, Next AI Revolution — Compassionate Artificial Intelligence, Medium, 2018 https://askwhy.medium.com/next-ai-revolution-compassionate-artificial-intelligence-557d5f49dc68

[355] J. Coleman, cit.

[356] Future World Alliance, Why Kindness is a Crucial Part of Responsible AI, Future World Alliance, 2023 https://futureworldalliance.org/blog/f/why-kindness-is-a-crucial-part-of-responsible-ai

actions implemented, both at the level of users and people in general, and at the level of artificial intelligence developers[357].

4.4. Challenges of AI regarding kindness

Kindness, along with other qualities such as empathy and additional core values, is a hallmark of human beings, an essential component of our humanity but also of our character. These elements not only define us at the level of identity as individuals, but also as a society. Kindness can manifest itself through acts of courtesy, gestures of generosity or altruistic behavior. It represents not only an action in its own right but an attitude that succeeds in expressing consideration and awareness of others. Through this value we are also able to create bonds of affection and social ties that implement mutual respect, a foundational trait for the development of cohesive sociality.

Both kindness and empathy are in fact not only so-called ethical or simply desirable behaviors, but are also purely human phenomena: both represent us and make us authentically human.

Thus, kindness represents a complex and unique characteristic that derives from human beings and is developed through experience, empathy and even emotional awareness. On the other hand, artificial intelligence represents the fruit of human ingenuity and was developed with the intent to solve complex problems and tasks efficiently.

However, although it is then possible to program artificial intelligence systems that can simulate kind behavior, such kindness remains superficial and simulated. In this case there always remains a missing piece namely, the authenticity and depth that characterize real human interactions. It is certainly possible to program AI systems that can provide kind responses or assist us with kindness in our daily tasks, but these actions are still driven by predefined algorithms and not by sincere emotions.

Artificial intelligence, in spite of this, can still play a significant role when it comes to creating a more efficient environment for us humans by automating repetitive tasks and supporting us in daily activities.

[357] Future World Alliance, cit.

Kindness in artificial intelligence is thus something simulated and the result of programming, which has little to do with the authenticity of feelings such as compassion and empathy. It can definitely bring it to life, but without competing with human kindness and connection[358].

Although artificial intelligence is able to pick up signals with respect to human emotions, it does not have the same effectiveness as human understanding, being based only on machine learning data it has been taught, such as in the case of the Humana Pharmacy algorithm mentioned earlier. In fact, the CEO of the company that developed the algorithm believes it can only act as a coach and not as a replacement for the operators[359].

In addition, a further problem arises from simulating kindness and other values such as, for example, empathy. This may indeed present the risk of deception and manipulation. It would therefore be important to associate kindness, an ethical framework within artificial intelligence systems so as to avert such risks.

Artificial intelligence, however, remains a technology developed and designed by humans and should therefore work for the benefit of humanity, while at the same time keeping alive the valuable traits that make us human and, above all, respecting us. This implies that we should consider the role of AI as a support for our activities, rather than a substitute for the human experience. To this end, it is necessary to promote the ethical use of AI, so that this powerful tool is used to improve our lives and succeeds in helping us in addressing challenges such as sustainability and many others. In conclusion, artificial intelligence can prove vital in our lives and bring enormous benefits to them, but it is essential that it be developed and implemented with a humanistic perspective in mind.

[358] J. Hinton, cit.
[359] N. Martin, cit.

Conclusions

In conclusion, this thesis has explored the history, definitions but also the ethical challenges and importance of kindness within Artificial Intelligence. AI, since the beginning of its history, has undergone several transformations, which have declined into periods of enthusiasm but also disillusionment, finally arriving at its current state of development. This technology is certainly not new, but it is clear that in 2023 the attention around this field of technology has increased significantly, especially with regard to certain gills such as generative AI, which has become the star of the last few months especially thanks to OpenAI and its ChatGPT.

If we refer to Gartner's Hype Cycle, we find different branches of this technology at different points. Gartner's Hype Cycle consists of a graphical representation of the level of media attention, maturity and adoption of specific technologies and is divided into five phases: innovation and technology trigger (entry into the marketplace often accompanied by high expectations); peak of inflated expectations (peak of enthusiasm); abyss of disappointment (enthusiasm begins to fade and expectations are not met); slope of enlightenment (we begin to understand how to leverage this technology to achieve tangible results in companies); plain of productivity (the technology is well established and its benefits are known and widely understood)[360]. Applications such as computer vision are already in the slope of enlightenment, others such as autonomous vehicles, deep learning, and Natural Language Processing are still part of the abyss of disappointment, and finally some such as Generative AI and Responsible AI are to date between the innovation trigger and the peak of inflated expectations[361]. This shows that we are actually still a long way from seeing the full potential of artificial intelligence being applied, and we also need to keep in mind the ethical challenges that need to be overcome at this juncture[362].

