UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE POLITICHE, GIURIDICHE E
STUDI INTERNAZIONALI

Master’s Degree in Global Social Policies and Security Issues

THE EU AI ACT: SHAPING AI'S ROLE IN COUNTER-TERRORISM

Relatrice: Prof. VALENTINE LOMELLINE

Laureando: MATILDE MASOTTI

Matricola N. 2092744

A.A. 2023/2024

INDEX

LIST OF ABBREVIATIONS

AI – Artificial Intelligence

AML – Anti-Money Laundering

ANN – Artificial Neural Networks

CNN – Convolutional Neural Network

DEC – Digital Equipment Corporation

DDOS – Distributed Denial of Service

DP – Dependency Parsing

DOS – Denial of Service

EU – European Union

FR – Facial Recognition

GAN – Generative Adversarial Network

GPS – Global Positioning System

HAR – Human Activity Recognition

ICT – Information and Communication Technology

LAWS – Lethal Autonomous Weapon Systems

LISP – List Processing

LSTM – Long Short-Term Memory

ML – Machine Learning

MIT – Massachusetts Institute of Technology

NER – Named Entity Recognition

NL – Natural Language

NLP – Natural Language Processing

POS Tagging – Part-of-Speech Tagging

RNN – Recurrent Neural Network

RL – Reinforcement Learning

SAIL – Stanford Artificial Intelligence Laboratory

TDC – Dark Crawler

TR – Terrorism Financing

XAI – Explainable Artificial Intelligence

INTRODUCTION

This thesis explores the complexities and the wide opportunities of Artificial Intelligence, focusing in particular, on its employment in the prevention and fight against terrorism, and it proposes to reflect on which could be the limits between its strategic use and the potential, serious violations of human rights that could derive from such applications. We are living today in an era where technological progress proceeds at such a speed as surpassing, in several cases, the capacity of societies and institutions to adapt, regulate the use and fully understand the consequences; in this scenario, AI not only presents itself as one of the most impactful innovation of our times, but it also introduces itself as a fertile ground for ethical matters of crucial importance and the legal responsibilities on a scale never seen before. Despite the common perception of AI being a technology reserved for the most advanced fields, the majority of the population is not aware of being constantly surrounded by it, and in most cases, to use it in an almost oblivious manner: vocal assistants that apprehend our requests, recommending algorithms that influence our consumption choices, facial recognition systems that observe us and analyze in public and private spaces, predictive models that orientate our online experiences. This capacity of AI to insert into every aspect of contemporary life, almost without being aware of it, amplifies the implications of its application in more delicate sectors, such as the one of security, where decisions can have significant and long-lasting consequences on human life and individual rights. Through an in-depth analysis that extends from the first phases of the birth and evolution of artificial intelligence systems to more recent developments, going through the employment in strategic sectors such as economy, healthcare, transports, energy, and security, this thesis delineates an overview of the state-of-the-art of AI, dwelling on the applications and regulatory perspectives that emerge from the contemporary context. Meanwhile, the thesis navigates the ethical and judicial consequences of such a powerful and pervasive technology that, on one hand, offers innovative solutions for responding to security challenges and the prevention of terrorist threats, and on the other, it raises fundamental questions regarding the line between the strategic use of AI and the respect for human rights and individual dignity, recalling the urgency of reflecting on which are the limits to establish, but also on the accountability that should be shared among legislators, governments, enterprises, and citizens.

In the first chapter, the historical and technological foundations that have led to the current development of AI are delineated, starting from the pioneering theories to modern models that have shaped an ever-evolving discipline. The concept of AI, today omnipresent, dates back to the theoretical studies carried out by Alan Turing and the contribution of John McCarthy, who minted the term “artificial Intelligence” and introduced instruments such as the language LISP. The chapter continues with the work of the pioneers Yann Lecun, Yoshua Bengio, Geoffrey Hinton on the development of machine learning and deep learning, deepening into key architectures such as Convolutional networks for the visual recognition, Recurrent networks for the elaboration of sequential data and adverse generative networks, that have revolutionised the generation of synthetic data. Systems such as Adverse generative networks and reinforcement learning, introduced in the works of Jan Goodfellow on one side, and Rich Sutton and Andrew Barto on the other, have been critically analysed, underpinning the architecture of contemporary artificial intelligence system. Additionally, the importance of the Transformer model is described in the cornerstone paper “*Attention is All You Need*” by Ashish Vaswani which has transformed the elaboration of natural language, opening up the way to models such as GPT. Subsequently, the applications and concrete impacts of AI in sectors like finance, healthcare, energy, transport, and security are examined, where AI is improving efficiency, automation, and safety, offering a solid comprehension of its current employment and its transformative potential in modern societies. At the end of the first chapter, it is underlined the importance of a regulatory framework for the discipline, a sector where the rapid acceleration of technological capacities has started delineating what some scholars in their academic work define as a real “AI arms race” among global powers. Several states, aware of the strategic potential of AI for national security, economy, and global influence, have started competitive plans to develop and implement increasingly advanced AI technologies. This concept of technological competition, on one side incentives innovation, on the other it raises ethical and security concerns, making it urgent a regulation that can orientate the developments of AI toward common goals, respectful of all human rights. In this global race, the European Union has adopted a more restrictive and prudent approach, with the end goal of regulating AI in a way to preserve values such as transparency privacy, and protection of human rights. Such an approach, culminated with the new EU AI ACT 2024/1689, tries to promote the ethical and safe use of AI, limiting its most invasive and risky applications. As analyzed in the second chapter, this regulation represents a crucial step toward the efficient and homogeneous regulation of AI inside the EU. This Act proposes to balance technological innovation with the protection of the fundamental rights of citizens, delineating the rules and limitations to its employment in sensitive sectors. EU, indeed, distinguished itself for a proactive approach aimed to guarantee that emerging technologies respect democratic values and human rights.

The third chapter starts with a brief description of the evolution of terrorist threats, an analysis of the growing complexities of challenges that counterterrorism faces at the global level, and the role of AI in contrasting them. A framework of the transformations in the operations of terrorists is delineated, with an increasing tendency to the use of advanced technologies and the diffusion of propaganda messages through digital channels. In response, security forces have adopted more sophisticated tools, among which AI applications, which are indeed, applied in strategic and innovative ways, aiming to identify, monitor, and prevent suspected activities, particularly thanks to technologies such as facial and vocal recognition. These tools allow us to track with more accuracy the potential suspects and detect behavioral anomalies inside massive flows of visual and sound data. Alongside the recognition, the crucial role of Natural language processing is described - extensively shaped by foundational insights and methodologies of the discipline - employed to monitor the suspected communications and decoding encrypted messages, often conveyed through social media and hidden networks. These instruments are pivotal in identifying risk signals and intervening preemptively, representing a crucial step forward compared to traditional intelligence methods. In contexts of asymmetric conflicts of scenarios of hybrid warfare, AI is additionally used as an enabler, reinforcing the capacities of drones or autonomous systems, that today are employed not only for surveillance activities but also as real autonomous weapons. These devices can make decisions on the battlefields in real-time, increasing the efficiency of operations but posing at the same time, complicated ethical dilemmas, especially in terms of accountability of the perpetrated actions. The chapter further extends to the use of AI in war gaming, namely the advanced simulations employed to test and develop strategies in response to terrorist attacks. Through war gaming, security forces are capable of emulating extremely realistic scenarios, predicting the actions of adversaries, and improving defense strategies, aligning with approaches outlined in key security frameworks and documents used by security organizations such as NATO. This chapter, which constitutes the core of the thesis, provides a detailed analysis of the potentialities of such technologies in the context of anti-terrorism, enlightening the operational advantages of AI systems. At end of the third chapter, the thesis examines a complex scenario characterised by both opportunities and risks offered by AI. The first section analyses the technical challenges, such as the problem of false negatives and false positives that can compromise the reliability of AI systems, the problem of admissibility of AI-generated material in legal proceedings, the black box nature of AI systems, namely the lack of transparency in automated decisional processes, and finally, the dynamics and relations between the private and public sector, essential to optimise the resources and improve the efficiency of anti-terrorist operations. The second section, on the other hand, deals with the dark side of AI, namely its employment by terrorists. Here the malign applications are describe, such as the reinforcement of cyber capacities, the enabling of physical attacks and the other uses, such as the creation of deep fakes, online propaganda for radicalisation and recruitment. It is clear that the growing power of these instruments requires a delicate balance, since they risk compromising the fundamental rights of people, such as privacy and personal freedom, if not applied with extra attention and respect to the ethical limits.

This topic is the focus of the final chapter, which starts with the question: *is artificial intelligence a strategic asset for security or a potential risk to fundamental human rights*? The ethical and juridical implications of the use of AI by governments and private corporations are described, examining how these technologies, if employed without rigid controls, can lead to violations of fundamental human rights. The thesis elaborates, in particular, central topics such as the right to privacy, increasingly threatened by the massive collection of data, and algorithmic discriminations, that risk reinforcing pre-existing prejudices, translating to concrete disadvantages for some categories of individuals. An additional element is the right to a fair trial, which might be compromised by the use of algorithms in legal decisions and judicial applications, especially in the absence of transparency and clear accountability in case of errors. Particular focus is given to the responsibility and necessity of transparency in AI systems that support automated decisions, which could drastically influence the lives of people without adequate human supervision. This final part depicts also a critical analysis of the use of AI with the purpose of surveillance and repression, highlighting how, in this case in China, these technologies have been employed to monitor and repress minorities, through a pervasive control that constitutes a large-scale violation of freedom and dignity rights. This exemplar case underlines the risks associated with the use without ethical limits and pushes us to reflect on the necessity to develop a global ethical framework that guides the adoption of artificial intelligence. This framework not only has the purpose of protecting rights and freedoms, but also promoting a culture of collective responsibility and respect for human dignity, indispensable for AI to operate safely and equally, beneficially for the whole society.

Artificial intelligence, with its potential and intrinsic risks, position itself at the heart of the contemporary debate regrading security and protection of human rights. The geopolitical competition surrounding the AI is reaching level of intensity with no precedents. This rivalry is not solely limited to the technological development, but raises questions about security standards and ethical concerns. While moving toward a future where AI might assume an increasingly predominating role, it is crucial to consider how policies and practices could evolve to guarantee that the technology truly serves the common good, without compromising human beings.

CHAPTER 1 - ARTIFICIAL INTELLIGENCE IN TODAY'S WORLD

1.1 The birth and Evolution of Artificial Intelligence

"…every aspect of learning or any other feature of intelligence can in principle be so precisely stated that a machine can be made to simulate it."[1]

Recent progress in technologies known to be part of the Artificial Intelligence (AI) discipline, has been extraordinary. The success has contributed to the creation of great expectations on what can be done with current AI technologies. This enthusiasm to exploit AI systems' potential is not new. The above-mentioned citation was written at the end of 1955 for a conference in 1956, which marked the birth of Artificial Intelligence as a recognized discipline. The Dartmouth Conference was held in an important period for the developments of AI, which seems to be constant. Looking back in history, it is possible to divide the journey of AI into two main cycles of "boom and bust" in its first 45 years of life. During expansion periods, impressing demonstrations would capture the attention of the public and arouse great expectations and investments, often amplified by incredible narratives in the media. Decline periods, on the other hand, happened when the promised capabilities could not be generalized. Governmental programs would be cancelled, investments would cease and the media attention would decrease or became critique. Looking back, it is possible to understand that AI is enjoying a third expansion period, which seems to propose some common characteristics with previous ones, both socially and technologically.

The previous and following years of the Dartmouth conference experienced developments that still today influence the discipline. In 1949, Arthur Samuel, American informatics, began working in the field of automated learning, which made him capable of teaching computers how to play board games. By the end of the 1950s, he had developed a computer program able to beat human players in the game of checkers. The program to play checkers would use a form of mechanic learning: it would try to record every possible chessboard configuration with an associated score (victory probability). Nevertheless, this approach would not adapt to more complex games; the mechanic learning could not generalize to more complicated cases. At that time, this aroused an incredible

---

[1] Words from the Dartmouth AI Conference Call for Papers, August 31, 1955, a proposal for the Darmouth Summer Research Project on Artificial Intelligence, http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

reaction: a computer had beaten a human being in a game that necessitated both intelligence and strategic thought. Just before the conference, Allen Newell, American psychologist, informatics and mathematician, developed a program capable of printing images of a map, just by using the printer's fonts (letters, numbers, and punctuation marks) as symbols. This result influenced his colleagues at RAND[2], especially Herbert Simon, American psychologist, informatics and economist, who understood that computers are not simply quick calculators, but devices for symbolic manipulation and could be used to simulate decision-making and other aspects of human intelligence. Subsequently, Newell and Simon created a program called "*Logic Theorist*" that managed to develop logic tests of the majority of fundamental theorems contained in Principia Mathematica. In the mid-50s, a computer program capable of producing automatic demonstrations of mathematic theorems seemed to show that computers could simulate human intellect. They kept on working on the hypothesis of the "*A physical symbol system has the necessary and sufficient means for general intelligent action*"[3]. According to them, both computers and human minds are symbol systems: this idea is still a foundation of the AI's functioning[4]. Newell and Simon, after the success of their program "Logic Theorist", developed the "General Problem Solver" (GPS) to create a more versatile calculus program. However, despite some initial successes, the GPS could solve only simple problems and less efficiently than other programs created for specific issues. Similarly to the checker's program, it could not generalize to all cases.

An additional important development was the attempt to make the computer talk in English. The program ELIZA was designed to simulate the conversational style of a psychotherapist. ELIZA could accept human conversational inputs and separate main words to adapt them to predefined answers, by using a simple array of rules. This has been the first work in Natural Language Processing, a field that has progressed exponentially and that is still fundamental in today's works. Furthermore, in this period, many were the efforts to shape the animate brain's neural structure to duplicate human activity, namely trying to create models of the human brain's functioning, by studying the connections among neutrons and the way these connections allow one to think, learn,

---

[2] The RAND Corporation is an American nonprofit global policy think tank,[1] research institute, and public sector consulting firm

[3] Newell, A. &Simon,H.A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, 19,113–26.

[4] Richbourg, R. F. (2018). Deep Learning: Measure Twice, Cut Once. Institute for Defense Analyses. http://www.jstor.org/stable/resrep36394

and make decisions. The goal was to understand how the human brain functions, to then build machines and programs that can imitate, at least partly, the processes, making computers able to do things that physical people normally do. One example at the time was the Percepton[5]: developed by Frank Rosenblatt in 1958, it was an artificial neural network designed to be trained and able to learn how to recognize digits and specific images provided in an encrypted way. This system could improve its operations when exposed to new data, a characteristic similar to human learning. Percepton represented one of the first practical applications in the attempt of simulating the human brain, through the use of algorithms that simulated neutrons' activities. New York Times reported that the United States Marines financed this research with the expectation that Percepton could evolve until becoming "*the embryo of an electronic computer that the Navy expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence*"[6]. The period experienced also the development of computers that demonstrated operations similar to human ones but applied to the recognition of images and digits. This model created the foundations for artificial neural networks that now allow the functioning of deep learning. Many others were the developments dating back to this period, but the mentioned ones were the most significant. The researchers tried to habilitate machines to carry out tasks associated with human intelligence, such as perception, natural language communication, formal reasoning, and strategic thought used in gaming. Given the reactions to these first results, the governmental investment in AI systems started growing. The program to play checkers would use a form of mechanic learning: it would try to record every possible chessboard configuration with an associated score (victory probability). Nevertheless, this approach would not adapt to more complex games; the mechanic learning could not generalize to more complicated cases.

Another field that had some success was automated translation, especially in the sector of language processing, which led to the creation of more financed and long-lasting programs. The governments' goal was to translate documents and technical publications from languages such as Russian to English. After many years and dollars spent, the Automatic Language Processing Advisory Committee (ALPAC, 1966) reported "*We do not have useful machine translation and*

---

[5] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*, 386–408.

[6] Olazaran, M. (1996). A Sociological Study of the Official History of the Perceptrons Controversy. *Social Studies of Science*, 26, 611–59.

therein not the immediate or predictable prospect of useful machine translation"[7]. This created a situation of stalemate of efforts, resources, and investments in the development of these types of technologies. Other types of research as well were put under discussion: the British parliament, for instance, instructed James Lighthill to evaluate the progress of AI in the United Kingdom. The "Lighthill Report" concluded that the British progress was minimal and what was obtained, was derived from more traditional disciplines. In the US, DARPA[8] reduced the Sustainment to AI research, after years of programs that had not achieved the predetermined ambitious goals. This led to the beginning of the first "*AI Winter*". Research did not stop completely, but with a slower rhythm and less finances[9].

At the end of the 1970s, a new technology linked to artificial intelligence, known as "*expert systems*", started showing notable progress in the automation of human competencies. These systems were based on symbolic reasoning and on the extraction of knowledge by experts to replicate their judgments and conclusions about specific fields. Their symbolic and based-on logic rules (if X, then Y) nature, allowed them to explain their decisional processes, useful not only for decision-makers but also for the developers of such systems. A fundamental difference compared to the previous software was that the expert systems focused on the resolution of specific problems, without trying to understand or replicate human intelligence in its whole. They were designed to face a particular issue, such as the configuration of a computer or diagnosis of a disease, by using knowledge gathered by field experts. This knowledge was organized in bases of knowledge, a sort of archive of specialized information, that allowed systems to solve problems precisely inside the specific domain. A concrete example is the Digital Equipment Corporation (DEC), which lost money due to its vendors that could not configure complex computer orders correctly. To solve this issue, DEC developed an expert system, called XCON, that helped with the configuration. Thanks to XCON, DEC saved about 25 million dollars each year.

---

[7] Report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences National Research Council. Language and Machines, computers in translation and linguistics (August 20, 1965), Publication 1416. https://www.mt-archive.net/50/ALPAC-1966.pdf

[8] Defence Advanced Research Projects Agency

[9] Richbourg, R. F. (2018). Deep Learning: Measure Twice, Cut Once. Institute for Defense Analyses. http://www.jstor.org/stable/resrep36394

This led to a new wave of optimism, with new investment flows from both governments and companies. New enterprise arose to support the development and use of these expert systems. In 1985, it was estimated that this market passed from 80 million dollars in 1983 to 3-12 billion in 1990, reaching 50-120 billion in 2000, representing 20% of the Informatics industry's revenues.

At the end of the 1980s, the enthusiasm for AI again started decreasing. As reported in the Economist in 2002 "*Like big hairdos and dubious pop stars, the term artificial intelligence was big in the 1980s, vanished in the 1990s*". The expert systems technology was revealed to be difficult to maintain and not very adaptable to be expanded to new applications. Consequently, especially the government of the US reduced investments in the ambitious Strategic Computing Initiative, whereas other countries interested in this research, like Japan, modified their projects into the "fifth generation computer", eliminating every goal linked to AI. Industries that were born during the enthusiasm for such systems, and specialized in the construction of specific and highly profitable informatics machinery, went bankrupt since the return on investment for their clients was incredibly low. Also, other industries, that offered environments and tools to build expert systems failed and oriented towards other areas, such as the development of object-oriented technologies. This led to the AI discipline experiencing its second AI winter. Nonetheless, as during the first winter, research did not stop completely, but slowed down notably. One example is the American Association of Artificial Intelligence (AAAI), which has been a key event for the community. During the boom period, the number of articles advanced almost reached 900 for the conference in 1990. In 1997, a little more than 300 articles were actually presented. Researchers started defining their works with more specific terminology such as "*automated learning*", and "*neural networks*" avoiding the term "AI".

The reasons behind the decline of Expert systems are an object of debate. Some believe that they were simply integrated into standard technologies of decisional support; others reckon that problems were deeper. One thought is that each expert system was extremely efficient in a limited field, but completely useless in a wider context. One additional difficulty was the huge complexity and costs for the creation and maintenance of the knowledge bases, namely the array of rules and facts that represented the human experience. The dependence on specific knowledge and data allowed the expert systems to be successful, but at the same time to decline.

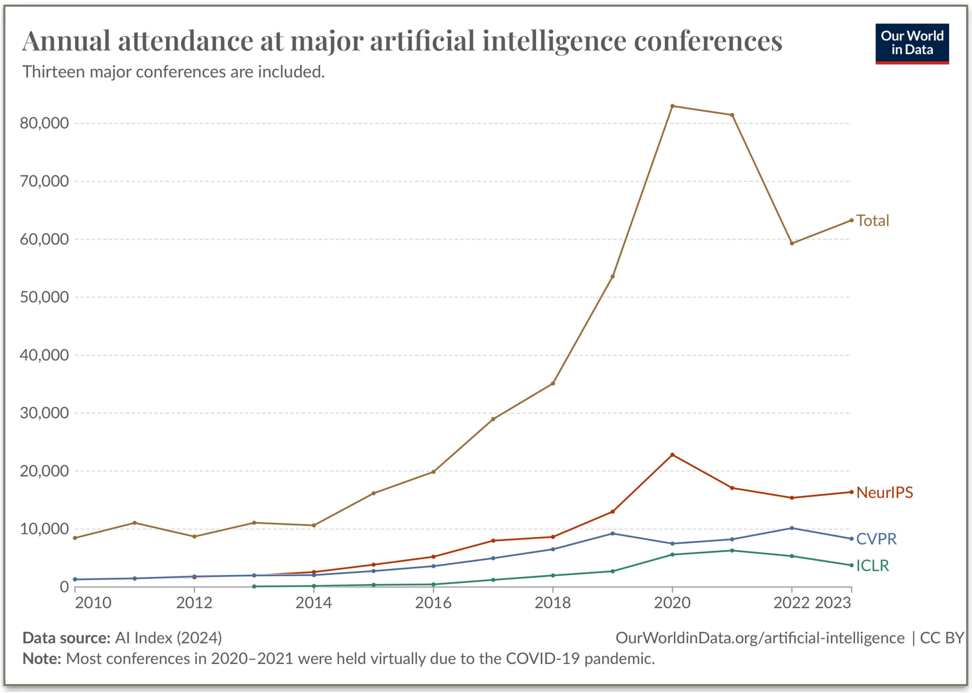


Figure 1. Attendance to the three major conference on AI[10]

In 1987, the IEEE organized its first conference on neural networks, attracting 1.800 participants. This was only the beginning of these networks' development, an enthusiasm that still grows today. The AAAI began organizing conferences in 1984, whereas the International Joint Conference on AI (IJCAI) is a paramount international event for this discipline. Again, the conference on Neural Information Processing Systems (NIPS) was rated in 1987 as a normal meeting, for then becoming the biggest conference on AI. Today, we are experiencing the third boom of AI, alimented by spectacular results obtained thanks to Deep Learning and Neural networks.

1.1.1 Alan Turing and John McCarthy: founding fathers of Artificial Intelligence

Alan Turing (1912-1954) was a British mathematician, logician, and cryptographer, considered one of the founding fathers of informatics and Artificial Intelligence. During the Second World War, he worked on the deciphering of the Enigma code, used by Germans, contributing notably to the victory of the Allies. Turing developed several important concepts among which the Turing machine, a theoretical model of computation that serves as the foundation of modern computer

---

[10] Data from Our World in Data https://ourworldindata.org/grapher/attendance-major-artificial-intelligence-conferences

science, and the Turing test, the criterion to assess AI by determining whether a machine can show intelligent behavior indistinguishable from a human one. Already in 1936, he proposed the concept of a *universal machine*, a programmable calculating machine that could work on different algorithms: he introduced the differentiation between "hardware" (the structure of the machines) and "software" (algorithm). Until that moment, the first models of computers were mechanic machines able to carry out a limited number of tasks and they were not programmable.

"*Can machines think*?", this is the question Alan Turing proposed in his work *Computing Machinery and Intelligence*, 1950. First of all, he proposed the clarification of the terms "machine" and "thinking", underlying that based on their common use, might lead to misunderstandings. Turing rephrased the matter in terms of a game, called "the imitation game". This game is played by three people, a man (A), a woman (B), and the interviewer (C). The latest is closed in a room, separated from A and B, and knows the man and woman with the labels X and Y: at the end of the game he need sto declared who is the woman and and who is the man, as "X is A and Y is B" or the other way round "X is B and Y is A". This imitation game is a mental experiment designed to explore the machines' capacities to exhibit intelligent behaviors. If in the game, a machine substitutes the man and manages to deceive the interviewer with his answers, this might suggest that machines can imitate human intelligence and thinking. The rules of the game set that the machine can play any behavior to deceive the interviewer: for instance, it could delay the answer for a calculus that could unmask its computational abilities. Turing affirmed that in 50 years (around the year 2000), it would be possible to program computers to perfectly play the imitation game and deceive the interview with more than a 70% chance. This methodology went down in history as the Turing test, and it represents still today a benchmark to understand whether an algorithm can be labeled as "intelligent" or not.

The 1950 paper[11] is considered an absolute masterpiece of knowledge and vision, not only because it formulated the concept of the Turing test, but the mathematician also directed further research towards intelligent algorithms, by introducing the concept of "machine learning". Turing's view is that the functioning of an adult mind depends on the initial state, namely the birth, and on the education the person received subsequently. Instead of attempting to replicate the cognitive

---

[11] Alan Turing. *Macchine calcolatrici e intelligenza* (1950) https://disf.org/files/macchine-calcolatrici-e-intelligenza.pdf

processes of an adult mind directly, by subjecting the program to a structured educational process, it is possible to gradually develop its reasoning capabilities to align more closely with those of an adult mind. This approach emphasizes a developmental trajectory that mimics natural learning and cognitive growth. Alan Turing laid down the foundations for the further development of AI, through his innovative concepts of machine learning. By exploring the potential of programmable machines and their learning abilities, Turing inspired many works, underlying the importance of computational theories in the creation of systems capable of imitating human reasoning.

The main turning point happened then in 1956, when John McCarthy, an American researcher and mathematician, organized a conference at Dartmouth College. On this occasion, McCarthy proposed the name "*artificial intelligence*", a term that would become a synonym for the entire discipline. His goal was to understand how to build a machine capable of imitating human cognitive functions, such as the use of language, resolution of problems, and learning. While the work of Turing was crucial in creating the conceptual foundations of the discipline, McCarthy contributed significantly to the practical and technological development of AI. Among his most significant contributions, is the creation of LISP, (Programming language) which became essential for Artificial Intelligence research and that is still in use today. Moreover, McCarthy developed the concept of "*timesharing*", a system that allowed several users to share the resources of a computer contemporarily, opening the way to the birth of computer networks, including Arpanet, the precursor of the Internet. In the meantime, Turing contributed to a significant advancement of computer science, also thanks to his work on the first digital computers and his contributions during the Second World War to the deciphering of codes. However, his vision of machines being able to learn and adapt would be further developed during the following years by researchers like McCarthy.

In the 1960s, McCarthy founded a project on Artificial Intelligence at MIT, where pioneering progress in sectors like robotics and automated reasoning were made. His students in this period, developed the first program able to play chess convincingly, demonstrating that machines could face tasks considered typically human. His contributions expanded also to the creation of a new laboratory at Stanford, the Stanford Artificial Intelligence Laboratory (SAIL), which became a center of excellence for the research on AI. Ideas and projects that were born inside the walls of SAIL significantly influenced the development of the technological sector in Silicon Valley, contributing to the creation of enterprises like Microsoft and Amazon. From the years 2010s on, due to the advent of machine learning and the growing computational power[12], AI has experienced a renaissance. This period was characterized by significant progress that revolutionized the social technological framework. In particular, the introduction of deep learning networks, which will be discussed later on, inspired by the function of the human brain, allowed machines to learn data in an increasingly autonomous and effective way. This led to the development of an AI system that is capable of overpassing human capabilities in specific tasks, such as image recognition, vocal recognition, or strategic games[13].

1.2 Machine and Deep Learning: the foundations of Artificial Intelligence

Machine learning technology feeds many aspects of modern societies, from research on the web to the filtering of social network content, to the recommendations in e-commerce websites, and it's more and more present in technological devices such as mobile phones and cameras. These systems are used to identify objects in the pictures, transcribe the talking into text, and combine news and products with the user's interests and needs. The conventional techniques of machine learning were limited to the capacity of elaborating natural rough data. For centuries, the construction of a pattern recognition system required accurate and expert engineering and expertise in the field, to design a tool that could extract rough data (such as pixels in an image) and transform it into an internal representation proper for recognition or classification. The layering of representations allows a machine to receive rough data and to discover automatically the necessary representations for detection and classification. Deep learning methods are learning techniques of multilayer representations, obtained by composing simple but non-linear modules, each transforming a layer (starting from the rough input like lines and shapes) into a representation of a higher level, slightly more abstract. With a sufficient number of these transformations, it is possible to learn complicated functions. For the classification tasks, superior levels amplify aspects of the inputs important for the discrimination and suppress irrelevant variations.

---

[12] The Computing power is defined as the number of the processing that is been initiated in a computer system and also implements the quantity of memory, speed of the processors, and the number of processors that are associated in the case of the computing system. In addition to that, the computer processing power that is associated in the case of technology tends to give a better output. Computing power is a crucial component that is mainly associated in terms of tasks that are beneficial for the user.

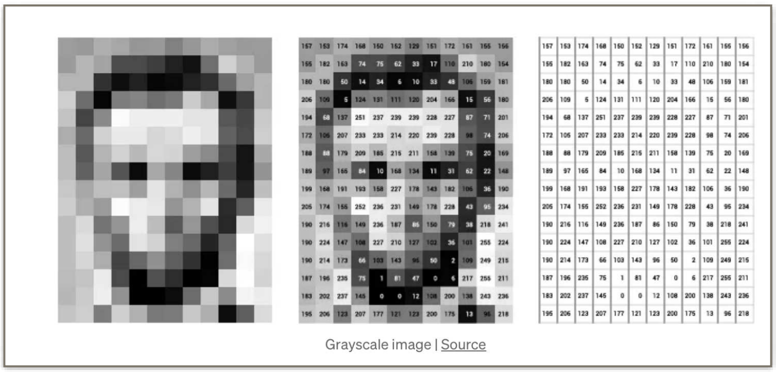
[13] Lawrence Livermore National Laboratory. (n.d.). *The birth of artificial intelligence* (AI) research. LLNL. Retrieved from https://st.llnl.gov/news/look-back/birth-artificial-intelligence-ai-research

Imagine the development of a deep learning system for the recognition of animals in images: through the first layer the system examines sizes, and identifies fundamental elements such as borders and lines; the second level recognizes more complex shapes, such as paws and ears, by combining the previously identified borders and lines: the third level integrates these shapes to recognize the animal. Each layer increases the complexity of the recognition, allowing the system to ignore the irrelevant variations and shapes, and focus on the distinctive characteristics.

Deep learning is experiencing huge progress in the resolution of problems that have resisted years of efforts from the AI community. It was revealed to be extremely efficient in discovering intricate structures in data, resulting in applicability to several scientific and governmental sectors. It surpassed records in image recognition and vocal recognition, and it has been important in reconstructing cerebral circuits and for foreseeing the effects of non-codified DNA mutations on genetic expressions and diseases. Moreover, it showed promising results in various tasks of Natural Language comprehension, classification of topics, sentiment analysis, question answering and linguistic translation. It will most certainly have further success since it requires little manual engineering.

The most common type of automated learning is supervised learning: it is a machine learning technique trained by using an array of labeled data. First of all, images with labels are collected, such as buses or cars. During the training, the system analyzes each image and generates an output, namely a score for each category. Subsequently, an objective function is used to calculate the error between the expected and the desired scores. The system then regulates its internal parameters, called weights, to improve the precision, continuing until when it is capable of generalizing on new data. The “weights” therefore, are used to calculate the output, starting from the input. During the training, the system tries to improve its results and reduce the errors by modifying these weights. This process continues until the system obtains stable results. Once the training is completed, it is time to verify if the system is capable of responding correctly to data never seen before, demonstrating its capacity for generalization.

Exemplifying what just described, imagine the creation of a system that recognizes fruits in photos. During the training, the human shows several pictures of specific fruits to the system, each time describing which image represents which fruit. The system uses the "weights" (numbers) to make these classifications. If the system makes a mistake, it modifies these weights to improve. After numerous images, the system will be able to recognize correctly the fruits, also in new photos. When talking about image recognition, computers use numbers to represent pixels in the images. Each image is composed of a grid of pixels, and each pixel has a numerical value that represents its color. In the fruit case, the system learns to recognize specific schemes in these numbers. For instance, it can be learned that a certain combination of numbers indicates the presence of a specific fruit rather than another.