In 2023, there are also important trends pointed out by the AI Index report to consider. Industry is outpacing academia when it comes to producing machine learning models due

[360] K. Terrell Hanna-I. Wigmore, What is the Gartner Hype Cycle? – TechTarget Definition, WhatIs.com, 2023 https://www.techtarget.com/whatis/definition/Gartner-hype-cycle

[361] J. Wiles, What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle, Gartner, 2022 https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

[362] K. Wilson, Beware the AI hype cycle – managing expectations in the age of ChatGPT, UKTN | UK Tech News, 2023 https://www.uktech.news/guest-posts/ai-hype-cycle-expectations-20230519

to the increased availability of data and funding[363]. However, this should not reduce the importance of institutions in this context. Artificial intelligence, moreover, is contributing to many scientific advances in fields such as medicine and energy, on the other hand, however, the recognition of potential environmental damage by AI systems is also growing, an aspect that definitely needs to be explored further[364]. There is also growing interest in AI among policymakers, and there is a concomitant increase in the number of laws containing the term "artificial intelligence," which is certainly a positive sign that may indicate a greater commitment on the part of governments to regulating AI in both ethical and responsible ways[365].

The AI index report also reminds us how important it is to address the ethical challenges that arise from artificial intelligence systems, in fact the trend regarding ethical controversies and incidents involving this technology is growing[366]. Indeed, it is essential to ensure that new technologies, especially one with high potential such as artificial intelligence, are both developed and implemented and used in an ethical and responsible manner, taking human values into account. In this context, it arises as imperative to consider factors such as privacy, security, equity, and impact on human relationships.

In addition to these principles that are already much discussed in order to arrive at responsible AI, it would also be important to incorporate a core human value such as kindness into AI, as it could play a crucial role in this context and succeed in promoting a more empathetic and humane approach to artificial intelligence. Indeed, the technologies we develop and use should also be a reflection of our human values and how we want to shape the future. Having brought the virtuous example of the new model of kind leadership and its positive impact within companies, similar effects but a much broader impact could be had by including kindness in AI systems. Such implementation, ideally as early as the design stage, would also succeed in having a positive effect worldwide if implemented on a large scale or in widely deployed AI products. Through this, ethical use of digital spaces could also be promoted, preventing cyberbullying and discrimination, thus helping to create a more equitable environment and society.

[363] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, cit.
[364] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, cit.
[365] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, cit.
[366] N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-… R. Perrault, cit.

In conclusion, the goal of this thesis has been to provide a comprehensive overview regarding artificial intelligence, highlighting both the ethical challenges involved and the potential value of kindness within AI systems. Despite the advances in AI, it is essential to ensure an open and collaborative dialogue regarding ethical issues and values to be incorporated into this technology. Policymakers, companies, researchers but also the public should work in synergy to succeed in ensuring that AI is used in a way that maximizes its benefits while minimizing its risks to humanity. At a time when we are moving toward a future that is increasingly permeated by artificial intelligence and in which it will play an increasingly fundamental role in our society, it becomes an imperative to continue to reflect on the proposed issues and to act responsibly, always respecting human values. Indeed, this is the only way to ensure that AI systems not only succeed in improving our lives, but also respect our fundamental values.

Bibliography

Accademia della Gentilezza, Parole e Comportamenti gentili, Accademia della Gentilezza, 2022 https://www.accademiadellagentilezza.it/words/

I. Ananthabhotla-A. Rieger-D. Greenberg-R. Picard, MIT Community Challenge: Designing a Platform to Promote Kindness and Prosocial Behavior, in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2017, p. 2352–2358

J. Angwin-J. Larson-S. Mattu-L. Kirchner, Machine Bias, ProPublica, 2016 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

J. W. Ayers-A. Poliak-M. Dredze-E. C. Leas-Z. Zhu-J. B. Kelley-… D. M. Smith, Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum, in JAMA Internal Medicine, 2023

J. Barroca, Surveillance and Predictive Policing Through AI | Deloitte https://www.deloitte.com/global/en/Industries/government-public/perspectives/urban-future-with-a-purpose/surveillance-and-predictive-policing-through-ai.html

B. Beasley, Be honest or kind? Do you really have to choose? // News // Notre Dame Deloitte Center for Ethical Leadership // University of Notre Dame https://ethicalleadership.nd.edu/news/be-honest-or-be-kind-do-you-really-have-to-choose/

L. Bertuzzi, AI Act moves ahead in EU Parliament with key committee vote, www.euractiv.com, 2023 https://www.euractiv.com/section/artificial-intelligence/news/ai-act-moves-ahead-in-eu-parliament-with-key-committee-vote/

M. Borzacchelli, Intelligenza artificiale e GDPR sono compatibili, ma servono norme più chiare: lo studio, Agenda Digitale, 2020 https://www.agendadigitale.eu/sicurezza/intelligenza-artificiale-e-gdpr-sono-compatibili-ma-servono-norme-piu-chiare-lo-studio/

Boston Consulting Group, Generative AI, BCG Global, s.d. https://www.bcg.com/x/artificial-intelligence/generative-ai

T. Bradley-K. C. Anderson-A. Hass, The Virtuous Cycle: Social Media Influencers' Potential for Kindness Contagion, in Journal of Macromarketing, fasc. 43, 2, 2023, p. 110–118

S. Bradshaw-P. N. Howard, DemTech | Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation, 2017 https://demtech.oii.ox.ac.uk/research/posts/troops-trolls-and-troublemakers-a-global-inventory-of-organized-social-media-manipulation/

G. Brunetti, L'etica della gentilezza e la sua funzione terapeutica, Riflessioni.it, 2021 https://www.riflessioni.it/finestre-anima/etica-della-gentilezza-e-sua-funzione-terapeutica.htm

J. Buolamwini-T. Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, 2018, p. 77–91

C. Caldwell, Understanding Kindness – A Moral Duty of Human Resource Leaders, in The Journal of Values-Based Leadership, fasc. 10, 2, 2017

R. Carrozzo, Diritto dell'AI: il ruolo dell'etica, AI4Business, 2022 https://www.ai4business.it/intelligenza-artificiale/diritto-dellai-il-ruolo-delletica/

M. Chui-V. Kamalnath-B. McCarthy, An executive's guide to AI - McKinsey, 2020 http://ceros.mckinsey.com/quick-guide-to-ai-12

J. Coleman, Next AI Revolution — Compassionate Artificial Intelligence, Medium, 2018 https://askwhy.medium.com/next-ai-revolution-compassionate-artificial-intelligence-557d5f49dc68

Consiglio dell'UE, Normativa sull'intelligenza artificiale: il Consiglio chiede di promuovere un'IA sicura che rispetti i diritti fondamentali, 2022 https://www.consilium.europa.eu/it/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/

Council of Europe, History of Artificial Intelligence - Artificial Intelligence - www.coe.int, Artificial Intelligence, s.d. https://www.coe.int/en/web/artificial-intelligence/history-of-ai

N. Crook-J. Corneli, The Anatomy of moral agency: A theological and neuroscience inspired model of virtue ethics, in Cognitive Computation and Systems, fasc. 3, 2, 2021, p. 109–122

D. Davies, Facial Recognition And Beyond: Journalist Ventures Inside China's «Surveillance State», in NPR, 2021

J. Delua, Supervised vs. Unsupervised Learning: What's the Difference?, 2022 https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning

D. D. Delzell- collaboratore di C. Post, The difference between natural kindness and Christian kindness, The Christian Post, 2022 https://www.christianpost.com/voices/the-difference-between-natural-kindness-and-christian-kindness.html

C. and T. (European C. Directorate-General for Communications Networks-Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Orientamenti etici per un'IA affidabile, Publications Office of the European Union, LU, 2019

G. Dotto Pagnossin, L'impronta gentile. Attraversare la vita in punta di piedi. Vale anche per il relatore pubblico, 2022 https://www.ferpi.it/news/limpronta-gentile-attraversare-la-vita-in-punta-di-piedi-vale-anche-per-il-relatore-pubblico