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Grayscale image | Source

Figure 2. Example of transformation pixel-to-numbers learning system of images.

Researchers have long tried to substitute the hand-designed characteristics with multilevel neural networks that can be trained automatically. Even if the concept is simple, it was not comprehended until the 1980s. These networks can be trained by using a method that is called "*simple stochastic gradient descent*", which allows the system to learn from past mistakes. A very crucial tool is backpropagation, essential for the training of feedforward neural networks since it allows the system to calculate how the modifications on weights influence the final results. In a feedforward network, the input is elaborated through different layers until producing an output. The backpropagation follows the feedforward process: after obtaining the output, it compares the desired result to calculate the error. Subsequently, the backpropagation uses this error to update the weights, improving the performance of the network during the learning process[14].

---

[14] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature, 2015, 521 (7553), pp.436-444. https://hal.science/hal-04206682/document

1.2.1 Convolutional Neural Networks (ConvNet or CNN)

Convolutional neural networks (ConvNet) are designed to elaborate and analyze complex data, such as colored images, which can be seen as numbers grids (pixels). A ConvNet analyses small groups of close pixels (local connections) and applies the same filter on the entire picture to recognize similar schemes in different areas (shared weights). Through what is called "pooling", the network simplifies the image reducing the less important details. Each subsequent layer of the network analyses the image in an increasingly sophisticated way, allowing it to recognize more or less complex objects and details.

Going back to the recognition of an animal in an image: the first layer of the network might individuate small details like borders and angles. These borders, if combined, might be recognized in the second layer as parts and shapes, for instance, the ear of the animal. However, individuating only the ears is not sufficient: other characteristics such as body or tail, must be recognized in the following layers. Only when all these elements are combined, the network can determine if the image contains a specific animal or not. The process, therefore, happens gradually, uniting the pieces together to obtain the entire picture.

Pooling must simplify the data, by grouping similar information. Subsequently, conditional and pooling levels are stacked on top of each other. The backpropagation is used to optimize all weights in the network, exactly similar to how it happens in a deep neural network. Deep neural networks exploit the fact that many natural signals have a hierarchical structure. This means that high-level features can be obtained by combining low-level features. For example, in an image, edges form motifs, motifs combine into parts, and parts form objects. This hierarchical concept is also present in other types of data, such as languages and sound. There structure of the convolutional and pooling layers in ConvNets is inspired by the visual cells in the human brain, particularly simple and complex cells. When an image is shown to a ConvNet model and to, for example, a monkey's brain, the activation of the higher-level units in the ConvNet can explain much of the electrical activity in certain cells of the monkey's brain.

In the 1990s, numerous uses of convolutional networks, starting from time-delay natural networks for vocal recognition and document lecturing, emerged. A system that reads formats used a

ConvNet trained as a probabilistic model to implement linguistic constraints. At the end of the 1990s, this system was capable of reading beyond 10% of checks in the US. Subsequently, optical character recognition systems were developed, and the recognition of handwriting was based on ConvNets.

Deep learning theory shows that, deep networks have two exponential advantages compared to traditional algorithms, which do not use distributed representations. Both advantages derive from the composition strength and depend on the structure of the data. Firstly, this type of learning allows to generalize to a new combination of values apart from the ones learned during training; Secondly, as above mentioned, Deep neural networks are not only used for visual recognition tasks but also for Language Processing. The level composition in a deep neural network means that each level learns to combine information in more and more complex ways.

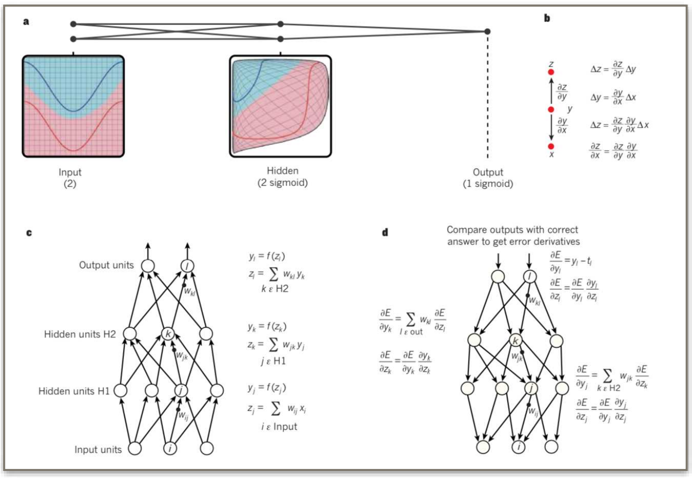


Figure 3. Multilayer neural networks and back propagation[15]

---

[15] A. A multilayer neural network with an input layer (2 sigmoid units), a hidden layer (2 sigmoid units) and an output (1 sigmoid unit)
B. The chain rule to compute gradients
C. Network diagram illustrating the forward pass through input, hidden, and output layers.
D. The back propagation algorithm, showing how to compute gradients layer-by-layer to adjust the weights

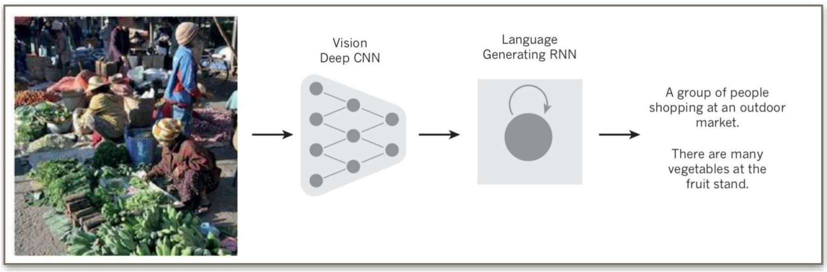

Vision
Deep CNN
Language
Generating RNN
A group of people shopping at an outdoor market.
There are many vegetables at the fruit stand.

Figure 4. From image to text. Captions generated by a recurrent neural network (RNN) taking, as extra input, the representation extracted by a deep convolution neural network (CNN) from a test image, with the RNN trained to 'translate' high-level representations of images into captions (top)[16].

1.2.2 Recurrent Neural (RN) and Long short-term memory networks (LSTM)

When backpropagation was introduced, its most enthusiastic use was the training of Recurrent Neural networks for tasks that involved sequential input, such as language. RRNs elaborate a sequence of inputs, one element at a time, maintaining in their hidden units a "state vector" that contains implicit information on the history of all past elements of the sequence, therefore remembering what has already been elaborated. RNNs are dynamic powerful systems: thanks to their progress in terms of architecture and training methods, RNNs are demonstrated to be efficient in foreseeing the following character of a text or word in a sequence, but they are also used in more complicated tasks. One example is the automatic translation between languages, such as English to French. Imagine that the network reads an English sentence word by word. At each step, it memorizes information about all words that have already been read, maintaining an internal memory that is constantly updated. At the end of the sentence, the final state of the network called the "*thought vector*", reorients the general idea or meaning of the English sentence. This vector is then passed to another network, that has been trained to generate French sentences. The de-codification network starts with proposing the first word of the translation based on the probability: for example, it could suggest the word "bonjour" with a higher probability compared to other words. Once the first word, is used as input to determine the second word of the translation, and so

---

[16] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature, 2015, 521 (7553), pp.5. https://hal.science/hal-04206682/document

on until the completion of the sentence[17]. The process continues until the system decides that the translation is complete, inserting a dot. In this way, the neural network transforms an English sentence into a French word sequence, following a probabilistic journey that depends on the comprehensive meaning of the original sentence.

It is also possible to "translate" the meaning of a picture into an English sentence. The codifier in this case, namely the part that analyzes the image, is a CNN that converts the pixels into an activity vector, which is the numerical representation of the image. Subsequently, another network, the decoder, an RNN, takes the vector and uses it to generate a sentence that describes the picture. Despite being designed to learn long-term, some empirical proofs demonstrated that it is difficult to memorize information for a long time. To correct this problem, it is possible to adopt the network of explicit memory, in particular the Long Short-term Memory (LSTM), which uses special hidden units, whose natural behavior is to remember input for a long period. LSMT is more efficient than RNN, especially when dealing with Vocal recognition[18].

1.2.3 R. Sutton and A. Bardo in "*Reinforcement learning: an introduction*"

Richard Sutton and Andrew Bardo are two pioneers in the field of Reinforcement Learning, a fundamental branch of AI, which focuses on how systems can learn and make decisions through interactions with the environment. Their main work "*Reinforcement Learning: an introduction*"[19] provided a theoretical and practical base to comprehend the key principles of reinforcement learning, such as the maximization of inputs from the environment and the management of exploration compared to the exploitation of such inputs. The importance of these researchers resides in the fact that reinforcement learning offers a powerful framework to solve complex problems, where decisions must be made in a sequence and the effects of actions are not immediately evident. This approach has a significant impact on several fields, from robotics to games, to the optimization

---

[17] Example: if the English sentence is "*how are you*", the network might consider more words as first one in the French translation ("comment", "commentaire, "commentairez"…), but it will give a higher probability to "Comment" because it is the most adequate word in that context. The process goes on until completing each word.

[18] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature, 2015, 521 (7553), pp.6-7. https://hal.science/hal-04206682/document

[19] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press. Pp. 2-10 https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

of complex systems. In a world where AI is becoming more and more pervasive in our daily lives, the discoveries of Sutton and Bardo keep on inspiring researchers and developers and contribute to creating increasingly autonomous and intelligent AI systems.

Reinforcement Learning (RL) is a multidisciplinary field that deals with the interaction between a learning system and its environment. Differently, from the supervised and non-supervised learning, RL requires the system to discover which actions generate the maximum result through direct experience. The main issues of RL include the dilemma between exploration and exploitation. The system has to balance the preferences for the actions that demonstrate being efficient (exploitation) and its necessity to explore new actions to obtain better information about its environment. This interaction requires a closed-cycle approach since the actions influence the future input and the results can depend not only on immediate actions but also on future states, generated by these immediate actions. A distinctive aspect of RL is its focus on goal-orientated systems that operate in uncertain environments. This characteristic allows such systems to adapt and improve over time, facing challenges that go beyond the isolated issues considered in other forms of machine learning. Moreover, the RL has strong interactions with other disciplines, such as statistics, optimization, and neuroscience, contributing to the development of models that accost artificial learning to the biological one. RL methods are inspired by the learning processes of human beings and animals, and in exchange provide more accurate psychological models of the animal learning and system of cerebral reward. To better understand this process, we can take as an example a robot that has to learn how to navigate in a labyrinth to achieve a goal, such as the exit door. The environment in the matter is clearly the labyrinth with walls, open spaces, and the exit door. The system, personified by the robot, has different options for movement, such as going straight, turning left or right, or stopping. According to the reward system, when the robot moves toward the exit door, it receives a motive reward, for example, +10 points, indicating that it is making progress toward the goal; if the robot runs against a wall it receives a negative score, indicating he carried out a useless action. Once the robot gets to the exit door, it could receive a significant reward for completing the mission. After several experiences in the labyrinth, the robot starts learning which action leads to positive rewards and which to negative ones. For example, if it discovers that turning left often leads to the door, it will start choosing that direction more often. During the learning process, the robot must balance exploration and exploitation: if it has already learned that turning left is often advantageous, it could continue doing so (exploitation). However, it could also try new directions (exploration) to discover whether there are more efficient routes.

Other than the environment and the “agent” (robot/system), it is possible to identify three other elements of a RL system: a policy, a reward signal, and a value function. The policy is the strategy used by the agent to decide which action to undertake in a specific state of the environment. In other words, it is a rule that connects the perceived states of the environment to the actions that the agent can undertake. The policy can be represented in different ways and it plays a crucial role in the behavior of the agent. Policies can assume different shapes, based on the complexity of the problem and the needs of the agent. They can be simple functions (the decision to always move toward the right, for example), research tables that lists states and correspondent actions, where each row represents a state of the environments and shows which action the agent should execute. Finally, policies can be complex calculus processes, namely being represented by more complex advanced calculus algorithms. For example, the agent could use a neural network to evaluate the layers and determine the best actions to undertake. A second element is the reward signal: it defines the goal of the agent in the RL. Every time the event undertakes an action, it receives a number that represents the reward. The main goal is to maximize the total sum of awards received over a long period. The rewards are useful to indicate which events are good and bad for the agent, and can be compared to the pleasure and pain experiences in a biological system Where the reward signal indicates what is good and what is not, the value function specifies what is advantageous in the long run. The value of a state represents the total reward quantity that the agent can expect to accumulate in the future starting from that a specific state. These functions provide a wider perspective compared to the immediate rewards, since they take into account future rewards deriving from the following states.

Going back to the robot example, imagining the robot is in an open space inside the labyrinth that has no immediate reward, therefore it might not seem useful. However, if the robot has already explored that part of the labyrinth and has already discovered and learned that that space is often followed by another space that contains a reward, then the value of that space increases. In this way, the value functions help guide the decision of the agent, orienteering its choices and awards actions that lead to more valuable states. Briefly, the RL is based on the interaction between the agent and the environment in a learning context. The main elements, above described, work together to allow the agent to learn to maximize the award over time, facing uncertainties and making strategic decisions. The learning happens through specific learning algorithms such as the Q-Learning and Policy Gradient Learning, that operate iteratively, updating policies and values based on past experiences.

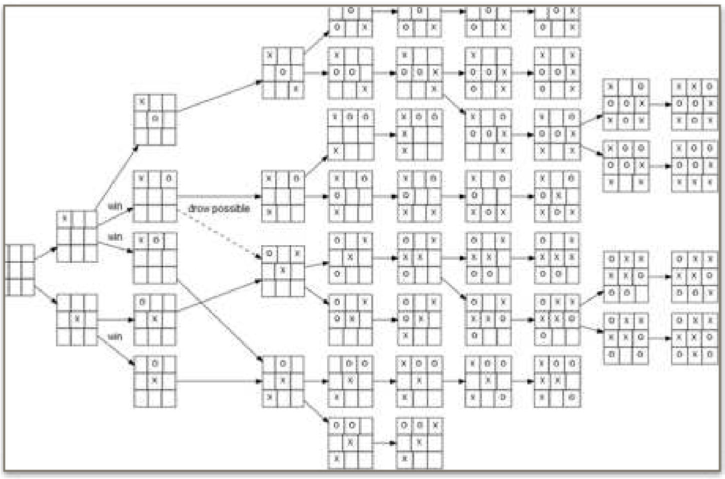


Figure 5. Schematic representation of Reinforcement Learning process while playing Tic-Tac-Toe moves[20].The solid lines represent the moves taken during a game; the dashed lines represent moves that we (our reinforcement learning player) considered but did not make. Our second move was an exploratory move, meaning that it was taken even though another sibling move, the one leading to e∗, was ranked higher. Exploratory moves do not result in any learning, but each of our other moves does, causing backups as suggested by the curved arrows and detailed in the text.

One of the most significant real-life examples is AlphaGo. Between 9th and 15th March 2016, a Go tournament happened between a professional player and AlphaGo, and informatics programs created by Google's DeepMind. The victory of AlphaGo became a significant moment in the history of Artificial Intelligence. It marked the first-ever case of a computer beating a human expert at the Go game. Normally, AI in informatics games uses a game tree (Figure 4) to determine the best move based on what the adversary could do. Nevertheless, for more complicated games like Go, determining the following best move becomes rapidly impossible because the game tree would contain an overwhelming number of nodes for a computer to memorize. This explains why this game has been seen as one of the major challenges for AI for a very long time. The majority of AI for board games use hand-created rules by AI engineers. Since these rules might be incomplete, they usually limit the intelligence of the system.

---

[20] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press. Pp. 2-10 https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

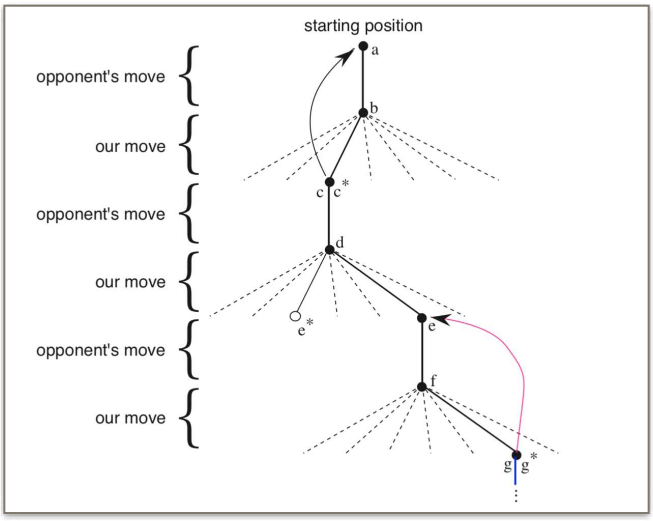


Figure 6. Game Tree for the Tic-Tac-Toe game[21].

All the previous AI methods that played the game, were based on some type of research on the game tree, combined with human-created rules. AlphaGo, on the other hand, widely uses machine learning to avoid using rules elaborated by humans and to improve efficiency. It also uses machine learning to individuate schemes and modify the actions, and it uses Deep Learning and neural networks to teach itself how to play. The AI of AlphaGo is based on its exposition to millions of positions and moves coming from games played by humans. In other words, the intelligence of AlphaGo is based on two different components: a research procedure in the game tree and neural networks that simplify such procedure. Inside AlphaGo neural networks are trained: the policy network and the value one. Both types take the state of the game as input and evaluate every single following move through different formulas, producing the probability of victory. The output is a probability value for each possible move. One of the most important aspects of AlphaGo is its capacity to constantly learn and improve its intelligence by playing numerous games against itself.

In the last years, this method of Reinforcement Learning has been applied to various contexts, from gaming to healthcare diagnosis, to personalized recommendations on streaming services. These are technologies that demonstrate how RL can face complicated problems, adapting and improving constantly over time, making it a precious tool for developing autonomous and intelligent technologies.

---

[21] ibid.,

1.2.4 I. J. Goodfellow and his revolutionary Generative Adversarial Networks (GAN)

Deep learning promises to discover complex and layered models in data such as images, audio with words, or natural language texts. The major success until now has been obtained in discriminatory models, that map the complex and multidimensional input and associate them to a specific category or label, as recognising an image and classifying it as “cat” or “dog”. This success is due especially to backpropagation algorithms, which use linear units, useful for learning data efficiently. Generative models, on the other hand, had less success due to the difficulty of approximating complex probabilistic calculus that are difficult to deal and optimise. Ian Goodfellow proposed a model to estimate generative models that could overcome these difficulties.

In Generative Adversarial Networks (GAN), there are two models of networks that compete with each other: a generative and a discriminative one. The generative model tries to create false data, whereas the discriminative one tries to understand whether the data are false or real. This competition makes it so that the two models increasingly improve, until when the data generated by the generative model seems real. This method is called “adversarial networks” and can be used with different models and algorithms. The GAN framework is simple to use when the generative and discriminative models are both multilayer neural networks. To teach the generative model to create data similar to real ones, the process starts with a random noise distribution coming from another computer: by using random generation functions available in language programs, the system creates random numbers that follow a precise distribution. The numbers of the generated noise are then used as input for the GAN generative model, which transforms those numbers into data that are similar to real ones. On the other hand, the discriminative model receives the data, which can be both real and fake data, and decides whether they are “artificially” generated or not. So the generative one always tries to deceive the discriminative one by producing data that seem more and more realistic. In this way, the discriminative model becomes gradually better in identifying real data. It could be seen as a competition between the two models, where each tries to surpass the other one. In the end, the generative models should be able to completely deceive the discriminative one, by producing data that are identical to real ones.

Generative Adversarial Networks had a significant impact on the development of artificial intelligence thanks to their capacity to generate extremely realistic synthetic data. This innovation opened the way to the creation of images, videos, and sounds improving machine learning techniques. GAN made possible also the increase of datasets for the training of models, facilitating the improvement of performances in complicated tasks, such as the recognition of images and the generation of content. Furthermore, the competitive approach between generative and discriminative models has inspired new methodologies of ML, contributing to a deeper comprehension of generative processes and enhancing the efficiency of AI algorithms[22].

1.2.5 "*Attention is All You Need* ", Ashish Vaswani, 2017

In 2017, the paper "*Attention is All You Need*"[23] by Ashish Vaswani and his team of researchers at Google, introduced the architecture of the Transformer, marking a fundamental turning point in the field of Natural Language Processing (NLP). This innovative approach eliminated the necessity of traditional sequential models, such as RNN and LSTM, introducing a mechanism of attention that allows the model to weigh dynamically the importance of the different words in a specific context. Thanks to its capacity to elaborate information in parallel, the transformer has accelerated the training process and improved its performance in language tasks. The architecture not only made possible greater efficiency and accuracy in the automatic translation and generation of texts, but also laid down the foundations for advanced models such as BERT and GPT. This has transformed the overview of Artificial Intelligence, making the transformer a central element in the development of modern linguistic technologies.

RNNs have been recognized as one of the best methods to face problems in sequences and transduction[24], such as the comprehension of language and automatic translation. Over the last few years, several researchers have continued to improve the models and the encoder-decoder structures. RNNs generally try to understand the context of a sequence by just looking at the past information and at the elements that are currently being analysed. However, this sequential structure makes it difficult to execute the calculus in parallel during the training, especially with longer

---

[22] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27. Retrieved from https://arxiv.org/pdf/1406.2661

[23] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems. https://arxiv.org/pdf/1706.03762

[24] The term "*Transduction*" in informatics and Artificial Intelligence, refers to the process of converting one type of data into another type. In specifics, in the context of NLP and machine learning models, the transduction implicates the transformation of input sequences into output sequences.

sequences, since there are limits to the memory that impede the contemporary elaboration or multiple sequences. Some recent studies have found ways to make this calculus more efficient, but the problem of sequentiality persists.

Attention mechanisms have become essential for sequence models and transduction tasks. These mechanisms allow to shape the relations among elements, independently of their distance in the sequences. In their study, Google researchers presented an architecture that avoids using recurrence and is based solely on the attention mechanism to establish relations between input and output. The transformer allows to execute calculus in parallel and can achieve high-quality results in the transaction in only a few hours of training.

The majority of translation models use the encoder-decoder structure. In this system, the encoder takes a sequence of symbols, such as words in a sentence, and transforms it into a shape (or vector), that contains all the important information about the original sequence. After transforming this sequence in the shape, the decor will use it to generate the desired output, such as the translation to another language. The decoder starts generating a new symbol sequence, word by word, based on what what has already eight transformed by the encoder and based on words already produced by the decoder. The encoder is composed of different layers that carry our two tasks: the first uses the attention method to understand which parts of the sequence are important, whereas the second is a network that further elaborates this information. The attention mechanism is crucial because it allows the model to focus on different parts of the sequence while creating the output sequence. In this way, the model can decide which information is more relevant to use in the translation. The transformer uses multi-head attention, meaning that it can look at different pieces of information at the same time, improving the quality of the translation. To convert the symbols into vectors (lists of numbers representing the words/symbol in a numerical space), the embedding techniques are used, and then probability methods to calculate the probability of the following symbol. Since the model does not use traditional methods such as recurrence, it is necessary to add information on the position of the symbols, to better understand the order of the words.

Let's make an example: imaging the translation of an Italian sentence "*Il gatto è sulla sedia*" into the English version "*The cat is on the chair*". The encoder receives the Italian sentence and converts it into a numerical representation. Each word is transformed into a vector (lists of number) through the embedding process. For example:

- “Il” → (0.1, 0.2, 0.3)
- “gatto” → (0.4, 0.5, 0.6)
- “è” → (0.7, 0.8, 0.9)
- “sulla” → (1.0, 1.1, 1.2)
- “sedia” → (1.3, 1.4, 1.5)

The same thing is done for the English words:

- “The” → [2.1, 2.2, 2.3]
- “cat” → [2.4, 2.5, 2.6]
- “is” → [2.7, 2.8, 2.9]
- “on” → [3.0, 3.1, 3.2]
- “the” → [3.3, 3.4, 3.5]
- “chair” → [3.6, 3.7, 3.8]

The attention is applied to determine which words in the sentence are important to understand the compressive meaning. For example, the attention could fall on “gatto” and “sedia”. The encoder produces a series of representation that summarises the information contained in the original sentence. The decoder starts generating the translation, using the information of the encoder and the generated words, to decide which word will come next.

The semantic space is a concept that represents the relation between words based on their meanings. In this space, each word is represented by a vector. Words with similar meanings are positioned closely. During the training of the system, the association of words is learned by analyzing many sentences of different contexts: this allows the model to understand the relation between words, facilitating the translation and other activities, since it recognizes and uses words that share similar meanings or contexts.

Going back to the example, during the training, the model is tested with many Italian and English sentences. For example, it might see the Italian sentence “*il gatto è sulla sedia*” and its English correspondent. The model learns that “*Gatto*” corresponds to “cat” and so on because the associated vectors have similar positions in the semantic space. After generating and associating all the words/vectors, the final result will be “*The cat is on the chair*”.

In this work, the Transformer represents a revolutionary model in the field of Transduction, capable of being trained significantly faster than the architectures based on RNNs or CNNs. In some sectors, this approach achieved a new state-of-the-art demonstrating the efficiency of attention mechanisms in improving performances. With further developments and applications, it has the potential of ulteriorly transforming the scenario of AI and NLP.

1.3 Applications of AI across different sectors[25]

Artificial intelligence is a technology that is transforming profoundly a wide range of sectors, thanks to its extraordinary capacity to analyze large quantities of data, identify complex patterns, and make autonomous or semi-autonomous decisions. This technology, based on advanced algorithms and machine learning techniques, is capable of elaborating information with a speed and precision impossible for a human being, contributing to the improvement of efficiency and effectiveness of processes in numerous fields. What make AI particularly powerful is its capacity to adapt and improve over time. By using enormous amounts of data and refining constantly its prevision and decisions, AI can face complex problems, automatize repetitive activities, and provide new innovative solutions. This led to a rapid spread of the technology in a variety of sectors, with tangible results that range from the optimization of companies’ operations to the acceleration of scientific discoveries. Its versatility allows AI to be used not only to better the operational and technical performances but also to transform entire industries. Over the years, AI demonstrated to have a transformative impact on key sectors such as finance, healthcare, transport, and security. These fields are at the base of modern societies’ functioning, and are assisting to a real technological evolution, where the AI is assuming a central role. In addition to improving efficiency and reducing costs, AI is also stimulating innovation, allowing the development of new solutions and approaches that were once unthinkable. The integration of AI not only optimizes existing processes, but opens up the way to new opportunities of growth, both in economic and social terms.

---

[25] JavaTpoint. (n.d.). *Application of AI*. JavaTpoint. Retrieved from https://www.javatpoint.com/application-of-ai

Nevertheless, together with benefits, the introduction of AI systems raises also ethical, social questions, and technical problems in their application, which will be further discussed in the following parts of this thesis. The technological progress must be accompanied by a careful reflection on how to best manage the changes that AI leads to, ensuring that the advantages are equally distributed and the potential risks are mitigated. Finally, AI is one of the most promising and disruptive technologies of our times, capable of defining the future of several sectors. Following are the most relevant applications of this technology in sectors such as finance, healthcare, transport, energy, and security, fields that are experiencing a transformation with no precedents thanks to the potential of AI technologies.

1.3.1 AI in Finance

Over the last 10 years, neural networks have been increasingly applied to diverse areas of finance. The main characteristics of these areas include the intensity of data, the unstructured nature, the high degree of uncertainty, and the presence of hidden relations. The majority of cases of AI application use a back propagation model with hidden layers. In many cases, the neural networks have overpassed traditional statistics models, such as discriminative analysis and regression. Furthermore, these applications have obtained significant success in the financial practice, such as the provision of government bonds, patrimony management, the selection of portfolios, and fraud detection. Artificial Neural Networks (ANN) are computational intelligence systems that simulate inductive power and the human brain's behaviors. These systems have the capacity of generalizing and see through noise and distortion, abstract essential characteristics in the presence of irrelevant data, and to "*provide a high degree of robustness and fault tolerance*"[26].

There exist numerous applications of neural networks in the business world, in particular in the financial sector. For instance, ANNs have been used to foresee the failure risk, by employing financial reports as input data. Odon and Sharda (1990)[27] have confirmed, by comparing these networks to traditional methods, that the ANNs can foresee failure with greater precision. These

---

[26] Lippmann, R. (1987). An introduction to computing with neural nets. IEEE ASSP Magazine, 4(2), https://doi.org/10.1109/MASSP.1987.1165576

[27] Odom, M. and Sharda, R. 1990, "A neural network model for bankruptcy prediction," Proceedings of the IEEE International Conference on Neural Networks, San Diego, California.

networks have been also applied to the evaluation of bond security, in the analysis of risk for mortgages and loans, in the forecast of the stock exchange, and financial analyses.

In the financial sector, Artificial Intelligence is revolutionizing both operations internal to finance institutions and users' experiences. Thanks to sophisticated algorithms of machine learning and Natural Language Processing (NLP), Artificial intelligence can deal with a wide range of activities with precision and speed that overcome human abilities. AI systems are used to analyze big volumes of transactions and detect suspected behavioral schemes in real-time. Advanced systems of machine learning can learn and identify fraudulent activities and potential risks, improving the security of transactions and reducing the time necessary to detect fraud. The AI algorithms allow financial institutions to provide personalized recommendations on products and services, investment offers, or loans. These systems analyze the financial profile of the user, and their preferences of risk and goals, providing suggestions that optimize the return on investments or improve the management of savings. Banks and investment enterprises use AI for the elaboration of complex documents, such as loan contacts or mortgage requests. The AI can extract relevant information from both structured and unstructured documents, allowing a more efficient and quick management of data. This reduces the manual loading and increases operational speed. In the investment field, algorithmic trading based on AI permits one to carry out transactions in milliseconds, analyzing contemporary market and macroeconomic data. This approach guarantees more informed and accurate investment decisions[28].

1.3.2 AI in Healthcare

Over the last few years, AI has started transforming profoundly the healthcare sector. Its potential impact is wide, but at the same time raises concerns regarding its evolution and its efficient integration in healthcare assistance. Currently, AI is confined to limited and well-defined tasks, but in the following years, it could expand significantly, improving the diagnosis, cure, and management of resources in the medical field.

---

[28] Fadlalla, A., & Lin, C.-H. (2001). An Analysis of the Applications of Neural Networks in Finance. Interfaces, 31(4). http://www.jstor.org/stable/25062724

AI applications have rapidly developed thanks to the improvement of computing-powered analysis of large quantities of data. Nonetheless, the use of AI in the cure of patients still presents some significant limitations. One of the major challenges is linked to the availability and computing power necessary to support AI in real-time: without these resources, the decisional process might be slowed down, compromising the timeliness of cures. At the moment, AI remains confined to what is defined as "Restricted AI", namely technologies capable of overpassing human capabilities only in limited tasks. The idea of a "general AI", capable of equaling or overpassing human cognitive abilities in multiple sectors, is still far away, and according to some people, highly speculative. In any case, AI is gradually entering global healthcare systems, thanks to strong investments by public, private, and philanthropic entities. A relevant example is the initiative for the public healthcare of Precision, launched by the Rockefeller Foundation together with organisations such as the OMS and UNICEF, with an initial financing of 100 million dollars. This initiative aims to bring AI technologies to low and middle-income countries, expanding the benefits of innovation also to those areas that, until now, have not had access.

Artificial intelligence has already demonstrated efficacy in activities such as the pattern recognition and analysis of medical records and images. Despite this, applications based on AI are not widely used in daily clinical practice. Currently, AI is going through an experimental phase for the diagnosis in oncology through radiological applications, such as thoracic, cerebral, and mammography imaging, or other areas such as dermatology and pathology. However, only few of these systems are subject to large-scale clinical trials. AI could also be used to evaluate the risk of chronic diseases, like cardiac ones or diabetes, providing more efficient prevention instruments. In the future, AI might accelerate medical diagnosis, allowing an early detection of conditions like ictus or breast cancer. Some exemplar studies indicate that AI could be used also to diagnose neurodegenerative diseases, such as Alzheimer's, years before the symptoms become evident. Despite the AI-assisted diagnosis seem promising in the short term, there are some concerns in the long term, especially regarding the precision of previsions. Mistakes in the precision might have serious negative consequences, like the limitation of the access to cure or the generation of useless warnings. There exists also a relevant ethical matter: the use of AI to predict disease, raises doubts regarding privacy and informed consent, especially if people are not aware of their data being used for such purposes (to explore the ethical debate, see the final chapter of the thesis). Additionally, the accuracy of AI depends on high-quality data, correctly noted and coming from diversified sources.