L. Eliot, Democratization Of AI Is Said To Be Essential For AI Ethics But The Devil Is In The Details, Including The Case Of AI-Based Self-Driving Cars, Forbes, 2022 https://www.forbes.com/sites/lanceeliot/2022/03/24/democratization-of-ai-is-said-to-be-essential-for-ai-ethics-but-the-devil-is-in-the-details-including-the-case-of-ai-based-self-driving-cars/

M. Esposito, Natural language processing (NLP), che cos'è, i progressi e le sfide dell'AI, Agenda Digitale, 2019 https://www.agendadigitale.eu/cultura-digitale/linguaggio-naturale-e-intelligenza-artificiale-a-che-punto-siamo/

European Commission, Ethics guidelines for trustworthy AI | Shaping Europe's digital future, 2019 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission, A European approach to artificial intelligence | Shaping Europe's digital future, 2023 https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Commission, High-level expert group on artificial intelligence, 2023 https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

European Parliament. Directorate General for Parliamentary Research Services., Artificial intelligence: how does it work, why does it matter, and what we can do about it?, Publications Office, LU, 2020

European Parliament. Directorate General for Parliamentary Research Services., The ethics of artificial intelligence: issues and initiatives., Publications Office, LU, 2020

L. Floridi, The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities, 2022

M. Fox, Be kind online, 2021 https://www.bupa.com.au/healthlink/mental-health-wellbeing/mental-health/be-kind-online

J. Fruhlinger, What is generative AI? The evolution of artificial intelligence, InfoWorld, 2023 https://www.infoworld.com/article/3689973/what-is-generative-ai-the-evolution-of-artificial-intelligence.html

Future of Life Institute, AI Principles, 2017 https://futureoflife.org/open-letter/ai-principles/

Future of Life Institute, Pause Giant AI Experiments: An Open Letter, Future of Life Institute, 2023 https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Future World Alliance, Why Kindness is a Crucial Part of Responsible AI, Future World Alliance, 2023 https://futureworldalliance.org/blog/f/why-kindness-is-a-crucial-part-of-responsible-ai

F. Gaetani, Coded Bias: la responsabilità dell'algoritmo., 2023 https://www.lisia.it/post/coded-bias-la-responsabilita-dellalgoritmo

C. Gerino, Google raccoglieva dati via WiFi: multa da 13 milioni di dollari per Street View, la Repubblica, 2019 https://www.repubblica.it/tecnologia/sicurezza/2019/07/24/news/google_raccoglieva_dati_via_wi-fi_multa_da_13_milioni_di_dollari_per_street_view-231905698/

S. Gibbs, Musk, Wozniak and Hawking urge ban on warfare AI and autonomous weapons, in The Guardian, 2015

F. Giffoni, Leadership gentile: caratteristiche e peculiarità, 2019 https://qsfera.it/blog/leadership-gentile-caratteristiche-e-peculiarita

Google AI, Responsible AI practices, Google AI https://ai.google/responsibilities/responsible-ai-practices/

J. Gorman, A Brief History of AI, From French Philosophy to Self-Driving Cars, Dell, 2019 https://www.dell.com/en-us/perspectives/a-brief-history-of-ai-from-french-philosophy-to-self-driving-cars/

C. Goujard-G. Volpicelli, ChatGPT is entering a world of regulatory pain in Europe, POLITICO, 2023 https://www.politico.eu/article/chatgpt-world-regulatory-pain-eu-privacy-data-protection-gdpr/

H. Guan-L. Dong-A. Zhao, Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making, in Behavioral Sciences, fasc. 12, 9, 2022, p. 343

M. Haenlein-A. Kaplan, A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence, in California Management Review, fasc. 61, 4, 2019, p. 5–14

R. Haghemann-J.-M. Leclerc, Precision regulation for artificial intelligenze, IBM, 2020 https://www.ibm.com/policy/ai-precision-regulation/

L. Hardesty, Explained: Neural networks, MIT News | Massachusetts Institute of Technology, 2017 https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

G. Haskins, The value of kindness in corporate leadership, Chartered Governance Institute UK & Ireland, 2018 https://www.cgi.org.uk/knowledge/governance-and-compliance/features/kindness-corporate-leadership

A. Hern, TechScape: Clearview AI was fined £7.5m for brazenly harvesting your data – does it care?, in The Guardian, 2022

J. Hinton, AI is no match for human kindness, but it can certainly bring it to life, 2020 https://razor.co.uk/insights/ai-is-no-match-for-human-kindness-but-it-can-certainly-bring-it-to-life-

Hitachi, Robot di comunicazione EMIEW4: Esprimere gentilezza che attrae le persone : Ricerca e sviluppo : Hitachi, s.d. https://www.hitachi.com/rd/research/design/product/emiew4/index.html

L. Honeycutt, Book II - Chapter 7 : Aristotle's Rhetoric, 2004 https://web.archive.org/web/20041213221951/http://www.public.iastate.edu/~honeyl/Rhetoric/rhet2-7.html

R. Hursthouse-G. Pettigrove, Virtue Ethics, in E. N. Zalta, U. Nodelman (a cura di), The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, 2022Winter 2022

D. Huyskes, AI ed Etica: L'importanza di usare dati rappresentativi, 2021 https://www.sas.com/it_it/news/leading-art-innovation/innovation-sparks/ai-etica-importanza-dati-rappresentativi.html

A. Iasimone, Gentilezza: significato, potere e storia a cura del prof. Canettieri, Scuola di Comunicazione Gentile, 2021 https://www.comunicazionegentile.it/gentilezza-significato-potere-storia-paolo-canettieri/

IBM, What is Computer Vision? | IBM, s.d. https://www.ibm.com/topics/computer-vision

IBM, What is Natural Language Processing? | IBM https://www.ibm.com/topics/natural-language-processing

InfoJobs, Un leader forte è un leader gentile, 2020 https://lavoroedintorni.infojobs.it/2020/11/13/un-leader-forte-e-un-leader-gentile

F. Ioli, Moxie, il nuovo robot coach per lo sviluppo socio-emotivo dei bambini, IngegneriaBiomedica.org, 2020 https://www.ingegneriabiomedica.org/news/robotica-intelligenza-artificiale/moxie-il-nuovo-robot-coach-per-lo-sviluppo-socio-emotivo-dei-bambini/

N. Joshi, 7 Types Of Artificial Intelligence, Forbes https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/

M. Karlin-B. Ozawa-De Silva, The Science, Theory and Practice of Kindness: A Brief Overview, UNESCO MGIEP https://mgiep.unesco.org/article/the-science-theory-and-practice-of-kindness-a-brief-overview

E. Kavlakoglu, AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?, 2022 https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

E. Kazim-A. S. Koshiyama, A high-level overview of AI ethics, in Patterns, fasc. 2, 9, 2021

KindWorks.AI, KindWorks.AI - igniting purpose, impact, and connection, 2021 https://kindworks.ai/

W. Knight, How malevolent machine learning could derail AI, MIT Technology Review, 2019 https://www.technologyreview.com/2019/03/25/1216/emtech-digital-dawn-song-adversarial-machine-learning/

C. Lamoutte, Responsible AI Begins with Human-Centered Design - AI for Good Foundation, 2022 https://ai4good.org/blog/responsible-ai/

Le virtù etiche in Aristotele | Platone 2.0 – La rinascita della filosofia come palestra di vita, s.d. https://www.platon.it/storia/il-bene-e-la-felicita/perche-e-cosi-importante-la-virtu/le-virtu-etiche-in-aristotele/

D. Leslie, Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, Zenodo, 2019

D. Lo, Will you lose your job to AI and tech like ChatGPT?, Fast Company, 2023 https://www.fastcompany.com/90881876/ai-chatgpt-take-jobs

G. Lughi, Robotica sociale: gli impatti de «La società dei robot» tra spazi ibridi ed etica, Agenda Digitale, 2022 https://www.agendadigitale.eu/cultura-digitale/robotica-sociale-gli-impatti-de-la-societa-dei-robot-tra-spazi-ibridi-ed-etica/

D. Marino, Intelligenza artificiale, usi o decisioni non etiche: il dibattito, Agenda Digitale, 2021 https://www.agendadigitale.eu/cultura-digitale/se-lintelligenza-artificiale-non-e-etica-le-ricadute-sui-nostri-diritti/

B. Marr, Barbie Wants To Chat With Your Child -- But Is Big Data Listening In?, Forbes, 2015 https://www.forbes.com/sites/bernardmarr/2015/12/17/barbie-wants-to-chat-with-your-child-but-is-big-data-listening-in/

B. Marr, Understanding the 4 Types of Artificial intelligence, Bernard Marr, 2021 https://bernardmarr.com/understanding-the-4-types-of-artificial-intelligence/