A significant risk is that, if data do not represent all populations or are low quality, AI might produce wrong or distorted diagnoses.

The use of AI in the medical field was first theorised more than 60 years ago. The goal was to create systems capable of helping junior doctors follow accurate diagnoses, especially in contexts where access to expert doctors was limited. Today, AI is employed to identify risk factors, suggest optimal treatments, and evaluate the costs and efficiency of cures. Nevertheless, the modern vision on AI is evolving towards a more human-centered approach. The goal is not only to support doctors anymore, but also to create human-machine teams that can work together to offer better solutions. AI technologies are becoming more and more patient-oriented, intending to favour shared decisions among doctors, patients, and the community. At present, experience and clinical knowledge continue to be indispensable: AI cannot substitute human judgment yet, but it can reduce the cognitive loads on doctors and help them make more informed decisions. AI can also be useful to integrate medical records and identify vulnerable patients, offering decisional support in complex situations.

Until now, public healthcare surveillance was based on the collection of proofs and their use to create mathematical models to make decisions. Technology has revolutionized surveillance thanks to the addition of digital traces, namely data that are not generated specifically for public healthcare reasons (blogs, videos, reports, or internet research) that can be applied to public healthcare. For instance, videos are rich sources of information that can be transformed into healthcare intuitions. However, public healthcare institutions have not fully exploited these sources yet. Google Flu Trends, for example, is based on queries of the search engine regarding complications, remedies, symptoms, and antiviral medicines for the flu, used to estimate and foresee flu activities. Data are useful only if the appropriate models are used. In the same way, machine learning and algorithms can result more precious and efficient when integrated with digital traces and human activity. In this context, surveillance is changing. For instance, researchers have been able to detect an increase in severe lung disease cases associated with the use of electronic cigarettes by analyzing disparate online information sources and using Health Map, an online data mining tool. Again, some Microsoft researchers have found early proof of adverse reactions to drugs through web registers. In 2013, the researchers detected collateral effects in different prescription medicines before being discovered by the American warning system of the Food and Drug Administration (FDA). OMS is developing EPI-BRAIN, a global platform that will allow data and public healthcare experts to analyze datasets for the preparation and response to emergencies. EPI-BRAIN aims to exploit Big Data and AI to mitigate the impact of epidemics. It will permit previsions and early detection of infective threats and their impacts through scenarios, simulation exercises, and intuition sharing to improve decision-response coordination.

Finally, the discovery and development of drugs are long processes: the development of a new drug requires, on average, a decade. AI might change drug discovery, transforming a laborious process into a more intensive one, in terms of both capital and data, thanks to the use of robotics and genetic, pharmaceutical, and organic models of disease and their progression. AI could be deployed in the discovery of drugs and their development to reduce waiting times. As of now, there is a collaboration between machines and humans, guided by humans or by AI with human supervision. In the following decades, along with the optimization of machine work, AI could further develop. The computing might allow the discovery and development of drugs by searching for new candidates and evaluating if they satisfy the criteria for such medicines, structuring unorganised data to identify structures and patterns, and re-creating the human body and its organs on AI chips. By 2040, some trials could be viral, without animals or humans, based on human body models. For this to happen, it will be necessary to face challenges of sharing and interoperability of data, in addition to ethical dilemmas that influence the use of AI in the development of drugs[29].

1.3.3 AI in Transports

Urban mobility has recently experienced a notable change together with the traditional transportation systems. Despite the urban areas occupy only 3% of the terrestrial surface, they consume 75% of natural resources and produce between 60% and 80% of global greenhouse gas emissions. The urban population will grow to 70% by 2050, compared to 50% of the current global urban population. The rapid urbanization of countries will have a dramatic impact on the environment, security, and management of urbanized cities. Many countries, like the US, EU, and Japan, have proposed the concept of "smart cities" to reduce the consumption of energy and

---

[29] World Health Organization. (2021). Artificial intelligence is changing the health sector. In WHO Consultation Towards the Development of guidance on ethics and governance of artificial intelligence for health: Meeting report Geneva, Switzerland, 2–4 October 2019 (pp. 3–7). World Health Organization. http://www.jstor.org/stable/resrep35680.7

optimize natural resources. All over the world, smart city programs have been developed and adopted to face future challenges efficiently. Innovative approaches are necessary to improve the quality of production and sustainability, other than reducing costs for intelligent production systems.

The congestion, accidents, and pollution problems caused by transport are becoming increasingly serious due to the enormous increase in travel needs, which include vehicular traffic, public transport, goods, and pedestrian traffic. To face and deal with such issues, Intelligent Transport Systems (ITS) have been developed, capable of integrating a wide range of systems, including detection, communication, information diffusion, and traffic control. Three essential components are necessary for an ITS to function: data collection, data analysis, and transmission of information. The data collection acquires observable information from the transport system for a deeper analysis of the current traffic conditions. Traditionally, inductive loop systems were used, that detected the presence of vehicles based on the current induced in the loop by vehicles in transit, and pneumatics tubes that detected the presence of vehicles based on the variations of pressure on the tube, to gather information on the traffic, such as the volume and the speed. However, due to the high implementation cost and the impact on the traffic during the installation, these methods are becoming less popular. Thanks to the progress in detection and imaging technologies, cameras and radio frequency identification scanners (RFID) are being considered more for the collection of traffic data. Cameras can be installed in different positions, videos can be then analyzed using image elaboration software expressly designed for these tasks (such as Autoscope), to determine information such as traffic flows, speed, vehicle typologies, etc… In this context, the automatic recognition of plates is a crucial research area, since through this stem it is possible to provide additional information, such as the selected routes and travel times.

ITS can be classified into two categories based on their functionalities: Advanced traveler information systems and advanced management systems. The first type aims to help travellers make decisions (for example the way, route, time) by providing different types of information. Among the implementations, the esteem/prevision of travel time and guiding systems are the most commonly studied areas, since they can influence directly the travellers' choices. With the advancement of data collection methods and communication technologies, information on travel times and routes can be provided more accurately and in real time. For instance, the image analysis of traffic conditions automatically provided by drivers through smartphone apps, can be used to determine in real-time the availability of parking slots along the streets. The second typology, the Advance Management System, aims to check or manage different infrastructures and operators inside the transport system in different situations, to guarantee the efficiency and safety of the system itself.

These ITS systems are included in the category of smart mobility in the wider frame of smart cities. In the literature, there does not exist a consensus on what constitutes a smart city, and there are different definitions. For example, some suggested that a smart city should monitor its components (streets, buildings, etc..) to optimize its resources, plan preemptive maintenance activities, and monitor security, maximizing the services for its citizens. Whereas others proved that smart cities are those that use information and communication technologies (ICT) to value human, relational, and social capital and environmental matters. Definitions depend on the stakeholders' backgrounds and the orientation of the governments. Despite this differentiation, the use of advanced digital and electronic technologies, the integration of ICT in urban infrastructures, and the improvement of stakeholders' interests, are the three common dimensions in smart cities.

Neural artificial networks (ANNs), with their capacity to execute non-linear mapping of entrances and exits through the consideration of hidden layers and adequate training, are proper systems to face transport issues where the relations among variables are not understood. In the literature, ANNs are commonly adopted in the esteem/ forecasting of states, detection of accidents, control of traffic/ infrastructures, and the analysis of behaviors. Similar to ANNs, support vector machines (SVM) are supervised learning models that analyze input data, but are more focused on the classification of phases/scenarios. Consequently, despite the SV has been applied to other issues other than transports, is mainly used for problems such as the detection of accidents. AI in transport systems implicates the application of machine learning, artificial vision, language processing, and methodologies to improve efficiency, safety, and sustainability. Other applications of AI are autonomous vehicles, where the AI uses sensors, artificial vision, and machine learning algorithms to navigate, perceive the environment, and make decisions without man intervention; predictive maintenance, by monitoring and analyzing data coming from vehicles and infrastructures to foresee the necessity of maintenance. This allows proactive maintenance and reduces unpredicted damages. Another application is the intelligent logistics and management of the supply chain, where the AI can optimize the logistic operations by analyzing large datasets, routes, predicting questions, and managing the inventory more efficiently[30].

1.3.4 AI in the Energy Sector

Energy is a fundamental element for the economic and social development of any country, and the development of energetic sustainment plays a crucial role in guaranteeing the availability of energy sources. Energy security is strictly connected to the efficiency and diversification of sources, so as the capacity to manage energy sustainably and competitively. In this context, the introduction of advanced technologies such as intelligent networks, Smart Grids and AI, is revolutionizing the energy sector. Smart Grids allow more efficient management of energetic resources, contributing in a significant way to the reduction of gas emissions and the increase of energy efficacy. These smart networks are capable of integrating reliable sources, optimizing consumption, and making the energy system more resilient, when facing challenges advanced by climate change and the growing global demand for energy. Artificial intelligence, in particular, represents an extraordinary opportunity for this sector: thanks to it, it is possible to collect, elaborate, and analyze large quantities of data in real-time, improving the forecasting of demand, the management of resources, and predictive maintenance of infrastructures. Systems based on AI can make autonomous decisions, granting a more efficient and safe management of energy sustainment, and reducing costs. Machine learning technologies allow systems to learn data and improve over time. This means that the energy models can become gradually more accurate and effective, anticipating the necessity of the system and a better allocation of resources. AI is also capable of contributing to the administration of unpredicted events, such as damages or emergencies, thanks to its capacity to react promptly and make informed decisions on difficult scenarios. Other than the planning, it is paramount to invest in the sustainability of energy transmission and distribution networks to guarantee reliable and competitive supply. Smart Grids and High Voltage Transmission infrastructures can reduce the losses of transmission and increase operational efficiency, enhancing the system's resilience.

In all sectors, in particular in the energy sector, everything is becoming client-oriented, and behind these developments are AI and Machine learning. Data have become as valuable as petrol. AI

---

[30] Bharadiya, J. (2023). Artificial intelligence in transportation systems: A critical review. ResearchGate.

algorithms applied to energy data improve reliability and create an intelligent energy sector. The learning techniques include statistics learning (ex. Bayes, Clustering), neural learning (the most used), and evolutive learning (ex. Genetic Algorithm). Hybrid methods are also used, by combing AI techniques to obtain better performances. Energy companies have to manage huge amounts of data, with issues linked to energy costs, production, and distribution. AI can deal with this data at a lower cost, by offering perspectives that could revolutionize the sector.

With the global population growth, the energy demand is on a constant rise. AI offers analytics and predictive capacities able to deal with extremely complex problems. Companies have to foresee demand fluctuations, overloads of systems, and potential damages with the greatest precision. AI also allows a better allocation of energy resources anticipating requests and preventing the emergence of problems, with energy savings for the consumers and personalized services. AI systems are widely used in solar and wind turbine centrals, to estimate power and forecast production. In Smart Grids, AI helps monitor variable parameters over time, optimizing the administration of nonlinear energy systems. With the technological advancement, the energy infrastructures face both risks and opportunities. It is crucial to monitor closely these innovations and integrate them into the long-term energetic strategies[31].

1.3.5 AI in Security

Finally, Artificial Intelligence is profoundly changing the security sector, expanding well beyond the traditional contexts, to include global, geopolitical, and social threats. Thanks to its automated analysis, prevision, and response, it offers advanced tools to face complex challenges in a more and more interconnected and digitalized world. Following, the thesis proposes an overview of the main security sectors, both internal and international, where and how Artificial Intelligence can be applied, which will be further explained in the second part of this thesis, especially when applied to counterterrorism strategies.

Firstly, cybersecurity is one of the areas where AI can have a more direct impact. Machine learning algorithms can constantly monitor informatics networks to detect anomalous activities, such as

---

[31] TÜR, M. R. (2022). Energy Supply Security and Artificial Intelligence Applications. Insight Turkey, 24(3), 213–234. https://www.jstor.org/stable/48733385

attempts of intrusion or malware, and respond in a proactive way to the attacks. The AI can also be used to reinforce the authentication systems and prevent identity theft or unauthorized access to sensitive data. Moreover, the automated analysis of vulnerabilities in the software and operational systems helps identify flaws before being exploited by hackers or cybercriminals, contributing to the creation of a more robust defense against digital threats. In the context of international security, Artificial Intelligence can be used to monitor geopolitical dynamics among states. Through the analysis of data gathered from open sources, such as news, political speeches, and social media, AI systems can pinpoint signals of possible international tensions, economic crises, or political instability. The predictive analysis can be applied to prevent conflicts, evaluate the efficiency of economic sanctions, and foresee governments' or non-state actors' reactions to global events. Since 2010s, along the development of new warfare scenarios, AI has become a key component also in military defense and cybernetic wars. The informatics attacks represent one of the most modern threats, with states and private actors that conduct offensive operations through the networks. AI can detect such attacks in real time and respond with appropriate and immediate countermeasures, like the access block or deviation of malware. Additionally, it can be used to simulate cybernetic war scenarios and prepare defenses against advanced threats. On the battlefield, it can be deployed to control autonomous drones, strategic planning, and the logistic management of military resources.

Artificial intelligence and predictive analysis can also be used jointly to prevent and reduce crimes in urban areas. AI systems can analyze historical data on crimes, human behavioral models, and other contextual factors to anticipate where and when new criminal acts could be perpetrated, allowing the armed forces to intervene preemptively and optimizing the allocation of resources. Smart cities use AI to manage video surveillance, traffic, and other critical infrastructures, such as power plants, water networks, and transport systems, which are targets for terroristic or informatic attacks. Some processes can be applied to climate change and natural disasters as well, which nowadays are considered a matter of international security. Finally, two sectors that are separate but often analyzed together are immigration and terrorism. In the context of the fight against terrorism, AI offers the capacity to analyze large quantities of data collected from different sources, including social media, encrypted communications, and financial flows. Through advanced algorithms, it is possible to identify schemes and suspected behaviors that could suggest terroristic activities. In parallel, AI plays a fundamental role in the management of immigration and border security.

Advanced technologies can speed up identification processes for individuals at the borders. Moreover, the predictive analysis can be used to monitor migration flows and anticipate crises, such as the afflux of refugees or the emergence of high-risk migrant groups. These tools not only improve national security but enhance a better use of resources, and allow authorities to react quickly and efficiently to emerging threats[32].

1.4 The impact of Artificial Intelligence on power relations and strategic stability

The term "arms race" was first used by Less Fry Richardson to describe the development of armaments in the period that preceded the First and Second World Wars and has been often used to describe the developments during the Cold War. This term seems to be re-emerging among journalists and scholars when discussing the competition among powers in the field of Artificial Intelligence. Making a comparison with the nuclear arms race, however, would be slightly imprecise: historically, the concept had a specific meaning, that could lead to misunderstanding regarding what is happening today with AI. During the Cold War, the arms race regarded the increase in numbers or physical capacities of armaments, to achieve or maintain supremacy towards the adversaries. AI, however, is not a weapon, it is not even a defined and unique technology. It is more proper to compare AI to electricity for instance. Indeed, progress in the AI field is mainly guided by civil needs. Leaders in this competition are not states nor public agencies and military services, but rather private civil enterprises, which are not under the direct or indirect control of the state, except for China. Another significant difference is that the competition among powers does not, for now, regard the acquisition of a large amount of advanced arms based on AI. Rather, it is more about the acquisition of the necessary resources to develop or maintain the advanced capacities of AI in all sectors. In this sense, it might be more appropriate to talk about the "capabilities race" to describe the competition among China, Russia United States, and other countries active in the sector.

Is there a "capability race" specifically for the AI? If so, what is the evidence regarding its nature and current state? These questions and the subsequent answers are fundamental concepts because they influence the way decision-makers consider national security needs. Affirming that AI is at the

---

[32] H. Jain, A. Vikram, Mohana, A. Kashyap and A. Jain, "*Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications*," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)

core of an arms race contributes to the elevation of its importance in the agenda of all national security experts all over the world (we are talking about a securitization process). This pushes politicians and armed forces to make decisions and undertake measures that might further militarise AI and fragmentise the AI field along national lines[33].

Several advanced military powers, including Australia, Israel, South Korea, Russia, and the USA have constantly blocked any progress towards a new treaty or declaration to avoid completely autonomous weapon systems in the various meetings held since 2014 regarding the Convention on Conventional Weapons of 1980. This highlights the beginning of an arms race in the field of AI technologies, which will mark future global power relations. For instance, in a study on autonomy, the defense science board of the US concluded that the defence department "*must accelerate its exploitation of autonomy—both to realize the potential military value and to remain ahead of adversaries who also will exploit its operational benefits*"[34]. In June 2017, China fixed the goal of becoming a global leader in the AI field by 2030, challenging the USA's supremacy. Whereas Putin's position was made clear at the meeting with Russian Children on Knowledge day "*Whoever becomes leader in this sphere [AI] will become the ruler of the world*"[35]. Therefore, the development of arms that rely on AI and autonomy will be a key characteristic of a potential arm race during the XX century.

For now, the development of autonomous nuclear weapons is not planned by any nuclear power, however, the use of artificial intelligence in command, control, and surveillance nuclear systems cannot be excluded. Different from the nuclear arms race, the AI race will most likely involve many more actors and will not be limited to states. Due to the scalability, efficiency, and ease in diffusion of systems based on AI, costs and resources and, last but not least, the psychological distance to perpetrate attacks, will be lower, potentially increasing the number of actors and attacks. This dynamic requires new approaches to arms control, with mechanisms that face both horizontal

---

[33] Boulanin, V. (Ed.). (2019). *The impact of artificial intelligence on strategic stability and nuclear risk: Volume I, Euro-Atlantic perspectives*. Stockholm International Peace Research Institute. https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf

[34] Defense Science Board, *Summer Study on Autonomy*, June 2016. office of the secretary of defense 3140 defense pentagon washington, dc 20301–3140, https://apps.dtic.mil/sti/pdfs/AD1017790.pdf

[35] "*Whoever leads in AI will rule the world*": Putin to Russian children on Knowledge Day', RT,1 Sep. 2017; and Maggio, E., 'Putin believes that whatever country has the best AI will be the rule of the world',Business Insider, 4 Sep. 2017.

proliferation (versus minor powers) and vertical one (versus non state actors). Also, the nature of the war will most likely change: since the democratization of the access to armed technologies, states and non state actors might be tempted to use surrogates to fight on their behalf. The application of AI in deterrence and nuclear strategies has the potential to reduce strategic stability. New technologies based on AI will introduce new offensive threats, since AI systems will be capable of completing tasks more efficiently, potentially exploiting the vulnerabilities of other AI systems. What is even more worrying is that the perception of these capabilities on the adversary's side might be destabilizing itself[36].

1.5 Regulation of Artificial Intelligence in the world

Over the last few years, the regulation of AI has become a top priority for international organizations and governments all over the world. With AI rapidly transforming the sectors of the economy, healthcare, security, and more in general daily life, it is evident the necessity to establish clear normative frameworks that can administer the ethical, legal, and social implications of this technology. To face these issues, international organisations such as G7, ONU or OCSE are trying to define regulations for AI, even though they found themselves in a race against the technological evolution. For instance, in November 2023, United Kingdom has hosted the first AI Safety Summit, with the goal of promoting the safe development of AI at the global level. This summit tried to establish an international consensus, but the different emerging regulations in the various jurisdiction create a fragmented environment for the enterprises. The approach to AI regulation varies notably among the different regions, reflecting the difference in political priorities, economic structures and social values. UK chose a more flexible approach, ruling on the existing regulators the application of general principles. In the United States, the normative approach towards artificial intelligence has been, until now, less focused and more fragmented compared to other regions, such as the European Union. There is not a unified federal legislation on AI yet. Contrary, the American government has preferred relying on a sectorial approach, with specific regulations for each sector and already existing norms. The National Institute of Standards and Technology (NIST), for example, published in 2020 a framework for the management of the risks associated with AI. This normative framework has no force of law, but it provides guidelines to help enterprises and

---

[36] Boulanin, V. (Ed.). (2019). *The impact of artificial intelligence on strategic stability and nuclear risk: Volume I, Euro-Atlantic perspectives*. Stockholm International Peace Research Institute. https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf

organizations develop AI systems responsibly. At the same time, the Congress started discussing the potential legislation to face more complex challenges linked to technologies, such as algorithm bias, privacy, and data security, with a particular interest in the protection of competition in the technology market. The USA maintains a “laissez-faire” approach, leaving technological innovation to develop with limited normative supervision. Other countries such as are still developing their won approaches. On the other hand, the Organisation for Cooperation and Economic Development (OCSE) promised a collaborative and multilateral approach to the regulation of AI. The guidelines of OCSE, adopted in 2019, aim to create a common framework among its members to guarantee the transparency of AI systems, the accountability of operators, the robustness of models, and the absence of prejudices in data. Even if these guidelines are not binding, they represent an attempt to create convergence among countries, promoting international cooperation on the administration of challenges advanced by AI. In Asia, Japan adopted a very proactive position regarding the regulation of AI. The government published its national strategy on AI in 2019, to promote the development of AI systems that must be safe, reliable, and human-centred. Japan’s strategies emphasize respect for human rights and transparency in automated decision-making processes. In collaboration with the private sector, it has tried to create an ecosystem where the AI can prosper without compromising the public’s trust. On the other hand, China has adopted a more aggressive and centralized strategy, making AI one of the key instruments in its plans for technological and economic development. China invested enormously in AI development, with the goal, as mentioned before, to become a global leader. However, International concerns are merging regarding China’s use of AI with the purpose of mass surveillance and social controls, raising questions about the implications for human rights and civil freedoms. The European Union distinguishes itself for a particularly structured approach to the regulation of AI. Its focus, on the balance between innovation and the protection of fundamental rights, poses the EU as a leader in the creation of a normative framework that can be taken as an example globally. The main provision will be discussed in the second chapter of this thesis.

This heterogeneity in the normative approaches at the global level creates significant challenges for technological enterprises that operate on an international scale. These find themselves in a position where they need to respect a variety of local and international regulations, that can increase conformity costs and operational complexity. At the same time, this fragmentation could incentivize a greater harmonization effort among jurisdictions, especially among advanced economies, to avoid normative unbalances that might limit global competitiveness or threaten the security of data[37].

The majority of jurisdictions have tried to look for a balance between the encouragement to innovation and investments, and the adoption of norms that protect form potential damages. However, there are some emerging trends that create impediments to this harmonisation process. Firstly, "AI" has different menacing in various jurisdictions: one of the fundamental challenges that every international company must face when developing a strategy of normative conformity on AI, is understanding what exactly constitutes AI. Unfortunately the AI definitions vary. The EU AI ACT, for instance, adopts a definition of AI systems based on the one of OCSE, but with some substantial differences due to a non clear formulation. In USA, various states have proposed their own definitions that differ one form the other. moreover, some countries (for example UK, Israel, China, etc…) have not provided a complete definition. Since different normative on AI have extraterritorial effects (meaning more than on regulation on AI can apply contemporarily), international enterprises high be forces to adopt an approach based on the most rigorous standard among the applicable ones. Additionally, emerging norma on AI assume different forms: somme are laws, some are executive decrees or expansions of already existing normative frameworks. This is aggravated by the different conceptual approaches charactering the normative scenarios: some normative are binding, some others are not. With the goal of creating regulations on AI that can adapt to future technological advancements, many jurisdictions have tried to include a notable flexibility. This happens both through the use of generic formulations and principles, and leaving space to further interpretations and applications by tribunals and regulators. Despite having the adaptation advantage of these formulations, it also bring the disadvantage of uncertainty, since enterprises have not the certainty of how to interpret their duties of conformity. Finally, a significant number of laws non directly applicable on AI, are applied to it anyway. These overlapping areas are, for instance, intellectual property, antitrust, dat protection, financial regulation, mineral extraction, etc… This overlap mans that many enterprises have to follow and learn not only AI regulations in general, big also rules that influence the use of AI in the context of their sector or activity[38].

---

[37] Kostiantyn Ponomarov, *Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview*, https://legalnodes.com/article/global-ai-regulations-tracker

[38] *The global dash to regulate AI*, White&Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker#introduction

CHAPTER 2 - EU AI ACT 2024/1689 AND THE ROLE OF THE EU IN THE REGULATION OF ARTIFICIAL INTELLIGENCE

During the last few years, the European Union (EU) has recognized the importance of a legal framework that can govern the use and development of Artificial Intelligence. As a response to the challenges and opportunities presented by this technology, the EU proposed a series of legislative initiatives aiming at establishing clear norms to ensure that AI is developed and used ethically and responsibly. The most significant proposal is represented by the EU AI Act 1689/2024, Regulation on Artificial Intelligence, which introduced a risk-based approach, classifying the AI systems based on their potential incidence on fundamental rights and security.

2.1 Context and objectives of the AI Act

The AI Act, formally known as Regulation (EU) 2024/1689, represents a milestone in the emerging technologies regulations at a global level, being the first European legislation governing AI. With the introduction of this act, the European Union demonstrated strong efforts in guiding not only technological innovation but also in guaranteeing that such innovation is intrinsically linked to the protection of fundamental rights and the promotion of social well-being. In a context where artificial intelligence is more and more pervasive in several sectors, from the economy to security, from health to education, the necessity of a legislative framework that balances the technological development and the protection of EU fundamental values is imperative. This chapter explores in detail the reasons, objectives, and implications of the AI Act, contextualizing it within European policies and global challenges linked to the use of artificial intelligence.

The adoption of the AI Act is the result of a complex process, led by a series of factors that highlighted the need for a normative intervention at the European level. During the last few years, artificial intelligence has experienced a rapid evolution, characterized by significant developments in learning machine capacity, data elaboration, and in the automation of complex processes. This exponential growth has led to the spread of AI systems in key sectors such as medicine, transport, finance, and security, offering competitive advantages and new opportunities for the economy. However, the speed of this technological innovation has raised concerns about the capacities of existing legislation to properly deal with problems associated with AI. In the absence of an adequate normative framework, the indiscriminate use of AI might bring violations of fundamental rights, social inequalities, and irreversible damage to the privacy and security of individuals. An additional crucial motivation for the introduction of such an Act is the necessity of preventing normative fragmentation within the European Union. Before its introduction, Member states were prone to develop the most disparate national regulations for disciplining the use of AI, creating a splintered legislative scenario. This normative divergence not only compromised the functioning of the European single market, but it created legal uncertainties for the enterprises that operated at a cross-border level. The lack of harmonized regulation made innovation difficult for these enterprises, reducing the global competitiveness in a strategic sector such as the one of AI.

A paramount aspect of the Act is the protection of fundamental rights. The European Union has a longstanding tradition in the protection of human rights, as written in the EU charter of Fundamental Rights and other European treaties. The introduction of AI technologies, especially those based on complex algorithms and automated decision-making processes, could significantly influence these fundamental rights, among which are the right to privacy, non-discrimination, freedom of expression, and equal access to public services. The AI Act aims to ensure that the development and implementation of artificial intelligence in the EU conforms to these fundamental principles, establishing rigorous criteria for the AI systems that present a high risk of inference with the rights and freedoms of individuals. The Act is not only a response to the risks associated with artificial intelligence, but it also represents the opportunity for the EU to establish a global standard for ethical and reliable AI.

One first goal of the Act is the development of an artificial intelligence focused on humans. This means that AI should be designed and used for improving the well-being of people, sustain democratic values and promoting social equity. It should be transparent in its operations, reliable, understandable for the final users, and should include control mechanisms that allow adequate human supervision. This principle is particularly relevant in critical applications such as healthcare, justice and public services, where the decisions taken by the AI have a direct and significant impact on people's lives. The implementation of the AI Act requires a significant effort from the national and European authorities, as well as from private actors. It is important to develop clear guidelines and tools for supporting enterprises to adapt to the regulation's criteria. Furthermore, it is necessary the monitoring to guarantee that AI technologies respect the ethical and security principles established by the act. This might create new agencies to control or reinforce the already existing ones and the adoption of new audit and risk assessment techniques. It represents an ambitious attempt from the European Union to regulate artificial intelligence in a way to promotes technological innovation and the protection of fundamental rights and the public interest. With this regulation, the EU tries to establish a normative framework that is up to the challenges advanced by the emergence and rapid evolution of AI, reassuring that the benefits of such technologies are distributed equally and the risks adequately mitigated.

2.2 Definition of Artificial Intelligence in the AI Act

In the context of increasing deployment of AI technologies, it remains crucial to define clearly what is meant by the expression of "AI system", both to guarantee legal certainty and to provide an international alignment that allows a wide government convergence. The definition of AI has been object to numerous debates at the normative and international levels, and in 2024 the European Union formalised its interpretation with this regulation 1689/2024, establishing a legislative framework that disciplines the use and development of AI systems with particular attention to ethical, social and legal aspects.

The EU AI Act introduced a detailed definition of an AI system, describing it as "*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*"[39]. One distinctive aspect of these systems is their capacity to generate output, such as prevision, contents, recommendations, or decisions, based on the received inputs. Such outputs have the potential to influence both virtual and physical environments, which makes them particularly powerful and at the same time complex to regulate. Another important element is the concept of autonomy: this does not necessarily imply a total detachment from human control, but it highlights how an AI system can make decisions or undertake actions independently from direct human intervention. This autonomy is variable, and it can manifest in different degrees based on the specific application of the system in matter. For instance, some systems can require constant interaction with the user, whereas others can operate in

---

[39] European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 24 September 2024 on Artificial Intelligence*. Official Journal of the European Union. Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689/oj

a completely autonomous way once activated. This characteristic distinguishes AI systems from traditional software, which are generally bound to predetermined rules and do not show any skill of apprehension or adaptation.

The AI Act underlines that inference capability is one of the key aspects that distinguishes AI systems from traditional software. The inference, in this context, refers to the ability of the system to elaborate the inputs to generate models, algorithms or both elements, that subsequently produce outputs destined to influence the operational contexts. This process is at the base of machine learning, one of the main AI techniques, where the systems learn data to achieve specific goals, but it learns also approaches based on logic and knowledge.

Another aspect regards adaptability, namely the systems' capacity to modify their behaviors after the phase of deployment, through processes of auto-learning that allow them to evolve based on data gathered and cumulated experiences. Such capacity permits the systems to improve their performances, resulting in particularly useful in dynamic applications, such as the management of energetic demand or predictive maintenance in industrial sectors.

An additional concept is the difference between implicit and explicit objectives: AI systems can operate following clearly defined objectives or, in other cases, act based on implicit objectives that emerge from the operational context or the training of the system. These objectives can differ from the purpose declared, highlighting the importance of monitoring and checking the use of AI systems so that their behaviors do no escape human control and the purposes which they had originally been designed for.

2.3 Structure of the EU AI Act

The regulation (EU) 2024/1689 on artificial intelligence is structured with several sections and articles that delineate the norms and requirements for the development and use of artificial intelligence systems in the European Union. The first part establishes the scope of the regulation, namely the improvement of the internal market functioning and guaranteeing high levels of health protection, security, and fundamental rights. Several key definitions are delineated to clarify the terms used in the regulation, such as "artificial intelligence system", "providers" and others. A third section describes how artificially intelligent systems are classified based on the degree of risk they present. The categories include in descending order: Unacceptable risk, High risk, limited risk and minimal or no risk. The fourth section establishes specific obligations for both the providers and users of these systems, along with the criteria of transparency, security and responsibility. The fifth section provides the creation of an artificially intelligent office and our European Council on Artificial Intelligence, with the responsibility of coordinating the application of norms at both national and European levels, developing competencies and capacities in the field of AI, and furnishing support and assistance to member states and the other operators of the sector. The sixth chapter delineates sanctions for non-conformity to the dispositions of the regulation, which can include administrative fines, suspension or prohibition from operating in the EU market, and obligation to conform by a certain period. The final sections regards the final dispositions on the entry into force of the regulation and the future revisions. A calendar is established for the implementation of the dispositions, with specific deadlines for the creation of conduct codes and for the operability of the governance structures. This detailed structure aims to guarantee a balanced and coordinated approach to the use of AI, promoting innovation while protecting rights and human security.[40]

2.4. Classification and degree of risks

The AI Act adopts a risk-based approach to technology regulation to keep pace with the rapid advancements in AI. Consequently, compliance obligations are dependent on and tailored to the level of risk applicable to allow for adequate command. If any are determined on a system basis, an organization should integrate AI management into its overall frameworks for risk management compliance and governance. This allows the AI system to continue achieving its objectives while avoiding being reclassified into another category of risk. There are four categories of risks associated with AI systems as defined by the EU AI Act: Unacceptable risk, High risk, Limited risk, and Minimal or no risk. Artificial intelligence can have some degree of risk according to the situation it is in and the specific application and level of development it boasts in addition to being able to harm public interests and fundamental rights at times.

---

[40] European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 24 September 2024 on Artificial Intelligence*. Official Journal of the European Union. Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Unacceptable risky, meaning AI systems contradicting EU values of respect for human dignity, equality, democracy, rule of law, and in general fundamental rights. These systems are prohibited and include systems capable of manipulating and persuading individuals to perform undesired actions or decisions outside their normal range of choices; artificial intelligence systems that take advantage of the weaknesses of an individual or particular category of people based on their age, handicap or some specific socio-economic status; Biometric-based categorization systems that use an individual's biometric information, such as an individual's face or fingerprints, to determine or speculate their attributes such as political views, religion, ethnicity, and/or sexual preferences; "real-time" remote biometric identification systems used in public areas for law enforcement.

High-risk systems are identified among those which have a significant impact on the health, safety, and fundamental rights of people in the EU. Providers and deployers of high-risk AI systems must follow specific obligations such as risk management, human oversight, accuracy, technical documentation and record-keeping. These systems include remote biometric identification systems, biometric categorisation, and emotion recognition systems, safety components in critical infrastructure, like water, gas, and electricity, access to education and vocational training, evaluation of performance, and essential private and public services.

Limited risk systems mostly concern the lack of transparency in the use of AI. It is important that all AI systems interactions with people, such as deep fakes and chatbots, are developed following criteria of information to the user and transparency.

Finally, the minimal or no risk systems (including the majority of AI applications currently available on the EU single market, such as AI-enabled video games and spam filters) are not as regulated as the previous categories; however, no risk systems, through their use and expansion, may be changed in their nature and cross into a category with a higher degree of risk. Therefore, proper governance is essential.[41]

---

[41] FORVIS Mazars. (n.d.). *EU AI Act: Different risk levels of AI systems*. FORVIS Mazars. Retrieved from https://www.forvismazars.com/ie/en/insights/news-opinions/eu-ai-act-different-risk-levels-of-ai-systems

2.5 High-risk AI systems

According to article 6 of the EU AI Act[42], generally speaking, an AI system is considered "high-risk" when two conditions are essentially met: such a system is destined to be used as a security component of a product or the system itself is a product disciplined by EU legislation. Such systems must be subjected to a conformity evaluation by third parties before being sold or used. Some AI systems are always considered high-risk, unless the cases where they do not represent a significant risk to health, security or human rights. Providers who believe their system is not high-risk must document their assessment before selling or using it. The European Commission can modify the record of AI systems considered high-risk if they meet certain conditions, by taking into consideration factors such as the purpose of the AI system, when it is used, the elaborated data, damage caused in the past, potential damage, and its capacity to correct or invert results. Such list of AI systems can also be modified by the EU Commission, in terms of removal of systems that are no longer considered high-risk ones.

2.5.1 Requirements for high-risk AI systems

AI systems classified as high-risk must follow rigorous standards, adequate to the purpose and current AI technology, and providers, on their part, must respect and guarantee the conformity of the product to all applicable norms; they can integrate tests, reports, and documents requested by already-existing European norms, merging them in their current procedures, avoiding duplication of work and reducing administrative costs, since a single set of documents and controls can satisfy several different normative requirements.

The EU legislation on artificial intelligence imposes the adoption of a risk management system for High-risk systems, which must be continuously reviewed and updated during the entire life cycle of the system. This process should identify, analyze and evaluate the potential risks to the health, security and fundamental rights, by adopting adequate measures for mitigating them. The action shall reduce the risks to the minimum level, reassuring the residual ones to be acceptable and manageable. Furthermore, high-risk systems must be subjected to tests to guarantee that the work is

---

[42] European Parliament and Council. *Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act), Article 6*. Official Journal of the European Union, 13 September 2024. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

as expected and that the management and measures are effective. Special attention must be reserved to the impact that such systems might have on minors of 18 years old and vulnerable groups. To guarantee the quality of high-risk systems, it is essential to use a set of high-quality data for the formation, validation and testing. These sets must be accurately managed, considering also the collection processes, preparation and potential gaps, and data must be relevant, representative and with no error, adapting to the specific context where the system is used. In some cases, providers can supervise personal data to identify and correct errors, always concerning the protection of individual rights and freedoms.

The EU AI Act establishes that before the launching of a high-risk AI system, updated and correct technical documentation must be held, which demonstrates that the system satisfies the legal requirements and verifies the conformity, it shall include key elements that allow for a transparent evaluation. Small enterprises, including start-ups, will be able to present such information through a module that the European Union will make available.

High-risk systems must be able to register automatically relevant events throughout their entire duration. This guarantees the tracking of the system's actions. The registration should include details such as the use of databases, and who verified the results, to reassuring accountability and safety in the use of such a system. Secondly, systems must be designed to be transparent, so that the users can comprehend the functioning and use them properly; must be accompanied by clear instructions including information on the provider, capabilities, limits, and potential risks of the system, and must be able to illustrate how to interpret the results, possible modifications and how to maintain the system. Lastly, these systems must be projected to allow effective human supervision, to minimize the risk to health, security and fundamental rights. Supervision measures shall be proportionated to the context and risk of the system and allow the supervisor to comprehend capabilities and limits, deal with potential issues, avoid an excessive dependency on the system, and decide whether to interrupt the functioning if necessary.

2.6 Obligations for providers and deployers of certain AI systems

In Article 50 of the EU AI Act 2024/1689, a clear obligation for both the providers of AI systems to inform users every time that they interact with a system emerges, such obligation unless the use of the system is not immediately evident or expected for legal purposes, like the prevention or prosecution of crimes. This principle extends also to AI systems that generate synthetic content, like modified audio, videos, or texts: such outputs must be labeled as artificially generated to guarantee transparency. In particular, the use of deep fakes, emotion recognition, or biometric categorization systems, imposes specific notifications to users, apart from when such systems are legally authorized for security tasks or criminal justice.

An ulterior transparency obligation regards information about the purposes or modalities of generation and manipulation of contents, in a way that the public can recognize the contribution of AI in different contexts, except in legal cases or satin and artistic works. The AI office must facilitate the implementation of conduct codes at the European level, to support the application of these obligations, contributing to defining common rules for the detection and labeling of artificially generated content. This ensures that the evolution of AI maintains a balance between technological innovation and the protection of fundamental rights, forming a normative framework that can adapt to new technological challenges.

The article then specifies that the information relating to the use of AI systems must be provided in a clear and distinguishable way at the moment of the first interaction, in conformity with the predetermined accessibility requirements. Nonetheless, the underlined legislation does not influence other transparency obligations already established by the Union or national law.

CHAPTER 3 - ARTIFICIAL INTELLIGENCE AND COUNTERTERRORISM

3.1 Introduction to terrorism: evolution and political implications

Terrorism is not a new concept: the idea of destabilizing the enemy through psychological means is present in numerous classic strategies, as demonstrated in historical strategic texts and documents. The very first systemic use of terrorism in history goes back to the first century A.C. with hitmen, knows also as Zealots. These were mostly part of a radical religious movement that opposed the Roman occupation of Palestine. Their actions included casual homicides in public spaces and tactics aimed at creating chaos to weaken the Roman resistance and dissuade them from maintaining control over the region. However, Roman power was revealed to be insurmountable, and hitmen, in the end, retired to the Masada fortress, where they preferred committing suicide rather than being captured. About one millennium later, in Persia, another of the most impactful terrorist groups emerged: the order of Hashashin, or Assassins. This group, an Ismaili Sect led by a spiritual leader called "the old mountain man", has challenged the Turkish sejulchidi authority for over two centuries, the authority that took over the Arabs from the control over the Middle Eastern region. Protected by a mountain fortress, assassins perpetrated numerous relevant political homicides, aimed at high-ranking figures. Among their main successes, is the murder of the Persian grand vizier Nizam al-Mulk in 1092 and of Corrado di Monferrato, one of the leaders of the Third Crusade, in 1192. Their fame spread also in Europe, where the idea that they could be employed to solve political disputes linked to the succession process started spreading around as well. Nonetheless, despite some tactical achievements, assassins couldn't achieve their political and religious goals and, in the XIII century, they were annihilated by the Mongols with surprising ease.

It is difficult to establish how much of the assassins' techniques had been copied by the Mongolian empire, but the terror tactics occupied a relevant role in general strategies. The invincible Tamerlano used terror as the main instrument to conquer cities, creating massive skeleton pyramids of those who had not given up in the battle, a technique developed by its son Miran Shah. Such techniques were not so common in Europe, but the 30-year War, religiously motivated, experienced several terroristic tactics. One century later, the French Revolution gave officially a name to the acts of political violence: terrorism. Despite this form of terrorism was regarded as the use of terror by governments, substantially it annunciated the beginning of a new era for terrorism. Therefore, it is possible to say that the use of terror as a common instrument of modern politics dates back to the

French Revolution, at the same time when more positive evolutionary concepts emerged, such as human rights and democracy. The American Revolution, on the other hand, did not approve of the use of terror, however, one of the first foreign policy decisions by the new American republic with resident Thomas Jefferson could be seen today as characterized by a “Terror War”. It regarded the creation of a fleet that managed to free the Mediterranean through terrorism sponsored by the state, conducted by barbarian pirates for many centuries.

Albeit the ideological forces unleashed by the French evolution have in the end, sustained several terrorism waves in the XIX and XX centuries, the emergence of modern terrorism was made possible by the conjunction of different important events, among which the Industrial Revolution, the collapse of ancient empires such as the Austrian, Russian and Ottoman ones, and the emergence of concurrent ideologies like liberalism, nationalism, anarchism, nihilism and marxism. Mostly, terrorism in the XIX and XX centuries remained laic, with religious terrorism that re-emerged in the 1980s after the Iranian revolution in 1979. Modern terrorism was born at the same time as modern democracy, and since then, it has never stopped challenging it.

Although there were isolated terrorist acts already in 1800, the first wave of modern terrorism started at the beginning of the 1860s, a decade that experienced the rise of the nihilist ideology in Russia and the invention of the dynamite by Alfred Nobel. Nihilists have not committed significant terror acts; however, thanks to the influence of Sergei Nechayev, they prepared the ground for political violence aimed at destroying the existing social and political structures. The invention of dynamite, in this context, was of extreme importance for the emergence of terrorism. For many revolutionaries, dynamite was seen as a blessing since it allowed a small clandestine group to challenge the authorities. Anarchists were the most enthusiastic: not organized and with few means, aiming to destroy but not interest in reconstruction, saw this explosive as a way to create necessary chaos to unleash revolutions. Europe, Russia, and the US became all victims of anarchist terror waves. In the advanced industrial economies, the increase in tensions between workers and employers pushed the most radical syndicates to resort to terroristic tactics. In Russia, the liberal policies of Tsar Alexander II had the undesired effect of raising social tension and opening up the way to any type of terrorist activity. The same tsar was the victim of terrorists in 1881.

For contemporary observers, terrorism is inevitably linked to political violence against civil populations. Indeed, today the majority of definitions of terrorism revolve around the goal of civil populations. Nevertheless, the link between terrorism and these populations has not been always so automatic: at the beginning, modern terrorism had a strong affiliation with the ancient tradition of Tyrannicide, which sustained that a citizen had the moral duty to free the city or country from the unacceptable abuses of the tyrant, resorting to murder. This tradition was revitalized by philosophers and intellectuals of the Renaissance and became more and more popular with the advent of the revolutionary era at the end of the XVIII and XIX centuries. Consequently, the majority of terrorism victims in that period were heads of state, political leaders, or royal family members. As it happens today, multiple attacks and the media attention that they aroused, inspired numerous local terrorists to express their discomfort against society through isolated violent acts on behalf of anarchism. Mostly, the anarchists were local criminals with superficial ideological positions. The anarchists developed the propaganda by deed, referring to the idea that actions, particularly violent and revolutionaries ones, can serve as a form of propaganda to spread anarchist ideals and inspire others to unite in the cause[43]. Although the majority of terrorist acts carried out by anarchists were not followed by attempts to take power, they contributed significantly to the deterioration of political structures that were already weak in many countries. Maybe even more importantly, their success suggested that terrorism could revealed to be a significant tool to assist them in the journey towards socialism. With the collapse and weakening of the great empires, also nationalists saw terrorism as a technique useful to their cause.

Terrorism, which up until that moment had been confined to isolated individuals or small groups irrational by nature and with no clear political goals, became, at the beginning of the XX century, a popular technique used by various organizations with more concrete political objectives. Despite the majority of these groups having nothing in common with the anarchists, they adopted a great number of their ideas and tactics. The rise of these new terrorists marked also a distancing from the radiation of Tyrannicides. As a propaganda instrument or as a means to mine political structures through the creation of chaos, terrorists started targeting the general population rather than people in the inner cycles of political leaders and royal families.

The emergence and spread of democracies had a significant effect on the nature of terrorism. Before its advent, the state representative was the head of state. With democracy, the representative became also the anonymous citizen. With democracy, all the freedom instruments that shaped public

---

[43] Colson, D. (2017). *Propaganda and the Deed: Anarchism, Violence and the Representational Impulse.* American Studies, *55/56*, 163–186. http://www.jstor.org/stable/44982624

opinion arrived, among which, national and regional newspapers. The notable urbanization process verified at the end of the XIX century, meant that a large amount of people lived in limited space and became an easy target for the attackers. Terrorists quickly learned that a terrorist attack could have a formidable impact on public opinion, amplified by media mechanisms. In the same way, the government rapidly learned that it could exploit terrorism in a variety of ways. Indeed, the beginning of the XX century, represented a big opportunity for agents to perpetrate terrorist acts themselves, on behalf of their government, they infiltrated terrorist networks and, generally, engaged in dark political operations. The Tsarist police was an expert in this type of manipulation. Lenin and the Bolsheviks were not so intransigent regarding the systemic use of terrorism. Lenin believed that the use of terror had to happen at the right opportunity. Despite the Bolsheviks exploiting the political situation weakened by political violence and terrorism generated by others, they quickly understood how to manipulate terror instruments to reinforce their positions.

While anarchist terrorism was being eliminated, also in the United States, a more dangerous form of terrorism was taking shape: nationalism. Nationalist terrorism emerged as a violent strategy used by groups seeking independence and auto-determination, by exploiting the chaos to weaken the existing state structures. These groups saw terrorism as a means to obtain international visibility and destabilize dominant regimes, with the end goal of mobilizing the popular consensus and forcing political concessions. Despite being initially inspired by anarchist tactics, nationalist movements adopted terrorism to create disorders, making it a key component of their fight for independence and the overturning of imperial and colonial powers.

Before and after the First World War, terrorism manifested especially in the regions of the Austrian and Ottoman empires, in particular in the Balkans, where nationalists of different ethnicities used political violence. A key event was the assassination of the heir to the Austrian throne, Francesco Ferdinando, by a group of Serbian nationalists I Sarajevo, which initiated the First World War. A new generation of terrorists emerged in the inter-war period, when nationalist movements adopted fascist ideologies to obtain support from Germany and Italy. The Second World War saw some marginal terroristic activities; however, after the conflict, terrorism re-emerged, essentially as part of the fight for anti-colonial independence, that exploded in different regions of the world. In Palestine, zionist organizations used terrorism against English troops for instance. Terrorism was used complementary to guerrilla and political maneuvers. The majority of anti-colonial terrorist groups were heavily influenced by Marxist ideologies. In the 60s and 70s, a new generation of terrorists arose, often inspired by leaders such as Mao Zedong and Che Guevara, who believed in the possibility of a revolution through terrorism and urban guerrilla. These groups, active in Latin America, the United States, and Europe, caused many victims, but they collided with the reality that the Western world was not ready for a socialist revolution. In Latin America, terrorism contributed to the rise of military regimes, whereas in Italy Red Brigades assumed a primary role.

In the '60s, the Middle East experienced the emersion of terrorist organizations, that aimed to the creation of a Palestinian state through highly visible terrorist acts, destined to capture the attention of the International community. With new and innovative techniques, such as the hijacking of planes, Palestinian terrorism couldn’t achieve the empathy of the international public opinion because of terror acts such as the murder of eleven members of the Israeli team during the Olympic games in 1972. In that period, Palestinian terrorism, as many other movements at that time, was secular and the religious influence manifested only in the 80s and 90s.

With the decline of left-wing groups in the 1970s, new groups with revolutionary goals started operating, with the idea of a religious revolution rather than a socialist one. The Iranian revolution of 1979 generated various acts of political violence, including terrorism. In the same way, the soviet invasion of Afghanistan led to the rise of a new generation of religious extremists determined to destroy not only the Soviets but also the West. Since politicians are prone to think about short-term gainings rather than long-term responsibility, the American government sustained insurgent individuals and groups, that later on would turn against America. After the Soviet withdrawal, Afghanistan became a refuge for Islamist extremists and a sanctuary for terrorism. In this context, Al-Qaeda was born, and similarly ISIS.

Although it is possible to trace terrorism back to a couple of millenniums ago, the terroristic activity has generally been sporadic, and only from the mid-XIX century on, its impact has been global. The first incarnation of modern terrorism helped launch two political events of great importance: the Russian Revolution and the First World War. Events of that era defined in numerous ways the post-Cold war world. The impulse towards democracy that followed the end of the East-West fight created a fertile ground for the emergence of violent radicalism, especially in areas where modernization and economic growth could not radicalize. Therefore, the growing irrelevance of post-colonial dictatorships and the fragility of the nascent democracies provided radical groups with a favorable environment to challenge the status quo. The Middle East and Sub-Saharan Africa have been revealed to be fruitful for terrorism. Economic, social, and political tensions have powered the general resentment among Muslim populations towards the West, which radical Islamist groups have tried to take and leverage with success to a certain point[44].

3.2 Terrorism trends and Evolving Threats

Terroristic threats have received deep evolutions compared to the past, especially after 9/11. The advent of social media enormously amplified the power of extremist movements and terroristic organizations, providing them a platform for planning, recruiting new members, and creating online global networks. This has facilitated the direction and inspiration of distanced attacks, without the necessity of a physical presence (this topic will be further analyzed at point 3.6 of this chapter). Social media have therefore represented a diffusion channel of propaganda and recruitment, in particular among youth, making it more difficult for authorities to prevent radicalisation.

Over the last few years, extreme right-wing terrorism has emerged as a significant menace: this type of extremism inspired numerous devastating attacks all over the Western world, with targets that vary from Hebrews to Muslims, and ethnic and religious minorities. The white supremacist ideologies, fed by anti-immigration and hatred rhetorics, have been revealed to be particularly dangerous. The rebirth of rich-wing terrorism adds a new dimension to the already complicated situation of global security, further expanding the spectrum of menaces that authorities have to deal with.

Another significant change is represented by the so-called policy of “decapitation” adopted by the United States to hit high-profile individuals of terrorist groups like Al-Qaeda. Operations like the killing of Osama Bin Laden and Anwar al-Awlaki have inflicted hard blows on these groups’ infrastructures. Nonetheless, despite these operations having eliminated key actors, threats persevere. Thanks to their capacity to regenerate and maintain their influence through decentralized radicalization and propaganda. Other than these visible changes, a series of more subtle trends are rising and contributing to making the scenario of global terrorism even more complicated. An important transformation regards the growing blurred lines between national and international

---

[44] Chaliand, G., & Blin, A. (Eds.). (2017). *The history of terrorism from antiquity to ISIS* (Updated ed.). University of California Press. https://books.google.it/books/about/The_History_of_Terrorism.html?hl=es&id=U6swDwAAQBAJ&redir_esc=y

terrorism. In the past, these two categories were neatly distinguishable, but in the digital era this distinction has almost completely disappeared. International Jihadist groups, for instance the Islamic State, increasingly rely on local actors, often defined as “lone wolves”, to perpetrate violent acts in the countries of origin. For example, in May 2016, Abu Mihhamand al-Adnani, a speaker on behalf of the Islamic State, escorted his followers in the Western countries to commit attacks in their own countries, rather than uniting in the wars in Syria and Iraq. This call was heard and followed by violent acts all over Europe and the United States, demonstrating how global terroristic organizations can exercise their influence from miles away without physical presence.

In parallel, extremist movements that traditionally focused on local matters, are now expanding their capacities at the international scale. A significant example is the attack on Christchurch mosques, in New Zealand, in 2019. Brenton Tarrant, the author of this massacre, was an Australian white supremacist who had been traveling and meeting with other extremists in Europe before perpetrating the attack. The event represents an extreme case of what is defined “*the self-referential nature of extrem-right terrorism*”, where extremists seek inspiration beyond their national borders. Despite the phenomenon of lone wolves increasing, this does not mean that the structured terrorist organizations and networks, with leaders and command chains, have lost their importance or have disappeared from the international scenario.

On the tactical level, terrorists today use a variety of tools, that range from simple weapons to more advanced technology, to carry out attacks that, in some cases require minimum logistics. Attacks with knives, mass shootings, and the use of vehicles as weapons are examples of less sophisticated actions, but extremely lethal, that can be executed by individuals with no specific formations. However, these attacks, although simple, have been demonstrated to be as devastating and difficult to prevent. On the other hand, there are highly trained terrorists that use more complex technologies to plan coordinated and large-scale attacks, like the cases of Paris 2015 and Brussels 2016, where explosive jackets and suicidal bombs were fabricated.

New technologies have then made the elusion of national norms on the possession of weapons easier. One exemplary case of such technological evolution is the use of 3D prints to fabricate firearms, as demonstrated by the attacks in Halle, Germany in 2019. On that occasion, the attacker used arms printed in 3D and publicized a manifesto containing the description of the process, to show the ease of creating improvised guns. Other than the 3D print, the increase in the use of “ghost guns", untraceable firearms assembled through online purchased kits, constitutes a growing menace for armed forces, since these guns can easily bypass legal controls.

Contemporarily, there has been an evolution in terms of targets chosen by terrorists. The complex and well-organized attacks against highly representative entities, like embassies, airports, and military bases, have come to be less frequent. Today, terrorists tend to prefer easier and quicker actions, hitting public and crowded places, which are more easily accessible, they offer fewer opportunities for defense and less sophisticated planning. This change posed armed forces and decision-makers in a difficult position, who needed to find new solutions to protect such environments free from physical barriers.

Beyond these tactical and strategy developments, the overview of terrorist threats has diversified notably. Terrorism financed by states keeps being a significant concern, with some governments providing financial and logistical support to terrorist groups to promote their political objectives. A relevant example is Hezbollah, which received support from Iran. There exists also a threat linked to nationalist and separatist movements, that aim to create independent states or obtain auto-determination for specific ethnic and cultural groups. Despite movements like IRA and ETA have diminished over the years, some territorial conflicts and ethnic tensions persist, which could ultimately give rise to new separatist terror waves. Also, terrorism linked to specific causes, such as environmentalism and animal rights, started operating, even though these groups are prone to focus on damages to property rather than attacks with numerous victims.

Finally, cyberterrorism is increasing and constitutes a new threat in the spectrum of menaces. Terrorism is increasingly using the internet, not only to spread propaganda and recruit members but also to launch informatics cyber attacks against critical infrastructures. Cyberterrorism represents a particularly difficult threat to deal with since it allows terrorists to operate anonymously, from remote and unknown places, to attack financial systems, governmental websites, or essential services without direct physical involvement. The new digital reality not only confuses national and international distinctions but also created new virtual battlefields[45].

---

[45] B. HOFFMANN, J. Ware, (2020, September 24). *Challenges for effective counterterrorism intelligence in the 2020s*. Lawfare. https://www.lawfaremedia.org/article/challenges-effective-counterterrorism-intelligence-2020s

In conclusion, the overview of terrorist threats has become more and more complex and diversified, with actors that operate on more levels and exploit different tactics and toolset to carry on their operations. Emerging technologies, such as social media and cryptocurrencies, are facilitating the communications, financing, and execution of attacks, making it more difficult to intercept and prevent such activities. Terrorism is not a phenomenon limited to national or regional borders, but it has become global and requires, consequently, a coordinated and global response.

3.3 Evolution of Counter-Terrorism Strategies

The concept of counterterrorism is complicated to define, especially in western democracies. Paul Wilkinson observes that there is not an anti-terrorism policy that is globally applicable for democracies[46]. Each conflict that involves terrorism has unique aspects. Wilkinson and Louise Richardson[47] believe that western democracies should respect civil freedoms and the rule of law in their counterterrorism strategies. Although su piece of advice is commendable and consistent with democratic principles, it does not constitute a real counterterrorism strategy, just some principles to follow. The American army's manual describes counter terrorism as "*operations that include the offensive measures taken to prevent, deter, preempt, and respond to terrorism*"[48], definition that presents both strengths and weaknesses. On one hand, it recognises that counterterrorism is a comprehensive doctrine that covers prevention, dissuasion, responses, requiring the employment of all national and international resources. On the other hand, this definition is so wide that it dose not clearly distinguish the different elements of counterterrorism, that could complicate the development of efficient strategies, the administration of resources and the establishment of accountability. Despite these limitations, a global approach to counterterrorism has its advantages. It allows governments to recognise the complexity of response to terrorism, reinforcing the idea that there is not a single easy solution to the problem. Counterterrorism operations change based on the nature of the threat. International terrorism, in particular organisations such as Al-Qaeda, remains persistent and adaptable. Even if it is not possible to eliminate all forms of terrorisms, it is possible to adopt measures to interrupt, dismantle and defeat organisations that use it.

---

[46] Paul Wilkinson. *Terrorism Versus Democracy: The Liberal State Response*. New York: Routledge, 2006, p. 203.

[47] Paul Wilkinson. *Terrorism and the Liberal State*. Basingstoke: Macmillan, 1977, 1986; Louise Richardson. What Terrorists Want: Understanding the Enemy, Containing the Threat. New York: Random House, 2006

[48] US Army Field Manual, 2006, p. 4.

To better comprehend the contemporary efforts in countering terrorism, it is crucial to distinguish between counterterrorism policy and the counterterrorism operations that derive from it, other than understanding the different objectives and competencies involved. In many democracies, counterterrorism policies are based on a principle of "non concession" towards terroristic organizations, a position that tries to discourage further attacks. This approach emerged in response to particularly traumatic terroristic events, where concessions to terrorists might have produced a dangerous precedent. However, even if the non concession policies are a pillar in many countries, counterterrorism operations have evolved in a wider context, especially with the advancement of technologies and the use of asymmetric war techniques. Counterterrorism operations, which often involve police, military, and intelligence agencies, aim to prevent, interrupt, or respond to terrorist acts. These operations include a wide range of activities from surveillance to direct action, such as the targeted attacks against leaders or operations to dismantle terroristic networks. Nonetheless, the direct intervention can vary enormously based on the type of threat or the political context wherein is undertaken. For example, while some operations might focus on the use of drones to hit specific targets, others might focus on negotiations with rebel groups, according to the specific situation and the strategic goals.

Counterterrorism operations might also appear contradictory due to the different positions and strategies adopted by the diverse institutional actors involved. Often, sectors such as the Ministry of Foreign Affairs or the Ministry of Defence could adopt different approaches based on the terrorist organization they are dealing with. In some cases, the necessity to converse with some terrorist groups emerges, to achieve long-lasting political stability, especially in the case of groups with concrete political and social objectives. Nevertheless, this approach is not universally applicable, particularly in the case of organizations considered irreducible, like transnational jihadist groups with radical ideas, to which military and security solutions are the only acceptable responses. The complexity of modern counterterrorism reflects a significant evolution compared to the last decades. While in the past the operations were often limited to local or regional contexts today the fight against terrorism is a global matter that requires international coordination and adaptability. The main challenges include not only the capacity to prevent imminent attacks but also the capacity to comprehend and dismantle networks that feed terrorism, just like maintaining a balance between security and civil rights.

The evolution of counterterrorism in the last four decades has been profoundly influenced. The terroristic tactics have adapted to the geopolitical and technological contexts, and consequently also the counterterrorism strategies have been constrained to transform to respond to the new threats. Modern international terrorism started obtaining global visibility on the 22nd of July 1968, when members of the Popular Front for the liberation of Palestine hijacked a commercial flight from Rome to Tel Aviv[49]. This event represents a breaking point with traditional forms of terrorism, since the hijacking, as above mentioned, did not aim for criminal or financial purposes, but it depicted a symbolic and political action. This change of strategy, from the local to the global stage, inspired other terroristic organizations, such as ASALA[50], and Marxist groups such as Baader-Meinfhof[51]. These groups learned that operating at the global level could obtain a greater media resonance. In the 70s, these media terrorism tactics posed the governments under growing media and diplomatic pressure. Despite the relatively low number of victims, each hijacking or kidnapping was transformed into an international crisis followed by global media, forcing the governments to make difficult decisions, sometimes counterproductive ones. Facing these challenges, many governments started developing special units of rapid intervention, such as the anti-terrorism units. Germany, for example, created the GSG-9, one of the most elitist forces in the world. These specialized corps, trained to intervene in situations of hijacking or kidnappings, became a crucial instrument for counterterrorism operations. With the incrementation of globalization and the advent of transnational terroristic networks, counterterrorism strategy had to adapt. During the Cold War, many terrorist organisations found support from states that used them as proxies in the global ideological war. However, with the end of the Cold War and the emergence of new threats, like al-Qaeda in the 90s and ISIS in the 2000s, the terrorist scenario changed radically. Terrorists were not limited to hijacking and kidnapping anymore; new generations of terrorists tried to maximize the number of victims through suicidal attacks or the use of explosives. 9/11 has been a turning point in this sense, demonstrating that terrorist groups were capable of orchestrating devastating attacks on a large scale with global geopolitical effects.

---

[49] Hoffman, B. (2006). *The Internationalization of Terrorism*. In Inside Terrorism (REV-Revised, 2, pp. 63-80). Columbia University Press. http://www.jstor.org/stable/10.7312/hoff12698.7

[50] Armenian Secret Army for the Liberation of Armenia

[51] The Red Army Faction, also known as the Baader–Meinhof Group or Baader–Meinhof Gang, was a West German far-left militant group founded in 1970 and active until 1998

As a response, counterterrorism evolved towards a more global and preemptive approach. Other than the rapid intervention units, governments started developing complex networks of intelligence, surveillance programs, and international collaboration. The fight against terrorism has become a top priority of national security, and many countries have adopted policies aiming at preventing attacks before happening. This led to a debate on the fine line between security and civil freedoms and rights, since the massive data collection, electronic surveillance, and secret operations raised concerns regarding the respect of fundamental rights (this debate will be analysed in the final chapter of this thesis).

The evolutions led governments to re-evaluate their strategies, including the use of new military technologies, such as drones without pilots. In the years following the 9/11, attacks with drones, especially used by the USA, became one of the most used tools to hit leaders and terrorist networks' members in remote areas. This use is represented in the sector of the so-called Revolution in Military Affairs (RMA), based on the use of increasingly precise conventional weapons, reducing the human losses among troops and minimizing collateral damages. Despite short-term killings can be seen as a success, many leaders and fighters can be easily and quickly substituted and the collateral damage of such killings can alienate the local population, compromising the strategies more focused on populations, such as stability policing functions of counterinsurgency. The use of drones does not allow the possibility of surrender, often violating the sovereignty of other states. Drones, and other advanced technological devices and instruments, and the integration of artificial intelligence in the military and political systems, have further influenced and changed the strategies used to fight against terrorism. This evolution underlines the importance of adapting the response to the different threats. The efficiency of contrast operations not only depends on the technology used but also on the deep comprehension of social, political, and economic dynamics intrinsic to terrorism. Therefore, an approach that balances out all the aspects above mentioned supported by technological innovations, is fundamental to dealing efficiently with future challenges in the field of global security[52].