N. Martin, Robots Are Actually Teaching Humans To Be More Compassionate, Forbes, 2019 https://www.forbes.com/sites/nicolemartin1/2019/08/27/robots-are-actually-teaching-humans-to-be-more-compassionate/

N. Maslej-L. Fattorini-E. Brynjolfsson-J. Etchemendy-K. Ligett-T. Lyons-... R. Perrault, AI Index Report 2023 – Artificial Intelligence Index, Institute for Human-Centered AI, Stanford University, 2023

F. Meta, Artificial Intelligence Act, accordo politico al Parlamento Ue sulle nuove norme, CorCom, 2023 https://www.corrierecomunicazioni.it/digital-economy/artificial-intelligence-act-accordo-politico-al-parlamento-ue-sulle-nuove-norme/

Microsoft, What Is Computer Vision? | Microsoft Azure, s.d. https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-computer-vision/

I. Mihajlovic, Everything You Ever Wanted To Know About Computer Vision. Here's A Look Why It's So Awesome., Medium, 2021 https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e

Ministero del Lavoro e delle Politiche Sociali, L'importanza della gentilezza nel curriculum vitae, 2020 https://www.cliclavoro.gov.it/page/limportanza_della_gentilezza_nel_curriculum_vitae?contentId=BLG10599

L. Mischitelli, La strategia europea sull'intelligenza artificiale: stato dell'arte e scenari futuri, Agenda Digitale, 2020 https://www.agendadigitale.eu/cultura-digitale/la-strategia-europea-sullintelligenza-artificiale-stato-dellarte-e-scenari-futuri/

MIT Media Lab, Person Overview ‹ Joy Buolamwini, MIT Media Lab, s.d. https://www.media.mit.edu/people/joyab/overview/

F. Moioli, Does AI have higher ethical standards than humans? | LinkedIn, 2023 https://www.linkedin.com/pulse/does-ai-have-higher-ethical-standard-than-humans-fabio-moioli/

J. Mökander-L. Floridi, Ethics-Based Auditing to Develop Trustworthy AI, in Minds and Machines, fasc. 31, 2, 2021, p. 323–327

C. Montgomery-F. Rossi-J. New, A Policymaker's Guide to Generative AI, IBM Newsroom, 2023 https://newsroom.ibm.com/Whitepaper-A-Policymakers-Guide-to-Foundation-Models

V. C. Müller, Ethics of Artificial Intelligence and Robotics, in E. N. Zalta (a cura di), The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, 2021Summer 2021

OECD, The OECD Artificial Intelligence (AI) Principles, 2019 https://oecd.ai/en/ai-principles

OECD, OECD Framework for the Classification of AI Systems: a tool for effective AI policies, s.d. https://oecd.ai/en/p/classification

C. Papapicco, Google Mini: Italian Example of Artificial Prosociality, in Online Journal of Communication and Media Technologies, fasc. 10, 3, 2020, p. e202015

S. Q. Park-T. Kahnt-A. Dogan-S. Strang-E. Fehr-P. N. Tobler, A neural link between generosity and happiness, in Nature Communications, fasc. 8, 1, 2017, p. 15964

D. Petersson, 4 Main Types of Artificial Intelligence: Explained, Enterprise AI https://www.techtarget.com/searchenterpriseai/tip/4-main-types-of-AI-explained

A. Phillips-B. Taylor, Sulla gentilezza, Internazionale, 2018 https://www.internazionale.it/notizie/adam-phillips/2018/05/31/gentilezza

D. R. Polgar, Thanks, Robot! Humans are Showing Kindness with Their AI Helpers., Big Think, 2017 https://bigthink.com/sex-relationships/thanks-robot-how-should-we-be-treating-our-ai-helpers/

E. Prem, From ethical AI frameworks to tools: a review of approaches, in AI and Ethics, 2023

S. Primiceri, Marco Aurelio e il culto della gentilezza, L'AltraPagina.it, 2017 https://www.laltrapagina.it/mag/marco-aurelio-e-il-culto-della-gentilezza/

I. Qian-M. Xiao-P. Mozur-A. Cardia, Four Takeaways From a Times Investigation Into China's Expanding Surveillance State, in The New York Times, 2022

D. Rampado, ETICA E GENTILEZZA NEL LAVORO « Donatella Rampado, Donatella Rampado, 2019 https://www.donatellarampado.com/etica-e-gentilezza-nel-lavoro/

Reuters, IBM to pause hiring in plan to replace 7,800 jobs with AI, Bloomberg reports, in Reuters, 2023