---

[52] Rineheart, J. (2010). *Counterterrorism and Counterinsurgency*. Perspectives on Terrorism, 4(5), 31-47. http://www.jstor.org/stable/26298482

3.4 Artificial Intelligence applications for contrasting terrorism

During the last couple of centuries, terrorist groups and violent extremists have refined the use of digital technologies, exploiting online communication platforms to spread ideologies recruit members, and plan operations. The use of social media, encrypted messaging applications, and multimedia content in multiple languages has transformed the landscape of threats, making terrorists capable of acting at large scale with no physical barriers. These tools not only amplify the capacity of their activities but facilitate the coordination of complex operations and the maintenance of clandestine networks. As a response to this growing menace, security and anti-terrorism agencies are relying on artificial intelligence to improve operational efficacy. AI distinguishes itself for the ability to process rapidly large volumes of data and identify models that are hardly recognizable by human beings. In the context of the fight against terrorism, these technologies can automatise the surveillance of social networks, detect radical content, and track suspected behaviors. For instance, advanced algorithms can individuate financial transactions linked to terrorism financing, recognize radicalization schemes, and provide previsions on potential future threats. Moreover, AI can help armed forces filter enormous data rapidly, saving resources and reducing the necessary time to complete tricky investigations. This is particularly useful when dealing with the monitoring of communication networks and online platforms used by terrorists to recruit individuals or plan attacks. The capacity of AI to act in real-time allows it to intervene faster, preventing potential attacks before happening. The efficiency of an AI system depends on the quality of data on which the system is trained on: if the data contain prejudices, these might be replicated in the results, compromising the justice of operations. In a more and more digitized world, the strategies against terrorism will have to be technologically advanced to carry out more efficient and targeted operations, and ethically solid, granting that the use of AI is transparent and respectful of international provisions[53].

---

[53] United Nations Counter-Terrorism Centre & United Nations Interregional Crime and Justice Research Institute. (2021). *Countering terrorism online with artificial intelligence: Opportunities and challenges*. United Nations Office of Counter-Terrorism. https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf

3.4.1 Predictive analysis and detection of threats

Even if the immediate management after an attack is fundamental, it is always preferable to prevent it. For this reason, prevention often represents the heart of counterterrorism strategies. There are two main ways to prevent attacks: deterrence, based on the protection of infrastructures, application of security controls and punishment; and the negation of the ability to perpetrate an attack, through the early arrest of a terrorist before concretizing their plans, the fight against recruitment and radicalization of a potential terrorist, and the imposition of restrictions on freedom to move and actions of specific people. The capacity to foresee events allows to apply preemptive measures in a targeted way, minimizing the impact on the general population. For instance, an efficient forecast might endure that only violent extremists suffer coercive measures, whereas more conciliatory approaches be applied to those who are vulnerable to radicalization. Artificial Intelligence technologies aim to reproduce or overpass abilities that would require "intelligence" if carried out by humans. These abilities include apprehension, adaptability, sensorial comprehension, interaction, reasoning, planning, optimization of procedures and parameters, autonomy, creativity, extraction of knowledge and forecasts from digital data.

Over the last decades, the interest in the development of web-crawling tools for the collection of large quantities of online content has increased evidently. This interest has found application also in the study of terrorism and extremisms. Scholars started using standard web-crawling tools and developing personalized programs for the gathering of online data more efficiently. One example is represented by the Dark Crawler (TDC), a program designed specifically to search for extremist content. Web crawlers, also known as data scrapers or data parsers, are instruments used by search engines to map automatically the internet and collect information on each site and webpage visited. There are different web-crawling solutions that users can employ to start the gathering of data on a specific webpage. Once the website is selected, the crawler follows recursively the present links, capturing all the materials along the way. The recovered content is then saved on a rigid disk for future analysis. Often these systems collect all the desired information but do not include advanced analytical capacities. TDC, being a personalized program, offers major flexibility compared to the standard tools. It is capable of seeking extremist content based on keywords and parameters defined by the user. While it visits each page, TDC collects all the contents and determines whether they contain extremist language and material. This approach combines different methodologies developed over the years. It can be distributed on various virtual machines and the first step consists of the definition of the task and search parameters. Each task is equipped with parameters designed to prevent the crawler ending up on web pages that are not pertinent to extremisms[54].

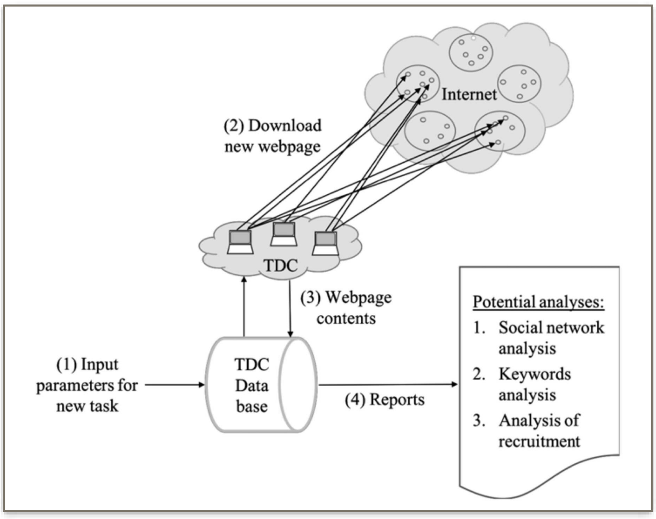


Figure 7. Overview of the Dark Crawler

Technological convergence advances two different ways to modernize conflict prevention: targeted behavioral monitoring at the individual and population levels, and automated and predictive behavioral and situational analysis. The large quantities of digital information generated by populations allow us to better comprehend the routinized behaviors through the analysis supported by artificial intelligence. Such analysis can help predict violent factors, such as violations of human rights or speech hate, and signal imminent crises through the monitoring of indicators such as armed group movements or changes in urban traffic. Advanced algorithms use the collected data to analyze individuals and activities in real time. Regarding the prevision, the most relevant AI system is the one that allows the extraction of data and forecasts from large datasets. This is linked, but separate, to the sector of big data analysis[55]. The collection of information for the analysis of conflicts and eruptions includes several types of datasets and techniques: Open source intelligence (OSINT) and social media intelligence (SOMINT), which provide data on the behaviors and intentions of populations. Human actions leave digital traces in real-time, such as transactions or

---

[54] M. Naz, S., & Shakil, M. (2019). *Searching for extremist content online using the dark crawler and sentiment analysis*. Advances in Information Technology and Computational Science, 24(16). https://doi.org/10.1108/S1521-613620190000024016

[55] Big data refers to datasets whose damsons overpass the capacities of normal software to capture, memorise, manage and analyse.

research and mobile data. Crowd-sourcing and analytical tools elaborate this information to produce tactics and strategic intelligence.

The unpredictable nature of terrorism is intrinsically connected to its goal to instill terror through attacks by unknown individuals in only apparently casual places and times. Nonetheless, an efficient counterterrorism strategy is based on the possibility of making predictions. The automated analysis of data is used to support the activities of intelligence and security services, especially through the visualization of data. Information can be collected and archived to be subsequently analyzed with the final purpose of revealing schemes and connections that can expose terrorist networks or suspected activities. Machine Learning approaches allow the interpretation and analysis of schemes normally inaccessible in large quantities of data. These approaches can include filters, analysis of relations among entities, or more sophisticated vocal or visual recognition tools.

Traditionally, the investigative techniques applied before an attack are managed by the security and intelligence services and focus on the exploration of partially uncovered plots or known suspects to find ulterior involvement or identify connections with terroristic networks; therefore, the access to data regarding a specific individual is contingent to its connection to one or more existing investigations. Recently, thanks to the development of artificial intelligence, it has become plausible to analyze the daily activities of all individuals to prevent terrorist events or identify terrorists, distinguishing the anomalies in the activities of specific subgroups. The vast quantity of digital information from every individual allows us to better understand these activities through analysis. The sources include metadata in communications and internet connection registers, but they expand also to the tracking of positions and activities, purchases, and social media movements.

Significant were the efforts, especially from the academic community, to develop models that can intercept the location and time of terroristic attacks. Among the basic approaches, it is possible to find the "*aftershock effect*", according to which the probability of an additional attack increases after each attack, a phenomenon observed also in crimes such as thefts. Other approaches predict the impact of external factors. Sophisticated models based on open-source information obtained surprising results in the prediction of other types of events, using social media-generated data and mobile apps. One example is the Early Model-Based Event Recognition using Surrogates (EMBERS), a system that continuously monitors data sources to extract emerging trends and elaborate them into a prediction of potentially disrupting societal events, such as protests and civil unrest[56].

Moreover, technological instruments might be re-proposed to identify the vulnerability to radicalization, an increasingly important field for technological enterprises in their effort to protect users. For instance, details leaked on SKYNET by the NSA, used in Pakistan in 2007, showed how quantitative methods can help predict involvement with terrorism. The system has analyzed metadata of 55 million Pakistani telephone users, classifying them into two behavioral categories: one showed a usage model similar to the one known for terrorist couriers, and the other one composed of all the other users. This model showed the capacity to reduce significantly the large population analyzed, identifying erroneously potential couriers only in 0.008% of cases. Nonetheless, it has produced numerous false positives: around 15.000 people have been erroneously identified as suspects and only 50% of real terrorist couriers were correctly individuated, demonstrating how this specific example of predictive AI for contrasting terrorism hints at the possibilities, more than providing concrete proof of feasibility. It is not realistic to expect an AI system to offer immediate solutions to complex problems.

3.4.2 Anomaly Detection Systems: Human Action Recognition and Crowd Analysis

In the last years, the recognition of human actions (HAR) aroused a growing interest in the field of computing vision, a sector that benefitted from significant progress thanks to advanced technologies and increasingly completed datasets. The possibility of recognizing human actions in different contexts opened up the way to develop algorithms capable of analyzing and interpreting human behaviors in an automated way[57]. Human action recognition systems are mainly based on the analysis of video sequences, where each frame provides visual information. Initially, these systems used basic artificial vision techniques, however, with the introduction of neural networks, in

---

[56] This system is supported by the Intelligence Advanced Research Projects Activity (IARPA) and Open Source indicators (OSI), whose goals is to predict changes at the population level using open source data feeds, such as tweets, web searches, news/blogs, economic indicators, Wikipedia, Internet traffic, and other sources. Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R. P., Saraf, P., & Wang, W. (2014). '*Beating the News' with EMBERS: Forecasting Civil Unrest using Open Source Indicators. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1799–1808. https://people.cs.vt.edu/naren/papers/kddindg1572-ramakrishnan.pdf (accessed 1 Oct. 2018).

[57] Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2022). V*ideo-based Human Action Recognition using Deep Learning: A Review*. arXiv. https://doi.org/10.48550/arXiv.2208.03775

particular, CNN and RNN[58], The recognition of actions has become more accurate and sophisticated. CNN is projected to elaborate images and can identify automatically the relevant characteristics, such as colours and shapes. In a HAR system, the CNN analyses video frames to extract useful information regarding the position and movement of objects and people. RNNs on the other hand, especially the Long Short-Term Memory (LSTM), are designed to manage sequential data. These are capable of remembering previous information and using it to interpret the context of an action. This is fundamental for the recognition of actions since several human actions are composed of sequential movements in time. Often, recognition systems combine CNN and RNN to exploit the strengths of both networks. CNN can elaborate frames to identify key characteristics, while RNN can analyze how these characteristics change over time.

Human Action Recognition is not only a field for academic research, but it also has important practical applications, especially in the context of security and counter-terrorism. These systems can be integrated into security cameras and surveillance systems to monitor the behaviors of individuals in public spaces. Through the automated analysis of videos, these systems can identify suspected behaviors or anomalous interactions that might indicate a dangerous situation, such as violent acts of mass events. When a system detects a suspected action, it can immediately send a warning to the competent authorities. This allows a quick and targeted response, increasing the possibilities to prevent serious accidents before happening. Furthermore, the analysis of historical data from HAR systems can help better understand the behavioral models. This information can be used to develop more efficient security strategies, train armed forces, and improve the administration of public events. HAR represents a paramount step in computing vision, with applications that go beyond the mere classification of movements. The ability to analyze and interpret human behavior offers powerful tools to enhance public security and contribute to the fight against terrorism.

3.4.3 Biometrics and Facial Recognition

The field of biometrics focuses on the identification and classification of individuals through physiological features and unique behaviours[59]. This recognition is often based on distinctive characteristics such as facial features and can be divided into two main types: physiological

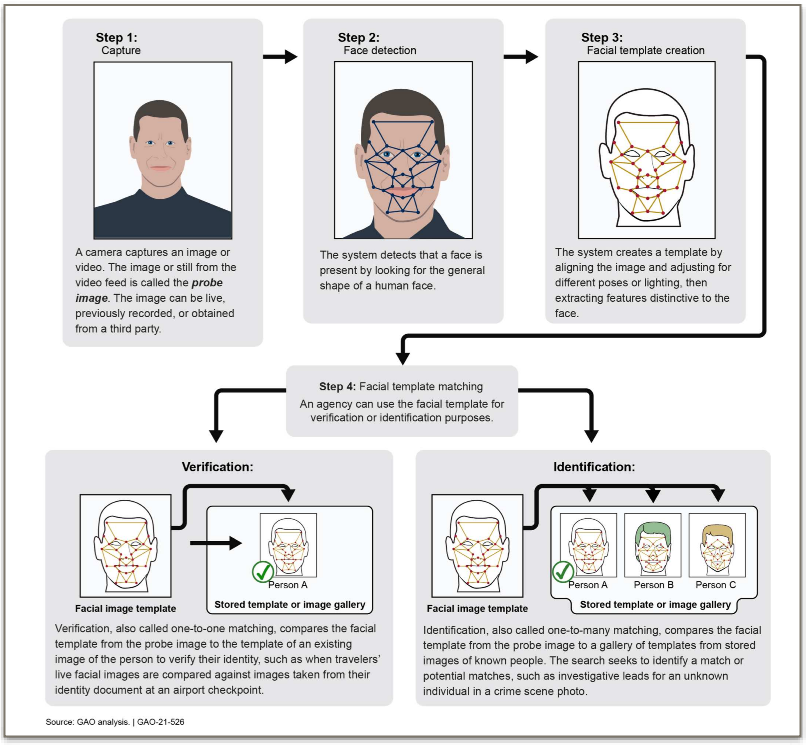
---

[58] Both CNN and RNN are Deep Learning Architectures

[59] Biometrics Institute. (n.d.). *What is biometrics? FAQs*. Retrieved October 7, 2024, from https://www.biometricsinstitute.org/what-is-biometrics/faqs/

biometrics, which deals with static physical attributes (for example digital prints or iris models, DNA, blood), and behavioral biometrics, which examines dynamic traits linked to the individual behavior (For example walking schemes, signature). The applications of Biometrics are wide and varying, ranging from healthcare to immigration and security.

In the context of counter-terrorism, among emerging technologies, biometrics systems, and in particular Facial Recognition (FR) as the most diffuse and promising technology, have been revealed to be efficient tools against extremist movements and groups. Due to its ability to quickly identify suspected individuals in crowded public spaces, Facial Recognition has found application in numerous sectors, among which surveillance, airport security, and armed forces. Thanks to the increasing availability of surveillance cameras in public spaces, it is possible to monitor vast areas and track potentially dangerous individuals with extreme precision. The use of this technology expands not only to the prevention of ordinary crimes but also in the fight against terrorism, allowing the fast identification of suspects, minimizing at the same time the impact on people's daily lives.


Step 1: Capture
Step 2: Face detection
Step 3: Facial template creation
Step 4: Facial template matching
Verification:
Identification:
Facial image template
Stored template or image gallery
Person A
Facial image template
Person A
Person B
Person C
Stored template or image gallery

A camera captures an image or video. The image or still from the video feed is called the ***probe image***. The image can be live, previously recorded, or obtained from a third party.

The system detects that a face is present by looking for the general shape of a human face.

The system creates a template by aligning the image and adjusting for different poses or lighting, then extracting features distinctive to the face.

An agency can use the facial template for verification or identification purposes.

Verification, also called one-to-one matching, compares the facial template from the probe image to the template of an existing image of the person to verify their identity, such as when travelers' live facial images are compared against images taken from their identity document at an airport checkpoint.

Identification, also called one-to-many matching, compares the facial template from the probe image to a gallery of templates from stored images of known people. The search seeks to identify a match or potential matches, such as investigative leads for an unknown individual in a crime scene photo.

Source: GAO analysis. | GAO-21-526

Figure 8. Face recognition technology using Datasets[60]

---

[60] Picture taken from the Report to Congressional Requesters, FACIAL RECOGNITION TECHNOLOGY Current and Planned Uses by Federal Agencies, August 2021, United States Government Accountability Office. https://www.gao.gov/assets/gao-21-526.pdf

Facial recognition systems operate by capturing images of people through cameras and analyzing them in real time. The images are then compared to a database of already-known individuals, such as suspected terrorists or wanted criminals, to identify potential threats. This technology functions through advanced algorithms of Deep Learning that create biometric models, or "templates", based on unique facial features such as the distance between the eyes, the shape of the nose or the countering of the face. When a correspondence is detected with a face already present in the database, the authorities are immediately warned to intervene. Facial Recognition systems are particularly useful in highly frequented and public spaces such as airports, train stations, shopping centers, and city squares. In these contexts, surveillance cameras are already present and can be integrated with facial recognition software, making the monitoring of individuals constant and automated. The tracking of terrorists through FR technologies happens mostly thanks to the use of a Watchlist, namely a database that contains the photos and biometric information of individuals known for their connections to terrorist groups. These systems are already implemented in various cities and critical infrastructures all over the world, as part of a proactive defense system against terrorism.

One of the benefits of facial recognition is its ability to operate in real-time without the necessity to interact physically with the monitored subject, making it a discreet but efficient method. This allows authorities to carry out security operations without significant inconveniences for the public. However, some limitations to this technology exist. The environmental conditions, lighting, and perspectives of cameras can influence the quality of the pictures, reaching the precision of the system. Additionally, individuals partly covered or modified through plastic surgery can elude the system. Despite these limits, the constant development of this technology is progressively improving the capabilities to face these challenges.

3.4.4 AI against terrorism propaganda and Natural Language Processing

Social media platforms offer infinite opportunities for the spread of terrorist materials, amplifying the recruitment pool with a strategy similar to the ones used by companies for their communication campaigns. The use of images and videos allows to hit a wide range of individuals and influence recruits more subtly. The capacity of propaganda is highlighted by the ease which such content can become viral on platforms such as YouTube or Facebook. One particularly significant example is the video game produced by ISIS, Sail al-Sawarem, designed to attract young users through violent themes. The extreme right also exploited gaming platforms like Call of Duty and Roblox to spread extremist messages and recruit members. This continuing evolution of online terrorist activity represents a growing challenge to armed forces, especially given the mole of contents uploaded every day on the internet. Only between 2015 and 2018, the EU Internet Referral Unit reported beyond 50.000 terrorist content, whereas in 2021 Twitter suspended more than 1.8 million accounts linked to the promotion of violence. In 2022 Facebook deleted more than 56 million terrorist content[61].

To face this threat, the use of advanced technologies such as artificial intelligence is becoming fundamental. Armed forces and Hosting services providers use AI to monitor and analyze large quantities of online data, detecting terrorist content automatically and more efficiently compared to traditional human resources. AI offers a series of sophisticated instruments, such as Natural Language Processing, which analyses multilingual texts and simplifies the identification of contents linked to extremisms. These tools not only identify dangerous contents but allow them to respond promptly to removal requests, respecting regulations such as the European "*one-hour rule*"[62], which imposes the removal of contents in one hour. Furthermore, AI provided support to moderators through techniques of "explicability", which furnish reasoning and explanation about the obtained results, making it easier to understand informed decisions on the removal of violent materials.

Natural Language Processing (NLP) is one of the artificial intelligence systems that most have developed over time. Thanks to artificial intelligence techniques such as Machine and Deep learning, NLP finds numerous applications. This NLP acronym or the expression elaboration of the natural language, means algorithms of artificial intelligence capable of analyzing, representing, and therefore understanding the natural language (NL)[63]. The purposes might vary from the comprehension of the content to the translation, until the production of texts autonomously starting from data or documents that are provided as input. Different from programming languages, which

---

[61] Netlaw.bg. (n.d.). *Artificial intelligence to counter cyber-terrorism*. Retrieved October 7, 2024, from https://www.netlaw.bg/p/p/a/paper-ai-to-counter-cyber-terrorism-3657.pdf

[62] European Union. (2021). Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online. Official Journal of the European Union, L 172, 79–109. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2021:172:FULL&from=EN

[63] Osservatori Digital Innovation. (2023). *Natural Language Processing (NLP): Come funziona l'elaborazione del linguaggio naturale*. Osservatori.net. https://blog.osservatori.net/it_it/natural-language-processing-nlp-come-funziona-lelaborazione-del-linguaggio-naturale

follow precise rules and are easily interpretable by machines, our language is not easily representable. Since humans interact with machines daily, it is necessary to create systems capable of understanding and answering humans. Here the computational linguistics comes into the game, namely the study of informatics systems for the analysis and elaboration of natural language, and it focuses on the functioning of NL in a way that it then elaborates program executable by machines. NLP deals mainly with texts, meant as a sequences of words that express one or more messages (ex. Pages, posts, tweets..). Speech processing (or vocal recognition) is a different field. The dialogue between man and machine includes different aspects, such as phonetics, phonology, morphology, syntax, semantics, and discourse as a whole, and as a consequence, several of the techniques developed, that will be briefly described below.

A. Text Mining, Tokenization, Stemming and Lemmisation

Text mining is a process of extraction of information and knowledge from data from unstructured texts. Text mining is a multidisciplinary field based on data mining, machine learning, information retrieval, and computational and statistical linguistics. Part of this process is information extraction, natural language processing, the classification of the text, and content analysis. This pre-elaboration is particularly relevant because it reduces significantly the dimension of the texts in inputs. Text mining is used to extract useful information, knowledge, or patterns from the unstructured document, it converts words and phrases into numerical values which may be linked to structured information in the database.

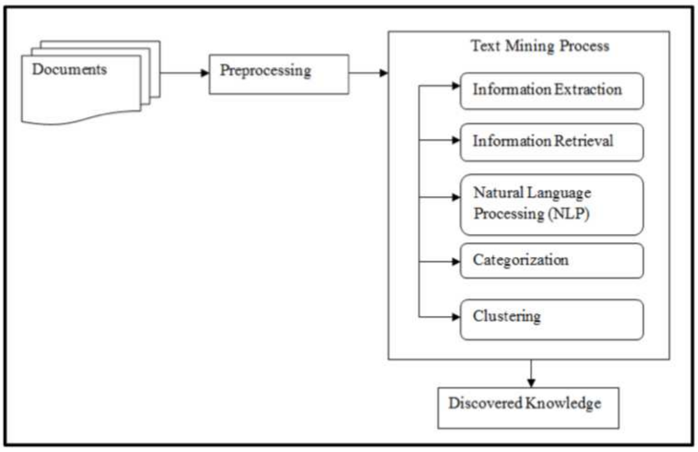


Figure 9. Text Mining Steps

The most important action in this phase is the process of Tokenization, a process of text, words, symbols (or other significant elements for the analysis) division into significant units, called tokens, which is a fundamental step of pre-elaboration in almost all NLP applications. This process is crucial in both linguistics and informatics, where it represents a part of the lexical analysis. Usually, tokenization happens at the word level, but it is not always easy to define what a word is. Its main advantage is its capacity to identify significant words or concepts inside a text, making it useful for other phases of elaboration, such the semantic analysis. When working with textual data, this process is indispensable. Textual data are, at the beginning, an array of characters, however, to execute efficient analysis, it is necessary to transform them into significant units. This step allows systems to elaborate correctly the documents[64].

If we consider the input "*Data Mining is the process to extract hidden predictive information from database and transform it into understandable structure for future use*." The tokenisation result would be "*Data, Mining, is, the, process, to, extract, hidden, predictive, information, from, database, and, transform, it, into, understandable, structure, for, future, use*".

There are several instruments that can be part of the process: NIpdotnet Tokeniser, which is an open source tool that manages efficiently complex texts and produce accurate and consistent documents; Mila Tokenizer, which focuses on the segmentation of words in different language than English; Pattern Word Tokenizer, which is particularly useful for those who works on data mining and text analysis applications with complex natural languages; or MBSP Word Tokenizer, which deals with the advanced injustice elaboration, optimal for projects that require an in-depth analysis.

The immediate following steps in text elaboration are the processes of Stemming and Lemmatization, two fundamental techniques in NLP that aim to reduce the world to its basic forms, even though they do it in different ways and specific gaols. Stemming is the process of reducing one word to its root form or base form. This approach allows to normalise the variations of a word, facilitating the research and improving the comprehension of the text. For instance, for the word "like", the steaming process will produce variations such as "likes", "liked", "likely" and "linking". In the same way as the word "chocolate", the variants will include "chocolates", "chocolatey" and "choco". The result is that all these words are reduced to the root form, facilitating the management

---

[64] Mohan, V., & B. Siddhartha. (2018). *Text mining: Open source tokenization tools - An analysis*.

and analysis of language. The importance of stemming resides in its capacity to ease the elaboration of textual data, shortening the research times and normalizing the sentence, which is particularly useful in applications such as research engines and information recovery systems[65].

Lemmatization, on the other hand, is a more advanced process that involves the assembling of reflexive parts of a word in a way that they can be recognized as a single element, known as limm or lexical form of the word. Different from stemming, lemmatization takes into consideration the meaning and context, connecting words with similar meanings to one unique form. For example, for the word “rocks”, the limm would be “rock”, and for the word “better” it would be “good”. This approach makes lemmatization useful in sectors where the meaning is crucial and the correct interpretation of terms is paramount[66].

B. Part of Speech Tagging

Subsequently, the Part of Speech (POS) Tagging comes, an important activity in the field of NLP that consists of assigning to each word of a text, an appraise morphosyntactic label, based on the context in which it occurs (noun, adjective, verb, etc…). This operation is crucial in some applications of NLP, such as the analysis syntax, translation and recover of information, vocal synthesis. The basis of POS tagging is that many words can be ambiguous regarding their grammatical role. However, in the majority of phrases, it is possible to clear a word correctly keeping in mind the surrounding context. For example considering the sentence:

“*The ball was shot by the man* ”

In this case the term “shot” is interpreted as a past participle because preceded by the auxiliary “was”. However, if the context was different, as in “the girl took a shot”, the word “shot” would a have a completely different meaning.

---

[65] B, S., Siddhartha. (2021). *An interpretation of lemmatization and stemming in natural language processing*. Retrieved from https://www.researchgate.net/profile/Siddhartha-B-S/publication/348306833_An_Interpretation_of_Lemmatization_and_Stemming_in_Natural_Language_Processing/links/6048467f299bf1e078696a3a/An-Interpretation-of-Lemmatization-and-Stemming-in-Natural-Language-Processing.pdf

[66] ibid.,

There are three main approaches to POS tagging, that can be classifies based on the type of knowledge used: linguistic approach, based on an array of rules written by linguists, that can vary from hundreds to thousands, and require a notable investment of time for their creation; statistic approach, most common one today, consist of constructing a language statistical model, by using co-occurrence frequencies of diverse linguistic sequences to clear out the words; lastly, the machine learning, which includes more sophisticated systems. With the help of these approaches, POS tagging systems can achieve a very high accuracy, surpassing the 96% in statistic taggers[67].

One additional technique, in the context of NLP, is Named Entity Recognition (NER), which represents a fundamental step in the extraction of data from texts written in natural language. This technique focuses on the identification and classification of people's names, organizations, localities, and other significant entities inside a text. The concept of Named Entities emerged in the context of the conferences on text comprehension in the 1990s, where the main goal was the extraction of information from unstructured content, such as newspaper articles. With the progress of NL elaboration technologies, it was clear that Entity Recognition was a crucial element of it. NER is essential for the automatic categorization of information since it allows to classify and organize data in defined hierarchies, facilitating access and research. This capacity of recognizing and classifying entities in complex texts improves significantly the accuracy of recovery systems and data analysis. This process is strictly connected to the above-mentioned processes of tokenization, stemming, and lemmatisation, which elaborate and simplify texts[68].

C. Sentiment Analysis

The use of keywords and expressions represents the very first step to individuate large-scale models in online extremist content. However, the use of single keywords can lead to misleading interpretations. For example, if on a webpage the words "arm" and "control" are found close to each other, it could be automatic thinking that the page deals with arms control, a topic that is not necessarily connected to extremisms. Nevertheless, if the same words are found in a sequestration context, the content could be relevant for an analysis regarding terrorism. In this sense, keywords

---

[67] Meyer, D. (2000). *Generalized additive models: An introduction with R*. Retrieved from https://link.springer.com/content/pdf/10.1023/A:1007673816718.pdf

[68] Liu, B. (2021). *Deep learning for natural language processing: A comprehensive guide to understanding deep learning for natural language processing*. pp. 810-828 Retrieved from https://link.springer.com/content/pdf/10.1007/978-981-16-3346-1.pdf

can give a direction regarding the content of the page, but cannot precisely define the context. To fully comprehend the text, some more computational tools are necessary, such as sentiment analysis.

Sentiments analysis, or "*opinion mining*",[69] is a branch of informatics science that deals with evaluating the opinions present inside a text organizing data in categories and assigning to each text a positive, negative, or neutral polarity value. This type of analysis allows a deeper comprehension of texts, distinguishing relevant and irrelevant materials. It is particularly useful in the context of terrorism studies because the quantity of online opinion data is increasing exponentially. This type of software allows us to deal with complex research issues, and identify and classify opinions expressed inside texts. The process usually consists of two separate phases: the text is first divided into sections to identify the objective and subjective content, and then the subjective content is classified as positive, neutral, or negative. Nonetheless, sentiment analysis has limits: human beings, for example, do not always agree on the sentiment expressed in the document, especially when expressed with irony or sarcasm. Consequently, an absolute precision can not be guaranteed. One of the most common tools used in the sentiment analysis applied to extremism and terrorism studies, is SentiStrenght, based on a lexical approach, namely the analysis of models and meaning associated with language. This software was designed to analyze short informal texts, but it also includes functionalities that allow us to deal with lower ones. It assigns positive, neutral, or negative values to the words of the text, the value that can be emphasized by "booster words" that amplify the sentiment through negations, punctuation, and other elements. One of its characteristics is the capacity to analyze the sentiment surrounding a specific keyword. For example, the sentence "I like dogs but I hate cats" can be analyzed for the positive sentiment surrounding "dogs" or for the negative one surrounding "cats". Words next to a specific keyword are compared with a sentiment dictionary and the resulting values contribute to the final score of the text. This process can be repeated more times to analyze different keywords and the relative sentiments. In the end, the scores are mediated to obtain a single sentiment value for the entire text.

Over the last years, we witnessed a significant change in the way online communities are studied, going from a manual analysis of specific content to the growing adoption of algorithmic techniques

---

[69] M. Naz, S., & Shakil, M. (2019). *Searching for extremist content online using the dark crawler and sentiment analysis*. Advances in Information Technology and Computational Science, 24(16). https://doi.org/10.1108/S1521-613620190000024016

to execute similar tasks at a wider scale. This evolution reflects the rise in the use of automated analytical approaches in sectors like criminology and criminal justice, and it is part of the phenomenon called “big data”, namely the availability of enormous quantities of online data. In the context of terrorism, many researchers have embraced using sentiment analysis and machine learning techniques to identify and analyze content with a wide range, allowing them to examine efficiently and in depth the way terrorists use the internet.