A. Rockwell, The History of Artificial Intelligence, Science in the News, 2017 https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

R. Roni, La gentilezza verso l'altro parte dalla vita interiore personale, Blasting News, 2022 https://it.blastingnews.com/opinioni/2022/09/la-gentilezza-verso-l-altro-parte-dalla-vita-interiore-personale-003565158.html

N. Routley, What is generative AI? An AI explains, World Economic Forum, 2023 https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/

L. Rowland-D. Klisanin, Cyber-kindness: Spreading kindness in cyberspace, Media Psychology Review, 2018 https://mprcenter.org/review/cyber-kindness-spreading-kindness-in-cyberspace/

S. Russell-D. Dewey-M. Tegmark, Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter, Future of Life Institute, 2015 https://futureoflife.org/open-letter/ai-open-letter/

L. Saettone, «Coded Bias», così abbiamo delegato il nostro razzismo agli algoritmi, Agenda Digitale, 2021 https://www.agendadigitale.eu/cultura-digitale/coded-bias-cosi-abbiamo-delegato-il-nostro-razzismo-agli-algoritmi/

L. Salonia, Gentilezza ai tempi di Covid: i piccoli gesti che ci fanno bene - Foto iO Donna, iO Donna, 2020 https://www.iodonna.it/benessere/salute-e-psicologia/gallery/gentilezza-ai-tempi-di-covid/

I. Sample-I. S. E. di Scienze-I. Sample, Computer says no: why making AIs fair, accountable and transparent is crucial, in The Guardian, 2017

R. Sanderson-J. Mcquilkin, Many Kinds of Kindness: The Relationship Between Values and Prosocial Behaviour, in Values and Behavior: Taking a Cross Cultural Perspective, 2017, p. 75–96

I. Sarker, AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems, in SN Computer Science, fasc. 3, 2022

SAS, Computer Vision: What it is and why it matters https://www.sas.com/en_us/insights/analytics/computer-vision.html

SAS, Cos'è il Natural Language Processing (NLP)? https://www.sas.com/it_it/insights/analytics/what-is-natural-language-processing-nlp.html

SAS, Natural Language Processing (NLP): What it is and why it matters https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

R. Schmelzer, Are We Overly Infatuated With Deep Learning?, Forbes, 2019 https://www.forbes.com/sites/cognitiveworld/2019/12/26/are-we-overly-infatuated-with-deep-learning/

S. serra-S. Greenhouse, US experts warn AI likely to kill off jobs – and widen wealth inequality, in The Guardian, 2023

joyce shen, Kindness & AI, Medium, 2022 https://medium.com/@joycejshen/kindness-ai-dd4272f88299

N. Spagnuolo, Leadership gentile: moda o cambiamento reale? - Il Sole 24 ORE, 2022 https://www.ilsole24ore.com/art/leadership-gentile-moda-o-cambiamento-reale-AE5RzwiB?refresh_ce

S. Srinivasan, Data Breach at Equifax - Case - Faculty & Research - Harvard Business School, 2017 https://www.hbs.edu/faculty/Pages/item.aspx?num=53509

M. T. Stecher, La storia dell'intelligenza artificiale, da Turing ad oggi, CyberLaws, 2018 https://www.cyberlaws.it/2018/la-storia-dellintelligenza-artificiale-da-turing-ad-oggi/

Stefanelli & Stefanelli Studio Legale, Raccolta fonti normative sull'AI, Stefanelli & Stefanelli Studio Legale https://www.studiolegalestefanelli.it/it/raccolta-fonti-normative-intelligenza-artificiale

R. Stevenson, Moxie the Conversational AI Robot Teaches Children Kindness, SAMA, 2021 https://www.sama.com/podcast/moxie-the-robot-teaches-children-kindness-conversational-ai-child-development/

G. Stratta, Leadership gentile: il nostro nuovo "way to be", 2021 https://www.enel.com/it/azienda/storie/articles/2021/05/leadership-gentile-nuovo-modo

G. Stratta-P. Cervini, Nuovi modelli di gestione: l'evoluzione verso la leadership gentile, 2022 https://www.hbritalia.it/giugno-2022/2022/05/31/news/nuovi-modelli-di-gestione-levoluzione-verso-la-leadership-gentile-come-le-aziende-possono-coltivare-una-fiducia-vera-il-caso-di-enel-15294/