D. Dependency Parsing

The Dependency Parsing (DP) represents an essential approach in the analysis of the natural language and it covers a particular importance when applied to complex contexts such as the monitoring and analysis of speeches and materials linked to terrorism[70]. This method focuses on the dependency relations between the worlds of a sentence, allowing the AI system to understand the syntactic structure without referring to syntactic rules. By using dependency analysis, AI systems can elaborate and interpret extremely difficult texts, identifying the grammatical relations between words. In terrorism scenarios, where communications can be ambiguous or full of meaning, dependence parsing permits to extract critical information, such as the intentions and reasons of the authors of the radical messages. Relations among subjects, verbs, and objects are mapped, facilitating a deeper semantic comprehension of communications. A significant advantage of DP, especially in languages with rich morphologic structures and flexible word orders, is its capacity to adapt to sentences formulated in nonconventional ways. In this way, AI can face texts coming from diverse cultures and linguistic contexts, identifying the underlining structures that might otherwise escape a more superficial analysis. Furthermore, the approach based on dependency helps intercept signs of recruitment and extremist rhetoric, offering tools to analyze messages in a way that is then possible to identify correlations among argumentations used by these actors. AI, using DP algorithms, can extract dynamically useful information to prevent and respond to terrorism.

An AI system or DP algorithm, after being trained on big quantities of linguistic data by experts, where dependencies between words have already been defined, uses statistic models or linguistic rules to choose the headword of the sentence, normally the main verb that governs the action of the sentence, and that guides the syntactic structure of the entire phrase, and connects to its dependents

---

[70] Jurafsky, D., & Martin, J. H. (2019). *Speech and language processing*. Retrieved from https://web.stanford.edu/~jurafsky/slp3/old_oct19/15.pdf

(the other elements of the sentence). After the identification of words and relations, each word is treated as a "node" and some arcs are drawn among nodes to represent the grammatical relations. These arcs are labeled with their grammatical functions like subject, object, modifiers, etc… Subsequently, a dependency tree is generated to provide a clear and structured representation of grammatical relations between words. Once this step is completed, the AI can use it for a variety of practical applications such as meaning comprehension, automatic translation, detection of violent contexts, etc… Practically, it allows the AI and NLP algorithms to learn the function and role of each word, facilitating the other tasks of language processing.

Lets take the following sentence as example of Dependency Parsing:

"*the police arrested the suspect who planned the attack*"

The dependency tree breakdown would be:

1. **arrested** (root of the sentence)
   - **police** (subject, depends on "arrested")
     - **the** (determiner, depends on "police")
   - **suspect** (object, depends on "arrested")
     - **the** (determiner, depends on "suspect")
     - **who** (relative pronoun, depends on "suspect")
       - **planned** (verb, depends on "who")
         - **attack** (object, depends on "planned")
           - **the** (determiner, depends on "attack")

Graphical Dependency tree:

```
arrested
    /     \
 police  suspect
  |        |   \
 the    the   who
                |
          planned
                |
            attack
                |
              the
```

E. Word Embeddings

Word embedding is a nonsupervised machine-learning technique used to learn a vectorial representation of words inside the body of a text. This methodology is particularly relevant in the context of Natural Language Processing for the analysis and the contrast to terrorism, as it allows to elaborate more precisely texts to individuate semantic relations among words that often appear together, providing an efficient analysis of suspected or extremist content. In the context of counterterrorism, Word embedding, such as the model GloVe (Global Vectors for word representation), is used to create numerical representations of words present in the body of documents (speeches, posts on social media, or intercepted communications). The algorithm generates a matrix of co-occurrence words starting from the original malign texts, applying a model of logarithmic regression to esteem the vectorial coordinates of each word. These numerical representations allow us to measure the semantic proximity among words, and therefore to identify terms that often appear in association, for instance, the word “attack”, “explosive” and “vendetta”.

One of the most useful aspects of word embedding in the anti-terrorism field is that it allows to identify subtle semantic relations between terms that might not be detected by traditional analysis. This allows the detection of suspected language patterns also when extremist contests are encrypted or hidden. The algorithm GloVe has been applied to the body of speeches and documents, with parameters set to represent each word with vectors at 100 dimensions. The context of words was defined as an interval of 10 previous and 10 following words, useful to determine which terms tend to co-occur in a suspected conversation or online discussion regarding terrorism. This type of analysis can be segmented for country, temporal period, or specific events, permitting to follow the evolution of the terroristic language in different geographical areas or determined historical moments, for instance after a terroristic attack or a security operation[71].

By applying these systems of word embedding, security agencies can obtain a sophisticated representation adaptable to the language used by extremist groups, individuating patterns or new emerging trends. Moreover, not only it make the monitoring more efficient, but it allows to anticipate potential threats through the automated analysis and the comparison between the most commonly used words. The way each word is transformed to a numerical vector is similar to process of automatic translation described at the point 1.2.5 of the thesis, when describe the work "*Attention is all you need* " by Ashish Vaswani.

F. The case of radical conversations by ultra-right groups in France[72]

The society we live in is characterized by a digital dimension that transforms radically the way we live, think, and communicate. Today's internet is a fertile ground for extremist and controversial discourses, which are designed to become viral, exploring the specificities of the digital world to spread out rapidly. Extremist groups exploit online violence such as humiliation, cyberstalking, and hatred, through various digital platforms. As a response to this issue, diverse countries have introduced legislation and institutions to fight against cyber hatred. This type of violent content, often towards the youth, uses emotional manipulation strategies. The study carried out by Séraphin Alava, Nawel Chaouni, Nagem Rasha, proposed to examine how digital technology favors radical recruitment. 5.215 sentences coming from radical groups have been analyzed to understand their methods of recruitment, which range from indoctrination to the attraction of new members. The research was carried out using sentences collected in France, Belgium Switzerland, and Quebec over two years, from websites, forums, social media, etc.. The methodology used considers also the

---

[71] Chaturvedi, A., & Mukherjee, A. (2020). *Cross-national analysis of global security discourse using word embeddings*. Retrieved from https://preprints.apsanet.org/engage/api-gateway/apsa/assets/orp/resource/item/5ecd3ff1fd965c0018b47287/original/cross-national-analysis-of-global-security-discourse-using-word-embeddings.pdf

[72] Séraphin Alava, Nawel Chaouni, Nagem Rasha. *Extreme digital content: study of radical conversations by ultra-right groups in France*. 2024. https://hal.science/hal-04561644/document

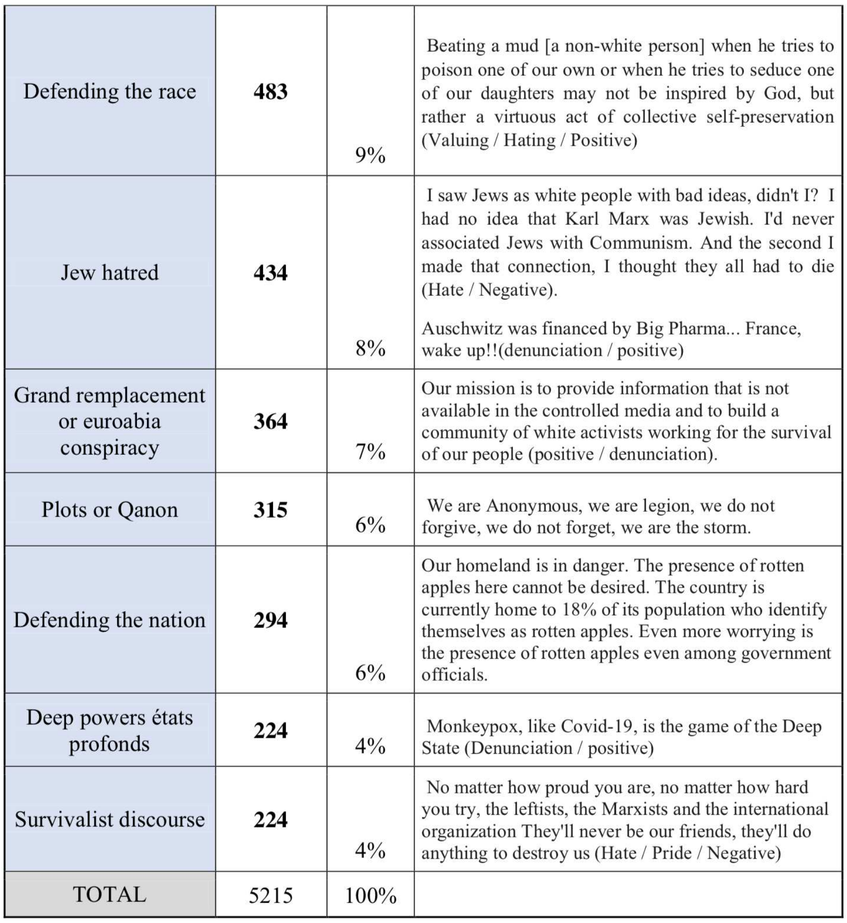
appropriation level of the radical speeches, distinguishing between extremist leaders' discourses and those adopted by already radicalized individuals. The material was then classified based on its nature, separating offensive and violent content, such as insults, incitements to violence, racism, and sexism. By using tools of semantic analysis and text mining, textual classification and thematic analysis new carried out, aiming to identify recurrent topics and understand the motivations that guide these conversations, with the final goal of studying the main factors of online radicalization. In addition, sentiment analysis and discourse strategies analysis were conducted to better comprehend the persuasion mechanisms used by these groups. Finally, the last step of this research was integrating the result with an online ethnography analysis, which requires the participant observation to immerse in the studied context, identifying the social dynamics and interactions that favor radicalization.

A radical discourse can be categorized based on its radical intensity, which depends on its nature and emphasis on the elements the discourse is composed of. This categorization can include levels like disinformation, hate speeches, and incitement to violence. The thematic analysis of these speeches can help understand the intersections of radical logic and ideology movements. Online radicalization presents specificities compared to the offline one for ultra-right groups. Firstly, the history of the digitalization process of ultra-right actions has been characterized by the rapid adoption of digital technologies to spread out their ideologies. This digitalization has amplified the scale and sped up the diffusion of extremist messages. Moreover, the variety of topics discussed online is notably greater than offline radicalization, which usually focuses on specific national matters.

The extremist actions can vary and differ based on their ideology, motivation, internal organization, external support, actors involved, goals, and resources. Online violent discourse is facilitated by the use of pseudonyms, anonymity, and persuasion efforts that allow groups to attract new followers. Social media, online forums, and instant messaging apps have become essential communication channels. These speeches receive transformations and transpositions over time. For example, antisemitism is gradually being subtitled by hatred towards Muslims, which represents the first element of respect, with 945 analyzed sentences out of 5.215. This evolution reflects a change in the targets perceived as dangerous. Similarly, the defense of the homeland is transforming into hostility towards migrants, especially towards non-western ones. This highlights xenophobia. This extreme ideology reinforced the condition that violence is necessary to preserve the race supremacy.

By analyzing data it was possible to see how the ultra-right bloc is heterogeneous, comprehending a variety of ideologies and targets, from identitarian moments to white supremacist groups. This diversity complicates the analysis of the discourses, since the specifies for each group need to be taken into consideration. Data extraction led to regular visits to platforms such as Facebook, Telegram, and Twitter and the identification of 446 key terms, maintaining only the sentences that did to contain short answers, eliminating sentences that repeated the same ideas as previous ones, to concentrate the ideas and facilitate the analysis.

	Number of sentences	**% by category**	**Example sentences**
Hatred of Muslims	**945**	18%	Your daughter will wear a headscarf (Hijab), your son will wear a beard and your daughter will marry a man with a beard. They reproduce faster and faster (Negative/Hate). A practicing Muslim who believes in the teachings of the Koran cannot be a loyal citizen in France (Hate/Negative) No one minds Hindus, Sikhs, Jews, Jehovists or anyone else of any other faith. Racist Muslims are not here to integrate and get along with others, they want to dominate, kill infidels and think that raping white children is their spoils of war in a conquered country. Islam is evil!(Hate / Negative)
Incel or antifeminist remarks	**714**	14%	I suffer from depression, anxiety, disorders and I'm not sure what gender I am anymore. I'm unemployed and in debt. I waste all my money on weed and Star Wars memorabilia. I have virtually no control over my impulses. Now, here's how I'm going to fix this country (Denunciation / Negative / Positive) You like men because you're gay. I like men because I consider women inferior. We're not the same (Hate / Positive)
Hate for feminists	**644**	12%	When a man can't possess a woman, he's bitter. When a man can use a woman, he's delighted (Masculinist/Human/positive) Feminism is a cancer. Repeat after me *fuck feminism* men are not slaves (Masculinist / Negative)
Antisciences	**574**	11%	Remember when they told you to wear your m@sk and that it would protect you from C0VID? They lied!!! (antisciences / negative)



Defending the race	**483**	9%	Beating a mud [a non-white person] when he tries to poison one of our own or when he tries to seduce one of our daughters may not be inspired by God, but rather a virtuous act of collective self-preservation (Valuing / Hating / Positive)
Jew hatred	**434**	8%	I saw Jews as white people with bad ideas, didn't I? I had no idea that Karl Marx was Jewish. I'd never associated Jews with Communism. And the second I made that connection, I thought they all had to die (Hate / Negative). Auschwitz was financed by Big Pharma... France, wake up!!(denunciation / positive)
Grand remplacement or euroabia conspiracy	**364**	7%	Our mission is to provide information that is not available in the controlled media and to build a community of white activists working for the survival of our people (positive / denunciation).
Plots or Qanon	**315**	6%	We are Anonymous, we are legion, we do not forgive, we do not forget, we are the storm.
Defending the nation	**294**	6%	Our homeland is in danger. The presence of rotten apples here cannot be desired. The country is currently home to 18% of its population who identify themselves as rotten apples. Even more worrying is the presence of rotten apples even among government officials.
Deep powers états profonds	**224**	4%	Monkeypox, like Covid-19, is the game of the Deep State (Denunciation / positive)
Survivalist discourse	**224**	4%	No matter how proud you are, no matter how hard you try, the leftists, the Marxists and the international organization They'll never be our friends, they'll do anything to destroy us (Hate / Pride / Negative)
TOTAL	5215	100%	

Figure 10. Results of the research on ultra-right online conversations and discourses[73]

From a semantic perspective, the ultra-right discourse results mainly offensive and negative, using denigratory and offensive terms to stigmatize determined communities and ethnic or religious groups. Numerous war metaphors are used to communicate the idea of a conflict between "Us" (group members) and "Them". This discourse tends to dehumanize the individuals belonging to groups that are considered dangerous or simply diverse and touches often on topics of exclusion, ethnic purity, and racial supremacy. They are generally opposed to immigration and multiculturalism. Some speeches incorporate conspiracy theories to explain social and political problems. In terms of sentiments, the rhetoric coming from these groups aims to arouse reactions of fear and anxiety in the public and exploit the apprehensions linked to insecurity, loss of cultural identity, and the threat of outsiders. They evoke hatred and anger to strengthen cohesion. Conceptually, these speeches are characterized by conflictual duality, oscillating between victory and victimhood narratives.

---

[73] Séraphin Alava, Nawel Chaouni, Nagem Rasha. *Extreme digital content: study of radical conversations by ultra-right groups in France*. 2024. https://hal.science/hal-04561644/document

Concluding, this study provided a significant example of the dynamics of online conversations among extremists, underlining the importance of artificial intelligence in the analysis of complex phenomena and demonstrating how AI can extract patterns from enormous quantities of data. This approach emphasizes that the necessity of monitoring and contrasting such discourse to prevent the escalation of hatred and violence in society, leading to a wider recruitment of individuals.

3.4.5 AI contrasting Money Laundering and Terrorism Financing

The use of artificial intelligence is becoming a key instrument also in the contrast to financial crimes such as money laundering and terrorism financing. Generally speaking, AI, based on automated learning algorithms, can analyze efficiently enormous quantities of financial data, identifying models and anomalies that might be difficult to detect with traditional methods. The expression money laundering means the process of legalisation of funds obtained illegally, often transforming them into apparently legal ones. Often it is a process divided into more phases, that include various financial transactions aiming at covering traces. Terrorism financing, on the other hand, is the derogation of financial resources, or equivalents, to sustain terroristic activities. Criminals use a variety of money laundering techniques to make the tracking of money more difficult. Among these, there is the division of large amounts of money into several smaller transactions, or the investment in real estate, artworks, or commercial activities. Money laundering (ML) and terrorism financing (TF) generate different negative effects. At the social level, they lead to a distortion of economic competition, an increase in criminality, and a loss of trust in financial institutions. The economic effects comprehend the destabilization of financial systems and loss of capital, whereas, from the security perspective, these phenomena might be devastating since TF allows extremist groups to pursue goals that mine societal stability.

Artificial intelligence can revolutionize the fight against money laundering, by analysing vast sets of financial data. Traditional methods are often limited in the scale and complexity of data they can scan. AI using advanced automated learning algorithms, is capable of processing enormous quantities of information and identifying subtle models that can indicate suspected activities. One fundamental area where AI can be used is the analysis of clients' behaviors. By monitoring and scanning the interactions of clients with financial services, systems based on AI can detect unusual models that might indicate an attempt at money laundering or terrorism financing. With the increasing use of blockchain technology, AI can also be deployed to monitor transactions inside networks based on blockchain, making money laundering more difficult[74].

In the context of the growing complexity of financial crimes, the collaboration between the public and private sectors is vital for the development and implementation of these advanced technologies to prevent illicit activities. Financial institutions, often in collaboration with governmental agencies and regulating institutions, are adopting innovative technological solutions, for instance those offered by SAS Anti-Money Laundering (AML)[75] and IBN Financial Crimes Insights, to improve their capacities of detecting and responding to emerging threats. This relation not only favours the adoption of more efficient systems but allows a greater sharing of information, a better comprehension of crime models, and more conformity to the law.

These specialized platforms, SAS AML and IBN, offer advanced solutions based on artificial intelligence and machine learning to face these challenges. SAS AML is an analytical solution designed to support financial institutions in the prevention and monitoring of money laundering and terrorism financing. It provides tools for scanning transactions in real time, generating conformity reports, and facilitating internal investigations. It is a highly adaptable platform that permits institutions to personalize their risk models based on their specific needs. It collects data from different sources, including transaction registries, client profiles, and public information. Through the analysis of this information, it identifies schemes and anomalies that might indicate illicit activities. It functions with customizable risk models to evaluate the probability of suspected behaviors, models based on key variables such as transaction history, client profile, and geo-localization. IBM on the other hand, is a platform that integrates artificial intelligence and machine learning to improve financial security and therefore combat against money laundering and terrorism financing. It uses machine learning algorithms to analyze data and identify behavioral schemes. The predictive models can recognize normal and abnormal behaviors and foresee potentially dangerous transactions based on historical data and risk profiles.

---

[74] Agnieszka, P., & Marcin, M. (2021). *The role of artificial intelligence in counteracting money laundering and financing of terrorism*. ASEJ - Academic and Scientific Journal, 12(1), 23-35. https://asej.eu/index.php/asej/article/download/779/802/1226

[75] SAS Institute Inc. (n.d.). *Anti-money laundering*. https://www.sas.com/en_sg/software/anti-money-laundering.html

3.4.6 AI as a tool of Hybrid Warfare

> *"It is imperative to keep an eye on new technologies and their potential for future development and disruption, and to analyse these developments with regard to their relevance in a hybrid warfare context. Their relationships must be understood before their implications become manifest in the context of hybrid warfare. In this regard, the technological revolution requires orchestration. This should not be left primarily to potential hybrid challengers"*[76]

The use of Artificial intelligence in wars has changed the concepts of AI itself. Conventional war methods have been substituted by modern equipment and techniques. Since the era of the atomic bomb, powerful states have avoided using conventional military power against each other. The method since then adopted, is the one of Asymmetric or Hybrid Warfare (HW). As AI, Hybrid warfare does not have a globally known definition: actors in the academic and military fields use this expression in different ways and with different meanings. The efficacy of this type of war resides in the use of guerrilla tactics by nonstate actors and in the use of paramilitary forces and deniable state actors, to achieve the political goal without overpassing the threshold. The variations of the definition are due to the fact that the concept is deduced from the observation of the enemy. With the evolution of warfare methods and techniques, also the concept is in a state of evolution. Battlefields and war techniques are becoming more and more blurred and ambiguous. Technologies such as AI have drastically influenced new skills and updated weaponry. The model of hybrid warfare combined with AI has been demonstrated to be revolutionary. Some defines AI as Key enabler of Hybrid Warfare[77].

Technological modernity had an impact on all the main aspects of hybrid warfare: synergy, ambiguity, asymmetry, and disruptive innovations, and it can be used to imitate, influence, and change the behaviors and strategies of groups, transforming the social and economic consequences

---

[76] Johann Schmid and Ralph Thiele, "*Hybrid Warfare – Orchestrating the Technology Revolution*". In Robert Ondrejcsak & Tyler H. Lippert (Eds.), STRATPOL. NATO at 70: Outline of the Alliance today and tomorrow, Special Edition of Panorama of Global Security Environment 2019, Bratislava December 2019, 211–225, https://www.stratpol.sk/wp-content/uploads/2019/12/panorama_2019_ebook.pdf

[77] RALPH THIELE, *Artificial Intelligence – A key enabler of hybrid warfare*. March 2020, Hybrid CoE Working Paper 6, COI STRATEGY & DEFENCE. https://www.hybridcoe.fi/wp-content/uploads/2020/07/WP-6_2020_rgb-1.pdf

of the conflict. AI can analyze big quantities of data coming from social media, communications, and people's movements, to individuate collective behavioral schemes. This can include the way people react to specific events or messages. Once these schemes are individuated, AI can simulate or replicate such behaviors strategically. For instance, in a situation of Hybrid Warfare, it can be used to generate content or actions (fake news on social media) that seem to be authentic, to imitate behaviors of real groups, and to influence the public discourse. Through personalization and micro-targeting, AI can spread information, propaganda, or disinformation to certain groups, based on what these groups already believe or fear. This can lead to mass psychological manipulation, influencing political opinions, social behaviors, and economic decisions of entire population segments. One practical example could be to use misinformation to create panic among communities, or to divide the population of a country on crucial questions during a conflict. The final goal of the use of AI in this context, is to modify the behaviors of people or groups, leading them to act in ways that are beneficial for the "attacker". This might mean pushing a group to protest against its government or distrusting one's allies. AI can do what is described above, monitoring in real-time the reactions to such actions and adapting its strategies to maximize efficiency. By changing group behaviors at a large scale, AI can have a significant impact on economic and social balances in a nation. For instance, if an entire population starts fearing its safety because of a well-planned disinformation campaign, this might destabilize the economy, bringing panic and social disorders to markets, or create internal societal divisions, weakening social cohesion and facilitating military actions and policies by the adversaries.

In contrast to terrorism, AI can play a crucial role in imitating, influencing, and modifying the behaviors of extremist groups. AI, as mentioned also in other applications, can analyze terrorists' activities, foresee their moves, and intervene strategically, both infiltrating their networks and altering their operations with false information. Furthermore, it allows identifying and direct de-radicalization campaigns towards individuals, contrasting terrorist propaganda.

One significant innovation in the field of modern warfare is represented by Autonomous Weapons systems (AWS), which are designed to operate without the direct intervention of humans and use advanced technologies of Artificial Intelligence. These are systems that once activated by a human operator, use the elaboration of data coming from sensors to select and engage targets with strength, without the necessity of human intervention. Systems of autonomous lethal weapons do not simply represent one or two typologies of weapons; rather, they constitute a category of Capabilities, namely a weapon system that incorporates autonomy in its critical functions, in particular in the selection and engagement of targets. The challenges associated with these weapon systems derive from this capability, which grants such systems a level of unpredictability that can generate a series of problems and unintended consequences, creating chain reactions in the context of conflict. States are increasingly developing and deploying weapons with autonomous functions; however, it is important to consider that certain rudimentary autonomous systems have existed for decades.

The most common weapons with autonomous functions are defensive systems. These include systems such as anti-vehicles and anti-person mines, which, once activated, operate autonomously based on activation mechanisms. More recent systems, that employ more sophisticated technologies, comprehend missile defence systems and guard systems, that can detect and engage autonomously targets and issue warnings. Other examples are hovering munitions (also known as Kamikaze drones), which include a built-in head that waits in a predefined area until a target is localized by a human operator or by automated sensors of the device, to then attack it. These systems emerged in the 1980s, but their functions have become more refined, allowing them to operate on lower distances, with heavier weights and the potential integration of AI technologies. Land and sea vehicles with autonomous functions are merging as well. These systems are mainly designed for reconnaissance and gathering of information but could be deployed also proactively.

As mentioned above, LAWS necessitate autonomy to execute their functions without direction and input from humans. Despite AI is not a prerequisite for the functioning of such systems, its integration could strengthen their capabilities. Not all systems include AI. The autonomous capacities can be provided through predefined tasks or sequences of actions based on specific parameters, or by using AI tools to derive behaviors from data, allowing the system to make decisions autonomously or adapt its behavior based on the changing circumstances. AI could be also engaged in a role of assistance in systems directly managed by humans. For example, a computing vision system, managed by a human operator, could use AI to identify and recall the attention on significant objects in the visual field, without having the capacity to respond automatically to those objects.

The military applications of AI are mostly developed along two levels: operational and strategic. At the operational level, the AI importance is manifested in robotics, autonomy, swarm drones, big data modelling, and intelligence analysis to monitor and localize devices, vehicles, and troops. At the strategic level, AI manifests in the capacities of ISR (intelligence, surveillance, and recognizance), in the C3 systems (command, control, communication, and intelligence), and in the missile's defence reinforced with technologies of automatic target recognition (ATR) supported by the machine learning, and in the offensive and defensive cyber capacities enabled by AI. These systems use machine learning techniques to detect, infiltrate vulnerabilities in the networks, and manipulate, deceive, and destroy them. Moreover, AI contributes to increasing the speed and efficacy of the OODA cycle (observation, orienteering, decision, and action), optimizing it in contexts of space defines and cyber war.

LAWS and artificial intelligence are more and more integrated into the strategies of counter-terrorism, representing a crucial element for security forces and intelligence services. These systems can be used to identify and monitor suspected objectives in real-time, analyzing large quantities of data coming from diverse sources, such as satellite images, digital communications, and social media. The use of autonomous drones and robots for surveillance allows operations of information collection in difficult areas, reducing the risks for the man personnel. In this environment, AI not only increases the efficiency of operations but contributes to greater precision, minimizing the risk of collateral damage and potential errors in the visualization of targets[78].

A perfect example of systems used in hybrid warfare situations is the Iron Dome, a aerial defence system developed by Israel to protect the country form aerial tacks, especially from missiles and attacks coming from militant groups, such as Hamas and Hezbollah. This system has demonstrated being a notable technological realisation, capable of foreseeing the trajectory if missiles and intercepting differs types of devices. It was tried for the first time in 2011. Its functioning is based on different advanced technologies such as radars and detection systems, to detect incoming threats; once the missile is identified, the systems calculates its trajectory and determine sit it will hit a inhabited area or not. If the radar foresees that the missile could hot a critical area, the system launches interceptor missiles to destroy the enemy’s device. This process happens in real time and requires a significant computing and analysis capabilities. Iron Dome incorporates advanced algorithms that can be almost considered a rudimental form of artificial intelligence. These

---

[78] Sheikh, H. (2022). *AI as a Tool of Hybrid Warfare: Challenges and Responses*. Journal of Information Warfare, 21(2), 36–49. https://www.jstor.org/stable/27199968

algorithms analyse data and are autonomously decisions, optimising the launch of intercepts to maximise the efficiency of the system[79].

3.4.7 Artificial Intelligence for Wargaming

Wargaming based on Artificial intelligence formulations represents a powerful tool to analyze and predict complex scenarios in political-military contexts, especially in situations that involve the mass destruction of weapons or high-intensity conflicts. This methodology integrates modeling, simulation, and wargaming (MSG) to create a dynamic environment where actors and strategies can be tested in realistic ways. Through the use of AI, simulations can analyse large quantities of historical and contemporary data, to provide a deeper comprehension of the conflict dynamics, allowing decision-makers to explore different options and foresee the possible consequences of their actions. AI-based simulations usually start with the construction of a model that depicts a situation or specific problem, for example, an international crisis or war scenario. AI plays a pivotal role in the elaboration of data and the execution of repetitive simulations that imitate the behaviors of human actors and complex interactions. The advantage of AI is that it allows to supplementing of both quantitative factors, like military or economic capacities, and qualitative considerations, like strategy and human behavior, hardly graspable by traditional simulations. This enables the obtaining of more realistic and reliable results. During Wargaming, AI can be employed to analyze decisions made by human participants, compare them, with simulated scenarios, and advance alternative solutions. This process helps explore crisis dynamics and also identifies weaknesses and opportunities in strategies and tactics. Furthermore, simulations can be used to train military and diplomatic personnel, providing a safe and controlled environment for the exercise of decisions under high-pressure situations. Another interesting aspect of AI-based wargaming is the use of advanced algorithms to refine continuously the simulations. After each war iteration, AI collects data and analyses them, allowing a continuing evolution of models and greater precision in previsions. This automatic learning cycle offers a progressive improvement of strategies and possible responses, making the wargaming a more and more accurate process over time[80].

---

[79] Richemond-Barak, D., & Feinberg, A. (2015). *The irony of the Iron Dome: intelligent defense systems, law, and security*. Harv. Nat'l Sec. J., 7, 469.

[80] SAGE Journal Article: Shukla, P. P., & Hasan, M. M. (2021). *A review of the ethical challenges in artificial intelligence systems*. Journal of Information Technology & Politics, 19(1), 26-46. https://doi.org/10.1177/15485129211073126

Recent progress in neural networks led to the development of deep reinforcement learning (DRL), which integrates deep learning techniques with Reinforcement learning. In this approach, deep neural networks are used to approximate value functions or decisional policies, allowing the management of complex and high-dimensionality environments, where traditional methods based on tables result inefficient. Instead of memorizing explicitly each state-action value, neural networks can predict the Q values starting from the input state, allowing the system to better generalize between known and new states, improving the prevision in little explored areas. Neural networks, moreover, offer a more compact and efficient representation in terms of memory and are capable of dealing with complex tasks thanks to their capacity to represent nonlinear functions. In other words, instead of memorizing all possible actions and results, these networks can predict which will be the results of a certain action based on the current situation. This allows the system to learn and adapt also in new situations. Despite the advantages of these neural networks combined with reinforcement learning, there are some challenges: data changes continuously, and this can make the learning unstable. Additionally, as in other DL techniques, the system might "learn too much" during some specific situations and then behave improperly in new contexts. Another limned that makes the sister a bit inefficient, is that it needs enormous quantities of data to learn properly.

In the wargaming context, the integration of neural networks and reinforcement learning can be very useful in creating virtual enemies that learn and adapt during the game, as a human actor would do. The Deep Networks that use neural networks to estimate the value of decisions "Q values", allow the AI to understand which actions are more beneficial in determined game situations. For instance, in a simulated conflict, the AI could decide when to attack or when to defend. Thanks to neural networks, AI analyses the current state of the game (positions, enemies, available resources) and predicts which action would lead to a better result in the long run. Over time, AI becomes smarter because it manages to generalize situations: even if it finds itself in unknown scenarios, it can make thoughtful decisions based on precedent experiences[81].

---

[81] arXiv Paper: Doe, J., & Smith, A. (2024). *Deep learning approaches to artificial intelligence ethics*. arXiv. https://arxiv.org/abs/2402.06075

3.5 Risks and Opportunities of AI in Counterterrorism

As previously described, the potential of AI applications to contrast the threat of terrorism is wide and is worth careful consideration from armed forces and anti-terrorism agencies. Nonetheless, AI cannot be seen as a quick and easy solution: despite being an attractive technology, there are several legal, political, and technical challenges that all actors must take into consideration before and during its implementation. Many competencies belong to the political responsibility of states or technology providers, however, it is fundamental also for armed forces and agencies to be completely aware of the challenges the AI might face, if the case, adopt the most appropriate supervision mechanisms.

3.5.1 False positives and false negatives

When the accuracy of an algorithm is optimized, the developers of AI can modify the threshold that establishes if a classification label is positive or negative. In other words, machine learning algorithms, when analyzing data, produce scores or probabilities to indicate how much an observation belongs to a certain category. For instance, in a model that aims to identify terrorists, the algorithm calculates the probability that an individual is or is not a terrorist. Regulating this threshold, it is possible to check two types of errors: false positives and false negatives. False positives happen when a positive result is assigned erroneously, the same happens for false negatives. Since a certain error margin is inevitable, and it is impossible to reduce at the same time false positives and false negatives, it is necessary to choose which correct to prioritize. This is not an easy choice, since both negative and positive can have significant implications. In a predictive model that identifies terrorists (imagining being the “positive label”), reducing false negatives means increasing false positives. This approach leads to individuating a greater number of potential terrorists, reducing the risk of dangerous individuals escaping the algorithm. On the other end, it also leads to an increase in cases where innocent civilians are erroneously identified as terrorists, with all the negative consequences that derive. Consequently, the choice of clarification threshold has a huge impact on the results of the AI system used. Thin sensitive contexts as counterterrorism. Where the decision to consider someone a threat or not can have consequences both for national security and the rights of citizens, finding a balance is a crucial challenge. Each decision leads to an ethical and strategic choice on how to balance security and justice[82].

3.5.2 Admissibility of AI-Generated evidence in legal proceedings

The admissibility of AI-generated evidence in courts is a complex matter that raises multiple questions regarding the validity, reliability, and transparency of such devices and systems. Differently from medicines or other technologies that must undergo rigorous approbation and test processes, AI algorithms are not subjected to independent systemic valuation before being used in civil or criminal trials. This leads to a lack of assurances regarding their efficiency and accuracy, creating a situation where AI tools can be employed without any test of their capacity to generate reliable output. Validity and Reliability are crucial in the legal context. Vanity refers to the capacity of an AI system to measure with precision what it had been designed for, namely if the system is capable of providing accurate results about its objective. Reliability, on the other hand, concerns the consistency of results: in other words, if the same system provides the same results under similar conditions. Both these aspects must be demonstrated so that an AI system can be considered reliable in legal circumstances. Unfortunately, the lack of clear standards for the testing of AI products means that many systems currently used, might not pass the necessary tests if subjected to a rigorous scientific evaluation. Moreover, the evidence generated by AI can result in insufficient in terms of authenticity. This complicates the admissibility but also poses serious questions regarding their integrity. As in the case of all digital proofs, also AI algorithm-generated ones can be vulnerable to manipulations, both intentional and accidental, creating an additional state of uncertainty. The possibility that and AI output can be altered, and the difficulties in verifying the integrity of such evidence, ulteriorly complicate the introduction to legal proceedings. In this context, security agencies and legal institutions have required clear guidelines on the admissibility of AI-generated proofs, which should evaluate the impact and specific results of the use of these tools, granting respect for human rights and the rule of law[83].

---

[82] A Joint Report by UNICRI and UNCCT, C*ountering terrorism online with artificial intelligence, An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia*, https://unicri.it/sites/default/files/2021-06/Countering%20Terrorism%20Online%20with%20AI%20-%20UNCCT-UNICRI%20Report.pdf

[83] Paul W. Grimm Maura R. Grossman Gordon V. Cormack, *Artifificial Intelligence as Evidence* (December, 2021), Northwestern Journal of Technology and Intellectual Property, Volume 19 Issue 1 Article 2. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1349&context=njtip

One example case these dynamics is depicted by the use of COMPAS, a system that, despite being accepted as a risk evaluation system, raised concerns regarding the lack of transparency on the methods used to calculate the scores of risk. His decisions based on such instruments, which influence substantially the sentences, cannot be fully justified if the details of the algorithmic functioning are kept obscured. This leads to fundamental questions on the right to a fair trial and the adequacy of proofs based on opaque technologies[84].

3.5.3 Intricacy of Human Communications

As already mentioned previously, a fundamental requirement for the development of reliable AI instruments is access to sets of labeled, accurate, and detailed data. Despite many challenges that are global in scope, South and South-East Asia present additional concern, given the enormous linguistic and dialectic diversity. There are an estimation of about 2.300 languages in all of Asia[85]. Despite the significant progress in the field of Natural Language Processing, the majority of efforts focus only on a few common languages: English, Chinese, Urdu, Farsi, Arabic, French, and Spanish. Additionally, languages are ever-evolving and adapting to social changes. This dynamism represents a big challenge for AI since AI systems cannot adapt autonomously to new contexts without large quantities of data to support such adaptation. Consequently, detests must be constantly monitored and dated. For many less-spread languages and dialects, the lack of available training data features a great difficulty.

A concrete example has recently emerged when Facebook algorithms, designed to identify and block malevolent content, have been deceived by a right-wing extremist group in the United States, that used an encrypted language to share publicly instructions on how to fabricate bombs. Not being able to decipher the code, the algorithms were not able to reveal the content. The context adds ulterior complexities because it influences the shades of human communication, like irony and sarcasm. To detect them is an extremely difficult task, also for human beings, since it varies for each person and depends on culture, facial expressions, and voice tone. In the sentiment analysis of

---

[84] Maria do Céu Cunha Carrão, *Artificial Intelligence in Criminal Proceedings: The admissibility of AI-generated evidence*, June 2023, pp. 33-39 https://run.unl.pt/bitstream/10362/145184/1/Carrão_2022.pdf

[85] Andy Kirkpatrick & Anthony J. Liddicoat. (April 2019). *The Routledge International Handbook of Language Education Policy in Asia*. Routledge.

NLP, sarcasm is a specific area of research, since it cannot be simply classified as positive or negative sentiment. Some studies have demonstrated that sarcasm can reduce the accuracy of automated sentiment detection by 50%.

Some approaches that try to deal with this issue include the addition of information layers to better capture the ratio between the talker and environment, such as the analysis of pragmatic characteristics (emojis, hashtags…). In May 2021, Twitter launched a new warning functionality for hate speeches, asking users to re-elaborate responses having potentially offensive words. The first tests showed that the algorithms had difficulties in grasping the shades of conversions and often couldn't distinguish between offensive language and sarcasm. Subsequent developments include the addition of data such as the relation between the user and the respondent[86].

3.5.4 Public-Private Sectors Relation

One of the problems that characterises the international view, not only in the field of counterterrorism, is the relation between the public and private sectors, particularly the role of the private sector in the fight against terrorism through the use of artifice intelligence. With the growing importance of online communications, the weight of the private sector is notably increasing thanks to the resources and skills it possesses compared to the public sector. Indeed, technological agencies can develop AI-advanced algorithms, able to analyze large quantities of data and carry out all possible tasks. As a consequence, public actors often seek partnerships with private entities to integrate and support their efforts in contrasting terrorism, since terrorist materials are uploaded on platforms owned and managed by these private companies. Online platforms not only host dangerous content but also manage the technological infrastructure necessary to monitor and analyze such content through AI systems. Furthermore, the private sector, plays a role of "guardianship", having the access to data necessary for the training of algorithms. Thanks to these resources, private enterprises own a greater capacity to design and develop technological devices to fight terrorism.

---

[86] A Joint Report by UNICRI and UNCCT, *Countering terrorism online with artificial intelligence, An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia*, https://unicri.it/sites/default/files/2021-06/Countering%20Terrorism%20Online%20with%20AI%20-%20UNCCT-UNICRI%20Report.pdf

This relationship is to always linear and simple, often due to the divergent goals. For instance, materials that might require action by armed forces could be deleted by the company owing them, because deemed to be obsolete and useless, reasoning in reality important evidence for investigations[87]. In this context, decisions by agencies to remove potentially useful content can comprise the efforts of authorities in preventing attacks or prosecuting individuals. Materials are to be archived in a way accessible for actors other than the company itself unless the authorities have a mandate or court order. This lack of access to historical data can be a significant issue for anti-terrorism authorities, who might have difficulties in establishing the existence or not of relevant evidence. The consequences of these dynamics are multiple and complex. On one side, the private sector can be a precious ally in the fight against terrorism thanks to advanced technology; on the other side, the lack of transparency and accessibility might jeopardize entire operations. It is essential to set a clear framework that regulates this public-private relationship.

3.5.5 The black box nature of AI

The Black Box nature of artificial intelligence represents one of the most significant challenges in the analysis and interpretation of modern technologies. The AI application, which ranges from facial recognition to medical diagnosis, operates on algorithms that are often obscured and inaccessible to users. This lack of transparency makes it difficult to understand how the systems achieve determined conclusions or decisions. The unreliability of such processes not only creates a barrier to the trust of users but raises also fundamental questions on the accountability and equity of automated decisions. This nature manifests in two main ways: the opacity of data and the complexity of algorithms. Machine learning algorithms, for instance, are designed to learn from data and improve over time, but the way they process information is often incomprehensible, also to the same developers. The decisions deriving from these algorithms can have impactful effects on people's lives, from the determination of loans to criminal surveillance. However, the relation between original data and final decisions is frequently obscured, leaving users without a clear comprehension of the logical basis or statistics that sustain such decisions. This phenomenon is not only a technical matter but also intertwines with sociocultural and political considerations. Users who interact with these systems, both normal citizens and subject matter experts, find themselves in

---

[87] Stuart Macdonald, Sara G. Correia, & Amy-Louise Watkin. (2019). *Regulating terrorist content on social media: Automation and the rule of law*. International Journal of Law in Context, 15(2), 190. Accessible at https://doi.org/10.1017/S1744552319000119

first of frustrated and disorientation when facing results that do not fully understand. The black box effect, therefore, manifests in the way people respond to these uncertainties. The difficulties of accessing data and underlying logic to algorithms indicate that the regulating agencies and researchers often can examine not evaluate critically the efficiency or equity of these tools. Information that could clear the functioning of a system is, mostly, confidential or incomprehensible. This lack of access creates an epistemological gap, where people are forced to navigate insecurities, relying on assumption and deception rather than tangible evidence.

This matter is tightly linked to the concepts of Explainable AI (XAI) and course transparency. XAI proposes to develop models and algorithms that not only are efficient but also interpretable so that users can understand how and why a certain system achieved a specific decision or conclusion[88]. This transparency requirement is fundamental to gaining the users' trust, guaranteeing that automated decisions are justifiable, and minimizing algorithmic biases which will be explained in this chapter. Having access to clear explanations is essential, especially in critical contexts such as healthcare, criminal justice, and financial decisions, where the consequence of such decisions can have a deep and long-lasting impact on people's lives. Nevertheless, the realization of XAI is not free from challenges and issues. Many ML models, especially when based on deep neural networks, are intrinsically complex, and present difficulties in expelling intuitively their operations. Even when explanations are provided, these can result in technically complex or vague, further alimentation the confusion among users. Additionally, explanations may be received as mere superficial justifications rather than genuine clarifications. Therefore, the goal of XAI is to be accompanied by a concrete effort for transparency, where clarity is not only an option but a necessity to grant ethics and accountability of the technologies. Only through an integrated approach that favors the compression and interpretability of AI systems, will be possible to efficiently face concerns related to the black box, promoting responsible use of AI in the society[89].

---

[88] Harmanpreet Kaur, et al. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. CHI 2020 Paper. Accessible at http://www-personal.umich.edu/~harmank/Papers/CHI2020_Interpretability.pdf

[89] Sternfeld, J. (2024). Chapter 9: Historical Evidence, Artificial Intelligence, and the Black Box Effect. Transactions of the American Philosophical Society, 112(3), 175–200. https://www.jstor.org/stable/48757509

3.6 AI used by Terrorists

Artificial intelligence is a powerful instrument, used the a variety of public and prate sectors to improve the quality of life and the well-being of society. Nevertheless, AI also presents a dark side: being a technology for general use, it can be used or abused by malign actors. In a recent Europol report, Trend Micro and UNCRI highlighted the different ways these actors, informatics criminals are already issuing AI both as a vector for attacks or as a surface of attack[90]. As Ai can be used for criminal purposed, it can also be employed by groups and individuals to intensify terrorist attacks or amplify their capacities to spread propaganda and incite to violence, al mentioned before. Thee are three main elements contributing to the concerns regarding the potential malign use of AI. The first one is the process of AI democratization[91]: This concept refers to the fact that technologies once accessible only to a restricted community with high resources and competencies, are becoming more and more accessible to all, thou huge investments or elevated technical skills. Many popular algorithms are open source and can be used with a limited technical education. The second aspect the the scalability, menacing the capacity of a technology to increase in dimness and volumes to manage a growing demand. Because of the particular scalable nature of AI, which deals with defense against dark uses, must prepare not only for individual attacks but also for a rise in the volume of simultaneous attacks. One exemplar case is the threat represented by drone swarms that fly automatically in synchrony. Even if the use of AI by terrorists fails, due for instance to a lack of competencies and knowledge, it still might have significant psychological impacts. Furthermore, where anti-terrorism entities must keep in mind civil rights and fundamental freedoms in the use of AI, terrorist groups do not have the same concerns. Terrorists have demonstrated to be able to adapt very rapidly and adapt to the ever-changing circumstances. From the technological perspective, this tendency to innovation is particularly evident concerning the internet and social media, which have been revealed to be extremely precious to terrorism. These tools have been used to radicalize, inspire, and promote violence, vindicate attack responsibility, recruit, collect, and transfer funds, acquire and move weapons, and make available tutorials and tools to members. In 2020, Europol and 17 countries have identified and evaluated the removal of almost two thousand URLs connected

---

[90] Trend Micro Reserach, United Nationas Interregional Crime and Justice Research Institute (UNICRI), Europol's European Cybercrime Centre (EC3), *Malicious Uses and Abuses of Artificial Intelligence*, https://documents.trendmicro.com/assets/white_papers/wp-malicious-uses-and-abuses-of-artificial-intelligence.pdf

[91] UNCCT & UNICRI, (2021). *Countering terrorism online with artificial intelligence: An overview for law enforcement and counter-terrorism agencies in South Asia and South-East Asia*, pp. 11. https://unicri.it/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

to terrorist content on 180 platforms. The diffusion of end-to-end cryptography (E2EE) on social media platforms, aroused significant concerns among decision-makers regarding the possibility that terrorists go dark and communicate eluding surveillance.

Even though technology has not been the only factor relevant to the evolution of terrorism, it certainly has played a pivotal role. The interaction between technology and terrorism is evident in three different ways: first of all, the nature of attacks has changed over time, from the use of knives and firearms to hijacking and vehicle attacks, with some terrorist organizations threatening the purchase of chemic, biological, or radiological materials. The adoption of the automatic rifle is probably the most significant modification linked to technological development. Secondly, the technological progress in the sector of transport and logistics has transformed the capacity of these actors and criminals, allowing them to increase the speed and magnitude of their operations, making them global. Finally, development in terms of ICT allows these groups to communicate rapidly and in a reserved way over growing distances. They improved the efficiency and effectiveness of attack and recruitment. The lack of proof of terrorists using direct Artificial intelligence should be interpreted as an indication that terrorists are indifferent or uninterested in this technology. These groups are aware of the revolutionary potential of such devices and systems. For example, in 20216, a video supposedly created by ISIL in Syria showed the groups that experimented with a rudimental version of an autonomous vehicle remotely controlled. There is the belief that the use of mannequins inside the vehicle was a way for these actors to replicate the thermal seal of human beings to further elude security systems. More recently, in March 2020, an ISIL follower diffused a video on Rocket.chat, a social platform used by ISIL to spread terroristic content, explaining how the software facial recognition could be used.

The list of malign uses of AI by the terrorists side is long; however it is possible to distinguish three main categories of action based on their goals: enhance cyber capabilities, enabling physical attacks and spread of propaganda and disinformation. Some of the actions that will be describe, can be used also combined together and for multiple purposes.

3.6.1 Enhancing cyber capabilities

Denial-of-Sevice attacks (DoS) or distributed denial-of-service (DDoS) have been popular attacks for decades, with the main goal of making temporarily inaccessible an informatics system connected to the Internet, consuming all its memory through multiple connection requests.[92] Aggressors usually see more machines, sometimes thousands, in the so-called botnet to direct the requests to the target. It is believed that between 2016 and the beginning of 2917, ISIL launch successfully its first series of DDoS, hitting mainly military, education, and economic infrastructures, demonstrating the seriousness of this threat. What makes the attacks attractive for cybercriminals and terrorists, is the possibility to launch them with little effort and their execution is relatively easy. The attacker doesn't need to seek specific vulnerabilities in the target: the fact that the system is connected to the internet is enough. For example, machine learning algorithms can be used to check Botnets behind the attacks or identify vulnerable systems through recognition techniques. The potential of machine learning applied to DDoS attacks is already exploited: in 2018, the platform TaskRabbit, an online marketplace for freelance workers, was subjected to a DDoS attack perpetrated u a hacker who used a botnet consoled by AI software. This attack hit 3.75 million users, provoking a signification violation of data.

Another tool used is malware, which comprehends a wide range of programs that penetrate inside a network or system, provoking damage or interrupting operations. Her malware examples are spear, ransomware, virus, and worm Trojan horse. Malware has long been used by several malign actors. AI can be exploited by malware developers to automatize the attack processes, improve efficiency, or create new forms of malware. It has already been used to create a poly form malware, capable of adapting and modifying itself to avoid its detection. Ransomware, a subset of malware, is one of the greatest information threats at the global level. Famous examples like WannaCry in 2017 demonstrate how quickly the ransomware can spread and cause damage on a large scale. The integration of AI could amplify these attacks, making them evermore sophisticated and targeted. Finally, other techniques of password guessing and breaking CAPTCHA can be strengthened through AI, significantly reducing the necessary time to violate systems and websites, allowing access to information, or the launch of other attacks. Also, encryption and decoration mechanisms can damage AI, improving the security of communication among gores and at the same time facilitating access to anti-terrorism agencies' information[93].

---

[92] *History of DDoS Attacks*. (Mar. 13, 2017). Radware. Accessible at https://www.radware.com/security/ddos-knowledge-center/ddos-chronicles/ddos-attacks-history/

[93] Chen, Q., & Bridges, R. A. (2017, December). *Automated behavioral analysis of malware: A case study of wannacry ransomware*. In 2017 16th IEEE International Conference on machine learning and applications (ICMLA) (pp. 454-460). IEEE.

3.6.2 Enabling physical attacks

Vehicles like cars or tracks, have often been used by terrorists to commit physical attacks, such as the one that happened in Berlin in 2016 and Barcelona in 2017. However, the advent of AI opens the way to more worrying scenarios. With its integration, autonomous vehicles could become lethal instruments in terrorist attacks, allowing the planning of attacks remotely, with the tackier risking his life. Autonomous cars, indeed, use AI to replicate the decision of the driver, dealing with swerving, acceleration, and braking, which make it possible for a terrorist to sedan attack while being physically present[94]. The potentialities of AI, together the the growing commercial development of autonomous cars, could also such actors to use the, as explosive devices, eliminating the necessity of a suicidal attacker. Also drones, with the integration of AI, offer new possibilities for attack. In particular, the combined use of AI with facial recognition, an extremely useful tool for security agencies and armed forces, could be exploited for target killings by the terrorist side as well. This scenario has already been shown in the viral video “Slaughterbots”, which shows how micro drones equipped with explosive and facial recognition can be used to eliminate selected targets with extreme accuracy. Even if this technology is not available yet, the elements already exist and technologically competent terrorists could assemble them to create lethal weapons. Another concerning area is the use of AI in the field of biotechnology to develop targeted biological weapons. The progress in terms of genetic sequencing and the growing access to genetic data through biobanks could allow terrorists to exploit AI to develop new forms of biological attacks. AI applied to genetic data, could theoretically allow the creation of pathogens designed to hit septic genetic groups. Despite knowledge has not reached this level yet, the combination of AI and biotechnology opens up the frontier to bioterrorism.

3.6.3 Other malign applications

Deep fakes, based on advanced AI techniques, progressed exponentially in the manipulation of visual and audio materials. Despite initially being known for the capacity to create fake videos that manipulated the aspect of a person, the technology has rapidly evolved, becoming capable of

---

[94] Jeffrey W. Lewis. (Sept. 28, 2015). *A Smart Bomb in Every Garage? Driverless Cars and the Future of Terrorist Attacks*. Smart. Accessible at https://www.start.umd.edu/news/smart-bomb-every-garage-driverless-cars-and-future-terrorist-attacks

replicating human voices and making it difficult to distinguish artificially created materials from authentic ones. In the field of images, GAN is used, to create videos where it seems a specific person is doing or saying things that in reality did not. Deep fake videos can overlap two different faces indistinguishably, or alter the facial expression and labial movements to synchronize them with an artificial dialogue. These techniques are widely used to create fake political speeches, deceiving the public and creating potential crises. The same happens for audio deep fakes. The machine learning algorithms are trained on audio registrations of a person to learn their vocal timbre, intonation, rhythm, and other details. Once the AI has enough data, it can create new vocal content that imitate perfectly the voice of that person.

The Internet offers a certain degree of anonymity, that facilitates some of its most dangerous uses, such as cyberbullying, manipulation, or minors exploitation. There is an esteem of more than 750 million fake accounts on Facebook, used to spread compromised content. AI technologies, such as GAN, which are the base of deep fakes, can be used to create highly realistic false images. This is strictly connected to the necessity for terrorism to obtain, alter, or falsify travel documents. Falsified passports are regularly used to facilitate international movements and for other administrative matters such as the obtaining of loans or visas.

In conclusion, Artificial intelligence, although offering infinite advantages in several sectors, brings a dark side that raises serious security concerns. The capacity to manipulate visual, audio, and behavioral data, as well as automate sophisticated processes, opens up the way to potential malign use of such technology, often difficult to detect and contrast. Its rapid evolution overpasses the legal and technological countermeasures, creating a reality where the border between true and false is more and more blurred. Facing these challenges, governments, companies, and societies must work together to balance innovation with accountability, mitigate risks, and prevent abuses[95].

---

[95] United Nations Office of Counter-Terrorism, UN Counter-Terrorism Centre (UNCCT), & United Nations *Algorithms an terrorism: the malicious use of artificial Intelligence for terrorist purposes*. https://unicri.it/sites/default/files/2021-06/Malicious%20Use%20of%20AI%20-%20UNCCT-UNICRI%20Report_Web.pdf

CHAPTER 4 - ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS: A TOOL FOR SECURITY OF A THREAT TO HUMANITY?

4.1 An Ethical Dilemma

Over the last few years, artificial intelligence has registered an enormous development, redefining the social, economic, and political dynamics at the global level. Its growing diffusion, welcomed with enthusiasm for the promising opportunities it offers, has generated a more and more heated debate on its impact on human rights. On one hand, AI has been recognized as a powerful tool capable of facing some of the greatest and most critical challenges that Humanity has to deal with, such as climate change, international conflicts, and medical discoveries. It promises to improve the quality of life, increase scientific knowledge, and strengthen human exploration. Nevertheless, on the other hand, the use of AI has raised a series of concerns regarding the potentially damaging effects, in particular on the level of fundamental human rights protection. The large-scale application of AI technologies has brought about complex implications for individual, collective, and social rights. The debate on AI and human rights focuses on an intrinsic tension between its potential for innovation and the risk of violations that derive from an unethical or poorly regulated use. According to several scholars (Raso et al., 2018; You, 2019), AI implementation has already created important concerns concerning decisions and results that risk compromising privacy, equality, and social justice. On one side, AI could improve crucial sectors such as healthcare, national security, and environment protection; on the other, it could facilitate mass surveillance, amplify existing discriminations, and limit individual freedoms, putting in danger some fundamental rights[96].

The increasing importance of AI has been confirmed by the significant investments that are carried out globally. According to the International Data Corporation (IDC), global spending is estimated to surpass 300 billion dollars by 2026, thanks to an annual growth of 27%. These numbers demonstrate how AI is considered a fundamental strategic resource, pushing many countries to develop national strategies to promote adoption and innovation. In the US, for instance, the federal government has increased notably the investments in research and development linked to AI and machine learning. In 2020, the federal balance included 4.9 billion dollars destined for non-

---

[96] Greiman, V. (2021). *Human Rights and Artificial Intelligence: A Universal Challenge*. Journal of Information Warfare, 20(1), 50–62. https://www.jstor.org/stable/27036518

classified projects on AI, with particular attention to sectors such as cybersecurity and advanced microelectronic development. Also in Europe and Asia, several governments are investing in advanced research programs. Countries like China, France, Canada, and South Korea have launched over the last years national strategies on AI, aimed at developing public services based on intelligent technologies and adopting more sophisticated armaments. These policies refit the growing geopolitical and strategic interest that surrounds AI since it is perceived as fundamental leverage for the economic and military powers.

The AI expansion has advanced new ethical, legal, and regulatory challenges. The use of complex algorithms and machine learning technologies to make designs that regard people's lives, raises accountability, equity, and transparency questions. For instance, facial recognition systems, increasingly used for public security purposes, risk violating the right to privacy and can be used to monitor, profile, and limit the movements of individuals without their consent. Moreover, AI raised significant doubts also in the field of international security. With the increase of automation and autonomous decisional capacity in military systems, the possibility of arms based on AI making lethal decisions without human supervision led to concerns regarding the international humanitarian law and norms of armed conflicts. Experts warn that AI could alter radically the power dynamics among states, as seen previously when talking about the AI race, creating new risks against strategic stability and global security[97].

4.2.1 Algorithmic Repression

Political repression is a classic concept, well developed in traditional scientific and political discourses, with different interpretations folded by the different contexts, cases, and disciplinary perspectives. Despite some consider it a manifestly violent phenomenon, others underline its more subtle manifestations. For political purposes, it is necessary to define operatively the concepts of "AI-guided repression" and "algorithmic repression"[98]. The concept of digital repression, as articulated by Steven Feldstein (2021)[99], comprehends a wide range of state-sponsored activities

---

[97] European Parliament. (2024). *The impact of artificial intelligence on human rights* [Study]. Retrieved from https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf

[98] ibid.

[99] Feldstein, S. (2021). *The rise of digital repression: How technology is reshaping power, politics, and resistance*. Oxford University Press.

that exploit information and communication technologies as powerful instruments to suppress dissent and control the population. This form of repression represents an evolution from traditional methods. Digital repression involves systemic efforts by state actors to use advanced technologies to monitor the digital tracks of individuals and groups. Through surveillance, governments can access a wide range of information generated online by citizens, from location data to communication models, from social networks to consumers' behaviors. This allows states to conduct not only large-scale surveillance, but also coerce populations by threatening the exposition and punishment for online activities, using personal data, extrajudicially gathered, in legal proceedings. Other than surveillance, digital repression includes the manipulation of information to shape public opinion and silence opposition. This can happen through the diffusion of state-sponsored propaganda, the use of botnets that spread disinformation, and the censorship of online content that is considered subversive or damaging to the state's narrative.

One insidious element of digital repression is its capacity to dissuade activities or convictions that challenge the state, without necessarily having confrontations. The conviction of being constantly observed can induce an auto-censorship among citizens, suffocating the dissents even before being expressed. This cooling effect on the freedoms of expression and association is a pillar of digital repression since it transforms subtly the behaviour of individuals, through the the perception of a threat of retaliation for anti-state actions or ideologies. This is a key concern for the European Parliament since many authoritarian governments ignore European concerns about the violation of digital human rights. Despite not being a manifested suppression of information, more sinister and deep disinformation networks that generate auto-censorship, often escape the attention of European monitoring systems. Nonetheless, digital repression can manifest also in more coercive and open ways. The state can launch cyberattacks against opposition groups, manipulate digital platforms to interrupt the organization of protests, or employ legal instruments to justify the arrest or prosecution of digital dissidents. The notion of "algorithmic repression" becomes crucial because it captures subtle but powerful ways in which technology and social media enterprises, together with state actors, can perpetuate the hegemonic control and suppress dissent. Algorithms that feed socials can silently suffocate the opposition through means such as the *algorithmic filtering* and the shadow banning, imposing a form of digital censorship that is as efficient as imperceptible. The algorithmic filtering is used by platforms such as Facebook or google, as personalised algorithms that tailor information based on what the final user wants, needs or based on the people they know: consequently, the search engine results are diversified for echo user, and two people with the same parameters might see different updates and informations, based on past interactions with the platform[100]. With term "shadow Banning", on the other hand, we refer to "*limit or eliminate the exposure of a user, or content or material posted by a user, to other users of the social media Internet site through any means, regardless of whether the action is determined by an individual or an algorithm, and regardless of whether the action is readily apparent to a user*"[101], namely a practice used mainly on social media, where contents of a user are hidden or limited without the user being aware. Despite posts or comments are still visible to the user, their public is significantly reduced, often making them invisible to other users or less evident in the research or feed algorithms. This tools is used to reduce the diffusion of content that is considered inappropriate or potentially dangerous without an explicit censorship or a total ban.

Margaret E. Roberts (2018) focused part of her work on the examination of the sophisticated information control in China, revealing not only the block of content but also the art of inundating the digital arena with noise (content flooding or hashtag hijacking), a technique to distract rather than confront. In her book "*Weapons of Math Destruction*", Cathy O'Neil explores how algorithms, in particular those used in big data, can perpetuate and exacerbate social and economic inequalities, leading to forms of repression, not necessarily confined to autocracies. O'Neil sustains that many of these algorithms, although appearing neutral and objective, are based on distorted data or wrong assumptions. This can lead to discriminatory results, such as the unfair targeting of determined groups for surveillance by police, the negation of opportunities based on opaque credit score systems, and the perpetuation of prejudices in the assumptions. O'Neil warns of the dehumanizing effects of these automated decisions, especially when used by those who have the power. She underlines how these algorithms, especially when employed by political actors, can systematically marginalize groups, manifesting as a form of political repression[102]. Safiya Umoja Noble (2018) does not focus on algorithmic repression per se, but exposes the insidious prejudices in search engines such as Google, revealing a technological infrastructure that not only reflects but amplifies also racial and gender prejudices[103]. The academic discourse around algorithmic repression is

---

[100] Bozdag, E. (2013). *Bias in algorithmic filtering and personalization*. Ethics and information technology, 15, 209-227.

[101] Nicholas, G., *Shedding Light on Shadowbanning*, April 2022, Centre for democracy and technology.

[102] O'neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

[103] Noble's, S. U. (2019). Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018. why popular culture matters, 166.

further amplified by interdisciplinary scholars who examine the multifaceted impact of digital technologies on freedoms and control. For example, Frank Pasquale (2015) provides a critical analysis of how algorithmic processes govern the economic opportunities and distribution of information in a society. These reveal the opacity of algorithmic decisions in crucial sectors such as finance or media, where the lack of transparency can lead to a form of repression that systematically individuals and groups without clear possibilities of appeal[104]. In the same way, Virginia Eubanks (2018) enlightens the socioeconomic dimensions of algorithmic decisions and explores how automated systems are employed in public services, resulting often in a new form of digital discrepancy that exacerbates existing inequalities, marginalizing the poorest and reinforcing systemic prejudices in only apparently objective technologies[105].

4.2.2 AI, Democracy & Rule of Law

The impact of AI on democracy is emerging as one of the greatest and potentially most dangerous challenges of our times. Despite the AI offers numerous opportunities to improve society, from economic productivity to infrastructure management, it raises crucial questions related to the functioning of democratic institutions, transparency of electoral processes, and the informed participation of citizens. Consequences not only regard the possibility of manipulation and political control, but also the risk of accentuating social and economic inequalities, the creation of vulnerable oysters and systemic failures, and concentrating the decisional power on the hands of a few big technology enterprises. The first critical point regards the role AI plays in the formation of the social and political discourse. In modern democracies, the capacity of citizens to access truthful and diversified information is essential for an informed decision-making process. However, the growing use of AI algorithms in digital platforms is transforming radically the way people consume information. Through systems of information personification, such as social media or search engines' feeds, AI determines which contents are shown to users, selecting what is more relevant based on behavioral analysis. This leads to significant consequences: on one side, these technologies can help citizens deal with information flows, providing the most pertinent materials; on the other side, they risk creating an extremely personalized informatics ecosystem that isolates

---

[104] Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[105] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

individuals, exposing them only to opinions and information that reinforce their preexisting convictions. This dynamic erodes informatics pluralism, essential for a healthy democracy since it limits the exposition to alternative perspectives and weakens the public debate.