O. Strelkova, Three types of artificial intelligence, 2017

C. Stroud, Cambridge Analytica: The Turning Point In The Crisis About Big Data, Forbes, 2018 https://www.forbes.com/sites/courtstroud/2018/04/30/cambridge-analytica-the-turning-point-in-the-crisis-about-big-data/

G. L. Teppati, Leadership gentile, Manageritalia, 2022 https://www.manageritalia.it/it/management/leadership-gentile

K. Terrell Hanna-I. Wigmore, What is the Gartner Hype Cycle? – TechTarget Definition, WhatIs.com, 2023 https://www.techtarget.com/whatis/definition/Gartner-hype-cycle

M. Toh, 300 million jobs could be affected by latest wave of AI, says Goldman Sachs | CNN Business, CNN, 2023 https://www.cnn.com/2023/03/29/tech/chatgpt-ai-automation-jobs-impact-intl-hnk/index.html

A. Toosi-A. Bottino-B. Saboury-E. Siegel-A. Rahmim, A brief history of AI: how to prevent another winter (a critical review), in PET Clinics, fasc. 16, 4, 2021, p. 449–469

L. Tremolada, Blog | La storia dell'intelligenza artificiale in due grafici, Info Data, 2022 https://www.infodata.ilsole24ore.com/2022/12/24/la-storia-dellintelligenza-artificiale-in-due-grafici/

Université de Montréal, The Declaration - Montreal Responsible AI, 2018 https://www.montrealdeclaration-responsibleai.com/the-declaration

G. Vaia-N. UL-Ain-E. Gritti-M. Bisogno, Ethical Systems for Artificial Intelligence, in Intellectual Capital, Smart Technologies and Digitalization, Springer, 2021

V. Vinge, TECHNOLOGICAL SINGULARITY https://frc.ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html

I. Vrtaric, History of Generative AI. Paper explained., Artificialis, 2023 https://medium.com/artificialis/history-of-generative-ai-paper-explained-6a0edda1b909

I. Wigmore-D. Shao, What is social robot? | Definition from TechTarget, Enterprise AI, 2022 https://www.techtarget.com/searchenterpriseai/definition/social-robot

J. Wiles, What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle, Gartner, 2022 https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

K. Wilson, Beware the AI hype cycle – managing expectations in the age of ChatGPT, UKTN | UK Tech News, 2023 https://www.uktech.news/guest-posts/ai-hype-cycle-expectations-20230519

World Economic Forum, The Future of Jobs Report 2023, 2023

Z. Yang, The Chinese surveillance state proves that the idea of privacy is more “malleable” than you'd expect, MIT Technology Review, 2022 https://www.technologyreview.com/2022/10/10/1060982/china-pandemic-cameras-surveillance-state-book/

A. Yela Aránega-C. Gonzalo Montesinos-M. T. del Val Núñez, Towards an entrepreneurial leadership based on kindness in a digital age, in Journal of Business Research, fasc. 159, 2023, p. 113747

H. Yu-Z. Shen-C. Miao-C. Leung-V. R. Lesser-Q. Yang, Building Ethics into Artificial Intelligence, arXiv, 2018

E. Yudkowsky, The Open Letter on AI Doesn't Go Far Enough, Time, 2023 https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/

C. Zaccarelli, Per la Giornata mondiale della gentilezza, ecco 8 startup che ne hanno fatto la loro filosofia, LifeGate, 2022 https://www.lifegate.it/startup-giornata-mondiale-della-gentilezza

A. Zampa, Biologia della gentilezza: 6 scelte per benessere, salute e longevità, LifeGate, 2020 https://www.lifegate.it/biologia-della-gentilezza-libro-interviste-lumera-de-vivo

R. Zazza, Blog | Metti etica e gentilezza nel curriculum e il lavoro del futuro è tuo, Alley Oop | Il Sole 24 Ore, 2022 https://alleyoop.ilsole24ore.com/2022/11/18/metti-etica-gentilezza-nel-curriculum-lavoro-del-futuro/

J. Zou-L. Schiebinger, AI can be sexist and racist — it's time to make it fair, in Nature, fasc. 559, 7714, 2018, p. 324–326

M. Zur, Homo Deus: After God and Man, Algorithms Will Make the Decisions, Yuval Noah Harari, 2019 https://www.ynharari.com/homo-deus-after-god-and-man-algorithms-will-make-the-decisions/