AI can also be used to manipulate political perceptions in subtle but efficient ways. The example of electoral micro-targeting, which exploits behavioral data to send personalized political messages to specific groups of electors, raised serious concerns in numerous recent elections, including the case of Cambridge Analytica and the American presidential elections in 2016. This type of personalized politics allows candidates or other actors to send targeted messages without adequate public transparency, bypassing open discussion and democratic control. Furthermore, it further aggravates the problem, since it deceives public opinion and mines the trust in media and in democratic institutions. This situation becomes even more complicated if considered that the majority of electors do not have the tools nor the skills to identify such manipulations. AI technologies, indeed, can act in invisible ways, influencing subtly public opinions without the citizens being aware of it. The combination of these technologies with disinformation campaigns organized by state or non-state actors represents a significant threat to the legitimacy of electoral processes and to the decisional autonomy of citizens. AI risks mining the fundamental assumption of each democracy: the informed and autonomous elector.

Another insidious impact of AI regards the systemic risks that its implementation introduces in social and economic systems. Along with the integration of AI into crucial sectors such as financial markets, energy, and healthcare, decisions that were previously made by a plurality of actors, are now centralized in a few automated systems. This creates a systemic vulnerability, where the failure of a single system can have catastrophic consequences. The case of the "flash crash" in financial markets is a clear example: interactions among high-frequency trading algorithms led to sudden collapses with no precedents of markets, with global economic impacts. In the same way, critical infrastructures such as electronic networks or transport systems, if controlled by automated systems, could be vulnerable to cyberattacks or bad functioning that could paralyze entire sectors of society. These vulnerabilities become particularly serious in a geopolitical context where AI and the arms race based on intelligent technologies, represent a concrete danger to international stability and collective security. Furthermore the centralization of control on the hand of a few actors, that could be both authoritarian governments or big technology enterprises, further accentuates these risks. The concentration of decision-making power, both political and economic, could lead to situations where crucial decisions for the general public well-being, are made without adequate democratic supervision. If not correctly regulated, AI could transform into a totalitarian control instrument, capable of monitoring, manipulating, and repressing citizens without them having the proper tools to defend.

Artificial Intelligence is significantly influencing also the principle of the Rule of Law, posing a series of challenges to public institutions and the procedural legitimacy of modern democracies. Public institutions, for their own nature, are subjected to higher standards than private ones, especially in relation to principles of justification, proportionality, and equity in their interactions with citizens. On one hand, AI can increase the efficiency of judicial and administrative institutions; on the other hand, it can compromise the transparency and trust in democratic institutions and mine the authority of law, generating less accessible and comprehensible systems for the public. One of the main risks regards the reduction of human control and the loss of transparency in decisions. For instance, in the justice field, the adoption of automated systems to solve online controversies or to evaluate evidence in courts could improve the efficiency of judicial systems. However, these systems, often administered by private enterprises, are governed by internal service and rules that do not offer the same assurances and procedural protections provided in public tribunals. This can lead to a loss of trust from the citizens' side toward the institutions, which risk appearing more distanced and opaque, with decisions made by "black box" algorithms, difficult to understand and contest. Another example relates to the control of online hate speeches. Traditionally, only tribunals were authorized to determine if a certain text constituted an illegal hate speech. Nevertheless, today the majority of decisions related to the removal of social media content are made by private AI systems, designed to generate decisions that do not fully respect the principle of the rule of law, creating a parallel justice that bypasses institutional controls. Clearly, AI does not represent solely a menace to this principle, since it can potentially individuate and prevent corruption phenomena, detection of suspected patterns, or the protection from cyberattacks on critical infrastructures, therefore reinforcing the safety and stability of democratic institutions. European Union's already recognized these potentialities, elaborating ethical guidelines for these AI tools in judicial systems. The European Ethical Charter on the Use of AI in the Judicial Systems, published in 2018 by the European Commission for Justice Efficiency, establishes some principles for guiding the development and implementation of AI technologies in the judicial field, among which transparency, equity, and protection of human rights[106].

A further crucial issue is the access to justice and availability of efficient remedies in case of violation of rights caused by AI. Many AI applications, developed by a limited number of technology companies, operate on global platforms that overpass national jurisdictions, making it difficult for citizens to obtain justice in case of abuse. This concentration of power raises fundamental questions on how to protect individual rights in an ecosystem that is incrementally dominated by non state entities. Many experts suggest the implementation of decentralized decision-making processes, where the control is distributed to multiple actors, both human and technological, reducing the risk of a chain collapse. To guarantee that these systems do not undermine human autonomy or have adverse effects, it is fundamental to provide adequate human supervision structures, realized by different governance mechanisms such as the concept of human-in-the-loop (HITL), Human-on-the-loop (HOTL) or Human-in-command (HIC), which in order, provides human intervention in each decisional phase of AI system, provides human intervention in the designing cycle of the system and monitoring of its operations, and finally, the last one allows comprehensive control over the system, leaving to the human the choice of interrupting, activating or deactivating the system[107].

4.2.3 Right to Privacy

The protection of the right to privacy has acquired great relevance in the contemporary context, especially with the advent of the digital revolution. In several countries, like the UK and the USA, privacy is seen as a fundamental protection against governmental, corporate, or individual intrusions in the private lives of citizens. However, there exist situations where citizens are forced to provide personal data to the state, creating a tension between public duty and the protection of personal confidentiality. The right to privacy is often defined as a fundamental human right,

---

[106] David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, & Morgan Briggs; With a foreword by Lord Tim Clement-Jones, *Artificial intelligence, human rights, democracy, and the rule of law*, May 2024. The Alan Turing Institute. https://edoc.coe.int/en/artificial-intelligence/10206-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-a-primer.html

[107] Albrecht, J. P., & Giebel, E. (2020). *The Impact of AI on Human Rights, Democracy, and the Rule of Law*. Retrieved from https://allai.nl/wp-content/uploads/2020/06/The-Impact-of-AI-on-Human-Rights-Democracy-and-the-Rule-of-Law-draft.pdf

intrinsically connected to the humanist tradition, that recognizes the intrinsic value of each individual. This right is essential to protect integrity and personal dignity, being also a basis for individual autonomy. The possibility of making decisions without external interference is crucial for the development of a democratic and free society, and this aspect assumes even greater importance in the digital context, where surveillance and data collection are increasingly invasive. The protection of privacy guarantees individuals the necessary space to develop their own identity, make personal decisions, and maintain control over which information is shared and how it is used. Without such protective measures, the individual becomes vulnerable to intrusions that might threaten their dignity and safety, such as identity thefts or manipulations. Furthermore, privacy is also essential to build trust among individuals, organizations, and governments: when people know that their personal information is protected, they are more available to share sensitive data when necessary, facilitating the functioning of society. Privacy is strictly linked to the freedom of expression and social trust. It protects individuals from prejudices, allowing free expression and participation in public debates, without fearing repression or surveillance. In a democracy, this is essential to guarantee that all voices can express, without repercussions. Contemporarily, privacy constitutes a limit to the institutions' power, both governmental and corporate, impeding the capacity to gather and use personal data in an abusive and indiscriminate way[108].

The difference between privacy and data protection is essential, since both concepts over even though regarding distinctive aspects of the personal sphere. Privacy, as above mentioned, refers to the right to maintain a certain degree of discretion regarding one's personal life, protecting the intimate sphere from external intrusions. Data protection, on the other hand, focuses on the correct management and treatment of personal information by organizations, companies, or governments, reassuring that data are used transparently and conformally to the law. In the current context, dominated by emerging technologies like AI and the Internet of Things (IoT), the dissection between these two concepts is gradually blurred. AI, in particular, analyses enormous quantities of personal data to predict behaviors, make autonomous decisions, or improve services, reducing the space of individual autonomy. For instance, platforms like Facebook and Google collect personal data to offer "free" services, but the real price paid by the users is their privacy since data are used for commercial purposes or to manipulate preferences and behaviors. A crucial element regards the impact of technologies on this difference: AI makes the line between data collection with content

---

[108] Shams, A. (2023). *Right of Privacy and the Growing Scope of Artificial Intelligence*. Current Trends in Law and Society, Volume 3, Number 1, 2023, Pages 1 - 11

and without it more shaded, complicating the protection of privacy. Without a clear regulation, there is the risk of sacrificing the protection of personal data to the benefit of speed and economic efficiency, introducing new forms of commercial exploitation of personal information[109].

Surveillance technologies have experienced an evolution with no precedents thanks to artificial intelligence, which made possible the constant analysis and monitoring of individuals and groups at a global scale. While traditional surveillance requires significant physical resources, today the algorithms and connected devices, can gather, analyze and interpret personal data in real-time, without human intervention. This represents a potential danger to democratic rights and individual freedoms since governmental authorities and private enterprises can use these instruments to manipulate or control citizens. For example, miniaturized drones, equipped with cameras and advanced sensors, can monitor individuals without them being aware. The use of these technologies is justified in the name of public security, to prevent terrorism or criminality, but the cost for privacy is enormous. Such technologies can be employed to spy, intimidate, or damage physically individuals, as demonstrated by military scenarios. This pushes us to wonder how much can security justify the erosion of fundamental rights, especially in a democratic society. Many countries use AI to feed a mass surveillance system that tracks the moments and activities of citizens, eroding their freedoms. Through facial recognition, behavioral analysis, and the collection of data from technological devices, authorities can profile people, and detect behaviors that are deemed suspect or undesirable. With the expansion of the Internet of Things, the daily lives of people are increasingly more connected, and this creates new challenges: devices such as intelligent thermostats, fitness watches, autonomous cars, or vocal assistants continuously generate data on the behaviors of users, creating a constant flow of information that can be exploited for commercial uses or surveillance. The border between daily life and privacy invasion becomes labile since these devices improve the efficiency of life but at the price of a constant collection of personal data. An exemplar case is represented by the cars connected to Tesla, which collect not only data regarding driving but also personal information like contacts, music preferences, and navigation chronologies. This information is used to optimize the driving experience, but can also be shared with third parties, exposing users to potential risks of privacy violation or unauthorized use of data.

---

[109] de Hert, P., & Papageorgiou, A. (2017). *Privacy and Data Protection*. European Data Protection Law Review, 3(1), 1-9. Retrieved from https://www.switchlegal.nl/wp-content/uploads/2019/10/Privacy-and-Data-Protection-edpl-3-2017.pdf

The General Regulation on the Protection of Data (GDPR), introduced by the EU in 2018, depicts a fundamental pillar in the protection of privacy and personal data. With the advent of the digital era, the EU has tried to create a legislative framework that allows individuals to have greater control over their own data, granting fundamental rights such as the right to access, rectification, cancellation, or portability of data. However, GDPR presents criticalities in its adaptation to AI systems, especially considering their complexity and autonomy. One of the major issues regarding this regulation is the fact that it was conceived in an era where AI technologies were less pervasive compared to today. It is based on relatively traditional concepts of data treatment, where the use of personal information is generally linear and predictable. Nevertheless, modern AI systems operate in a dynamic way: they learn and evolve autonomously, often modifying their own decision-making processes and algorithms over time, based on training data[110].

This phenomenon of total control over personal data is evolving and the change is manifesting in what is defined as "personal informative sovereignty", namely the capacity of individuals to manage and control their own data. In the current context, privacy is not a binary choice anymore (total sharing or absolute confidentiality), but it is located along a sharing spectrum, that varies according to the circumstances and social interactions. Artificial intelligence, with its capacity to gather and analyze enormous quantities of data through connected devices, intelligent sensors, and pervasive technologies, introduced new levels of social surveillance. This new configuration led to an implicit privatization of surveillance since many companies and governments have access to such data and use it for various purposes[111].

4.2.4 Right to Fair Trial

Modern tribunals are afflicted by administrative use that reduces the quality of judicial services offered to citizens, such as the accumulation of cases, and the length and costs of judicial proceedings. Despite the political and economic contexts of a country that can influence significantly the functioning of tribunals, internal factors, such as the lack of judges, excessive

---

[110] Regulation (eu) 2016/679 of the european parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[111] De Hert, P., & Papageorgiou, A. (2017). *Privacy and Data Protection*. European Data Protection Law Review, 3(1), 1-9.

inactivities of tribunals, and the presence of antiquated procedural rules, contribute to the inefficiency and the duration of trials. European states dedicate part of their judicial spending to technological instruments of ICT, that automate procedures such as the management of cases, communications, and organization of hearings, to improve the efficiency of tribunals. More recently, European courts have started investing in advanced technological instruments, such as artificial intelligence, to optimize their internal trials. The AI-based applications for the administration of documents, assignation of cases to judges, and the anonymization of sentences for publication are among the most advanced ones in the context of judicial administration. AI Algorithms, that use machine learning methods, differentiate from ICT traditional software for their interactivity and adaptability, other than a certain autonomy since they can execute internal modifications without the necessity of human interventions. However, despite the integration of AI having the goal of promoting access to justice and improving administrative efficiency, matters such as the opacity of algorithms, implicit judgments, and technical vulnerabilities aroused concerns regarding the risks associated with their use. Furthermore, European states fear that the attraction to AI solutions developed in countries like the USA and China, where the investments in research and development of AI are higher, can reduce European competitiveness. The preferential adoption of AI systems developed somewhere else could generate not only an economic disadvantage for European states, but also the risk that these applications are less regulated, and therefore, more susceptible to causing damages.

The European Commission has classified some AI systems used for the administration of justice as “high risk”, due to the potential damage that they could cause to the rights of citizens, such as the right to a fair trial. However, some AI applications that assist judges in the research and interpretation of facts and laws, are not considered high risk, for example, the systems that manage the anonymization of sentences or the distribution of resources. For their integration, it is essential that AI systems respect the right to a fair trial, as mentioned in Article 6 of the European Convention on Human Rights (CEDU). This article, binding for all member states of the Council of Europe, guarantees that tribunals and courts follow the defined legal rules and procedures. The article applies both to civil and criminal proceedings, and the principle of equity must be respected in all phases of the trial. The right to a fair trial includes the fundamental principle of access to justice. AI can improve this aspect by facilitating access to accurate legal information, like in the case of chatbots that help citizens. Nonetheless, if the prevision systems of cases results discourage people from resorting to tribunals, because of systemic biases or technological issues, this might limit their right to access, mining one of the pillars of a fair trial. Article 6 requires that each process is concluded in a reasonable time. The excessive delays in proceedings compromise justice. Ai could contribute to reducing such times, by automising determined activities. Nevertheless, if the implementation is not monitored correctly or if the tools are not efficiently integrated, the promise of accelerating the proceedings risks not being realized, with consequent violation of the principle of reasonable time. Finally, a fair trial requires judges to be independent and free from external influences, including those coming from governments and private corporations. If AI used in tribunals is developed by external entities with economic or political purposes, the influence of these parties might compromise the independence of the judiciary, posing under risk the concept of impartial justice. In short, access to justice, reasonable duration, and judicial independence are three key components of the right to a fair trial, and the implementation of AI systems must be carefully regulated to ensure that this principle are not compromised, but rather improved[112].

4.2.5 Algorithmic biases

In the contemporary world, automated algorithmic decisional systems govern numerous aspects of daily life, including fields strictly linked to social inequalities such as unemployment, governmental benefits, and criminalization. Despite these systems being designed to be objective and impartial, it is widely recognized that tend to reproduce or amplify existing social inequalities, reflecting prejudices intrinsic to the society. Machine learning algorithms, which dominate the current AI overview, are trained on data that reflect assumptions and human priorities, and when such data are influenced by structural inequalities, these are automatically and inevitably codified in the automated decision-making processes. This creates vicious cycles where automated decisions perpetuate inequality models already present in society. Researchers and AI developers are trying to face the issue of reducing biases in datasets and algorithmic processes, however, it is a complex challenge that cannot be solved only with technological solutions. Bias matters in these systems is not something new for the disciplines that study the reality between technology and social structures, but sociology has until now remained at the margin of the debate. The sociological experience in the analysis of social stratifications, power, and oppression, is now seen as essential to understanding and solving inequalities that automation risks further perpetuating. The AI developers

---

[112] Kterzidou, M. (2020). *Fair trial: A comparative analysis of the right to a fair trial in Europe and the United States*. Journal of Jurisprudence, 31(3), 1-26.

started recognizing the necessity to integrate competencies coming from social sciences to deal more efficiently with these issues.

The analysis of the concept of bias in AI puts us in front of deep questions that intertwine with social inequalities. Sociology has long tried to enlighten the structural inequalities present in modern societies, highlighting how these are reflected also in media and technologies. However, for AI developers, the comprehension of social injustices represents a conceptual challenge for which they are not ready. The problem of algorithmic partiality, often manifested in racist or sexist results, made necessary the development of technical solutions to mitigate the problem. Although attempts to reduce or eliminate biases in AI models, there doesn't exist a consensus over what makes an algorithm truly equal and impartial, or if this is even an achievable goal. In the field of AI, the concept of "Ground truth" represents the reality that exists out of the model. Often, the bias is defined as a deviation from this truth, with algorithms that produce imprecise or misleading results. *But what happens if the truth itself is impregnated with structural inequalities*? Traditionally, the AI bias has been treated as a problem of accuracy: an algorithm that reproduces faithfully the input data is considered free from partialities. The important question here would be whether the existing algorithms limit themselves to accurately reflect the inequalities, or whether they should be designed to change them. In other words, it is not only a matter of eliminating technical errors or improving the precision of a model, but also deciding if and how the algorithms should face and correct social injustices[113].

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) case, already mentioned, is one of the most known cases of algorithmic bias in the American judicial system. COMPAS is a risk evaluation tool used to predict the probability that an accused person will commit again a crime (relapse). It is used in different tribunals in the US to support decisions regarding cautions, sentences, and probation. It is a proprietary system developed by Northpointe that generates risk scores based on variables linked to the accused, among which are age, gender, criminal precedents, work status, and some answers to questionnaires. The scores influence the way judges and functionaries evaluate the future dangerousness of an individual. In 2016, an investigation led by the newspaper website ProPublica, enlightened issues of racist bias in the COMPAS system. The investigation revealed that the algorithm tended to overestimate the relapse

---

[113] Eubanks, V. (2020). *AI and Inequalities*. Sociological Forum, 35(3), 759-775. doi:10.1111/soc4.12962. Retrieved from https://compass.onlinelibrary.wiley.com/doi/epdf/10.1111/soc4.12962

risk when involving black people and underestimate the risk for white people. According to ProPublica, the Afro-American accused were doubly prone to be erroneously classified as high risk of relapse. The 45% of black people labeled as "high-risk" have not relapsed, compared to the 23% of white people. White individuals, on the other hand, were often labeled as low risk even though they had a major probability of relapse compared to black people. The thesis present in COMPAS is not directly attributable to an explicit characterization based on race, since the model does not use ethnic or racial variables. However, some of the variables used (the socio-economic context and the areas of residence) are connected to race because of the structural inequalities present in American society. This generates an indirect bias, where race is not an explicit factor, but it becomes so through other correlated variables. In other words, the algorithm, based on previous and historical data, reflects and amplifies the existing inequalities, producing results that perpetuate the racial prejudices already existing in the criminal justice system. The report of ProPublica initiated a heated academic and public debate. Northpointe, the company that developed COMPAS, contested the conclusions, affirming that the system was designed to be equal and that, measuring other accuracy scales (such as the general accuracy of the model), would not lead to a significant difference between white and black people. Nonetheless, the experts have underlined that the scale used to measure equity is central: while Northpointe focused on comprehensive accuracy, ProPublica focused on false negatives and false positives, highlighting significant racial disparities[114].

4.3 The Ethics of Counterterrorism

Counterterrorism professionals regularly deal with ethical challenges that require balancing out contrasting moral values such as security, privacy, and human rights. These situations imposed careful evaluation of the consequences of each action, in order to guarantee the respect of well-being and rights of the people involved. Despite the importance of ethics in the decision-making process in counterterrorism, there exists a lack of structural methods that can support professionals in this complex task. Recently, different research efforts have been started to fill this gap, developing tools and guidelines to help professionals navigate between moral dilemmas and conciliate often contrasting values that emerge in counterterrorism. These initiatives highlight the

---

[114] Angwin, J., Larson, J., Mattu, K., & Kirchner, L. (2016). *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

necessity to provide practical instruments that keep into consideration ethics as the core of the decision-making process.

In the context of counterterrorism, the use of artificial intelligence introduces a new dimension of secrecy of crucial importance. The anti-terrorism operations are increasingly based on advanced technologies for the analysis of data and surveillance, but this secrecy can isolate professionals and impede them from freely discussing ethical matters. The AI technologies can operate in a black box, making it difficult to evaluate the implications of decisions, and limiting the capacity of exchanging experiences and best practices of ethics among professions and institutions. The lack of open discussion and a support network makes it difficult to deal with ethical issues associated with the use of AI in counterterrorism. The nature of terrorist attacks is characterized by low frequency but high impact, an aspect that poses significant ethical challenges in the use of AI. Security agencies can feel pushed to implement rigid security measures to prevent attacks, sometimes sacrificing civil rights in the name of the protection of society. This tension between secrets and freedoms is one of the most debated topics in the ethics of counterterrorism[115]. Decisions in counterterrorism scenarios must be taken rapidly which further complicates the use of AI. Decisions in counterterrorism scenarios must be taken rapidly which further complicates the use of AI. Algorithms can provide information in real-time, but their reliability can lead to erroneous conclusions if data are distorted or incomplete. In high-pressure situations, the risk of judgment increases, and consequences if decisions can have a direct impact on people’s lives and their rights.

The use of AI in counterterrorism operations has often led to significant violations of human rights, both internally to nations and in military operations abroad. At the national level, many western democracies have adopted surveillance methods based on AI to monitor suspected activities. These tools have raised many concerns: a notable example are the mistakes in facial recognition systems, that have demonstrated a racial prejudice, increasing the risk of unjustified arrests and abuses, especially toward ethnic minorities[116]. Ai can also lead to mass surveillance, violating privacy rights

---

[115] Kowalski, M. (2020, March 12). *Ethics on the radar: exploring the relevance of ethics support in counterterrorism*. Retrieved from https://hdl.handle.net/1887/86282

[116] Amnesty International. (2020). *Public statement: Amnesty International calls for a ban on the use of facial recognition technology for mass surveillance*.

of citizens: for instance, the “dragnet surveillance”[117] allows authorities to collect data on individuals that are not suspected of any crime, violating the principle of proportionality. This type of surveillance moves away from a vision where civil freedoms are protected, creating an atmosphere of general suspicion that canned up eroding democratic liberties. At the international level, the employment of AI in military operations of contralto to terrorism, such as autonomous drones to hit terrorist targets, represents an ulterior delicate matter[118]. These automated systems, capable of making decisions without the direct human intervention, created enormous doubts from the ethical perspective. In particular, there is the risk that such operations lead to civil victims, creating what is defined as “*collateral damage*”. Often, indeed, the algorithms that control these drones are based on incomplete or imprecise data, that can induce to the wrong identification of targets. These actions not only violate the international humanitarian law, but canals feed hate sentiments and radicalisation in the hit areas, mining the foundational goals of counterterrorism. Moreover, the counterterrorism operations that use AI abroad can violate the rights of foreign citizens through unauthorised surveillance operations or preemptive attacks in sovereign states. These actions destabilise the principles of national sovereignty and put under discussion the respect for international laws on human rights. Countries involved in conflicts with terrorist groups, often justify the use of extraordinary measures for national security reasons, but this can translate to serious violations, such as arbitrary detentions, torture and lack of fair trials for the suspects[119].

4.4 Algorithmic authoritarianism

The analysis of algorithmic authoritarianism is becoming a topic of growing interest, not only in authoritarian regimes but also in democracies, highlighting a concerning global tendency. The countries that are labeled as concerning cases, exercise significant control over digital feeds, from internet services providers to online platforms, sustained by wider surveillance systems. Understanding these dynamics becomes essential to decipher mechanisms and the sector of authoritarianism and repression based on algorithms and artificial intelligence. Another crucial

---

[117] Slobogin, C. (2010). G*overnment dragnets*. Law and Contemporary Problems, 73(3), 107–143. http://www.jstor.org/stable/25766402

[118] Human Rights Watch. (2012). *Drone strikes: A review of the U.S. government's use of drones for targeted killings*.

[119] Töws, M. (2023). *Artificial Intelligence and Human Rights: A Human-Centric Approach to AI Governance*. Retrieved from https://tobias-lib.ub.uni-tuebingen.de/xmlui/bitstream/handle/10900/65097/2301-volledige-tekst_tcm45-535187.pdf?sequence=1&isAllowed=y

aspect is the use of unregulated advanced technologies (or under-regulated) by authoritarian states, which employ practices of repression and suppression of information from foreign countries. Over the last ten years, the use of emerging technologies by authoritarian regimes for the manipulation and interference of foreign information and repression, has been widely documented, capturing the attention of the EU. The legal and normative structures of these nations not only habilitate but also legitimize the use of repression, control, and censorship algorithms. This includes legislation regarding cybersecurity, information control, and national security, creating a legal context for such practices. In many of these authoritarian regimes, AI and algorithms are used deliberately in an unregulated way, justified with the excuse of fighting terrorism or limiting radicalization. These processes are spread in Russia, China, and Iran but also Middle Eastern, north African, and sub-Saharan contexts, in addition to some countries in South America. In conclusion, the analysis of algorithmic authoritarianism offers an overview of the current and future challenges to democracy and human rights. It is essential that democracies work to protect their own institutions and guarantee that emerging technologies are used ethically and responsibly, protecting the fundamental rights of individuals in an ever-changing world[120].

4.4.1 (Un)ethical Counterterrorism: the Case of China and the Region of Xinjiang

One of the most complicated and well-documented cases of the use of technology based on AI for algorithmic authoritarianism is the case of Xinjiang, in China, and the prosecution of the Muslim minority of Uyghur. Chinese authorities have implemented a wide apparatus of surveillance that uses technologies such as facial recognition and the predictive police to monitor, control, and hold the majority of the Uyghur population. The autonomous region of Xinjiang, inhabited mainly by Uyghurs, has a historical and strategic relevance for China, also thanks to its natural resources. However, in the last decades, the afflux of Han-ethnicity settlers has increased tensions in the region. At the core of the conflict, there are religious restrictions, economic inequalities, and attempts from the minority to maintain their own cultural identity. Until 1949, the Chinese government pursued cultural and economic assimilation policies, perceived by the minority as a threat to their cultural inheritance. These tensions have been further complicated by the influence of global jihadist movements, that have found support in the region, pushing Beijing to adopt even

---

[120] European Parliament. (2024). *The impact of artificial intelligence on human rights* [Study]. Retrieved from https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf

stricter security measures. Beijing's strategy on Xinjiang has passed from mere containment to the active suppression of information, with the goal of managing meticulously biometric and personal data of the citizens of the region. For such purposes, Chinese authorities have decided to exploit some of the most advanced AI applications, testing and widening the capacities of automated surveillance. The adoption of such technologies is the result of an accurate planning that unites two thought positions inside the Chinese leadership: on one side, the global growth of AI has been seen as an irresistible tool to obtain surveillance capacities with no precedents; on the other, the unique challenges in the Xinjiang have required an invasive, discreet and preemptive solution. In this context, AI has not been adopted as an impulsive decision, but as a strategic choice aiming to integrate traditional surveillance with an advanced predictive police program. This program is based on the use of a wide range of data collection infrastructures that comprehend biometrics and behavioral data, feeding algorithms capable of identifying potential dissidents before perpetrating explicit acts of resistance. This has marked a turning point in the surveillance politics of China, marking the beginning of an era where algorithmic governance started taking priority over traditional mechanisms based on human resources[121].

Consequently, numerous initiatives for data collection have been launched. A significant example is the program Sharp Eye, launched in 2015, which widened the initiative Skynet and started in 2005 for urban surveillance. Sharp Eyes exploits a wide range of data sources, among which surveillance cameras, cameras for the recognition of vehicles and plates, together with virtual identities like MAC and telephone numbers. These data are integrated through geographical informative systems and sent to platforms of social resource integrations in various provinces, including Xinjiang. Data fusion programs in China target specific social groups, especially those considered "personnel at risk", such as individuals who present petitions to the government, are involved with terrorism, or are considered threats to social stability. The Uyghur ethnic minority in Xinjiang is subjected to intense surveillance through these programs. Tools like the platform for integrated joint operations (IJOP) in Xinjiang connect the identity cards emitted by the government to the physical characteristics of individuals, monitoring behaviors deemed indicators of potential social instability. Furthermore, Chinese laws impose the cooperation between private enterprises and state security institutions. Among these laws, there is the law on cybersecurity 2016, the Law on national

---

[121] European Parliament. (2024). *The impact of artificial intelligence on human rights* [Study]. Retrieved from https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf

intelligence 2017, and the law on data security 2021. This environment of growing rigidity and centralization places a significant emphasis on political stability and requires the sharing of data with governmental authorities. Daily movements of residents have become part of a systemic surveillance ritual, characterized by frequent verifications at checkpoints or data collection stations[122]. These stations have been created to gather personal data, generally through the scan of identity documents, facial recognition, and inspection of personal communication devices. These checkpoints carry out two functions: underline the constant presence of state surveillance and collect detailed data necessary for advanced systems of predictive police. The algorithms of the predictive police have been developed directly from this collection of data and have facilitated the detection of models that indicate potential dissidents or non-conformities. Their goal is not to deal with already perpetrated crimes, but to predict potential threats to security, marking a turning point compared to traditional police methods, toward a governance model focused on the management and preemptive mitigation of risks. Because of this security strategy based on data, many Uyghurs have been captured in the system of predictive police and subsequently confined in "re-educational camps". Through a combination of methods and constant surveillance, the re-educational goal is composed of two phases: firstly, isolating and erasing opposition to the state; secondly, coercively transforming these individuals into productive subjects[123]. These detentions have been characterized by a lack of transparency, often carried out without formal accusations or legal proceedings, only based on ambiguous results of the AI systems' analysis. The constant surveillance has created a sense of constant observation inside the Uyghur minority, leading to a spreading phenomenon of auto-censorship and behavioral changes. The daily practices have been interrupted and conversations have become more cautious, with people adapting to unspoken boundaries imposed by the surveillance system. This had a huge psychological impact on citizens since the simple awareness of being observed has altered fundamentally the dynamics of the community.

A fundamental aspect in China is the social credit system (SCS), misunderstood as a single monolithic score for every citizen, it is, in reality, an articulated network of interconnected political and technological initiatives, finalized to shape individual and corporate behaviors. At its core, it

---

[122] Roche, G. and Leibold, J. (2022), *State Racism and Surveillance in Xinjiang (People's Republic of China)*. The Political Quarterly, 93: 442-450. https://doi.org/10.1111/1467-923X.13149

[123] Stefanie Kam, Michael Clarke, *Securitization, surveillance and 'de-extremization' in Xinjiang*, International Affairs, Volume 97, Issue 3, May 2021, Pages 625–642, https://academic.oup.com/ia/article/97/3/625/6219662?login=false

gathers social, behavioral, and digital data, combined with AI, to classify citizens and enterprises, rewarding or punishing them based on various parameters. It is composed of numerous pilot projects managed by municipalities and private companies, with the intention of integrating them into a cohesive national framework. Each project faces specific social and economic behaviors. For example, some programs monitor financial credibility, track reimbursements of loans, or financial fraud. In SCS, government and enterprises operate in a complementary way. The government relies on the innovation of the private sector to deal with enormous datasets, whereas companies benefit from the legitimacy and normative framework provided by the state. This symbiosis aims to create a more efficient and reactive social credit system. Private companies own vast quantities of data on consumers' behaviors and, developing their own credit systems, not only satisfy their own commercial interests but also contribute to the national initiative of social credit. The collection of data goes beyond public registers and it includes financial transactions, medical records, occupational status, and conformity to civil duties. This information offers a multidimensional profile of the public, private, and financial life of each citizen. Citizens can contribute with data, both voluntarily or as part of compulsory interactions with governmental services. For instance, the presentation of information to obtain licenses, social services, or participation in community activities can be tracked and considered in the system. The surveillance extends also to public transport, where systems monitor in real-time the conformity of individuals to laws. The key to this system's influence resides in its capacity to correlate data coming from various sources. The online behavior of a person can be correlated to the physical activities registered to provide a holistic vision. The integration of data involves the use of sophisticated algorithms and constantly refined machine learning mechanisms, which analyze large sets of data to identify patterns, make predictions, and generate scores[124].

As a consequence of this digital authoritarianism, a growing economic market of surveillance technologies has emerged. Chinese enterprises specialized in advanced hardware and software for surveillance, such as Hikvision and Dahua Technology, have found themselves at the heart of a flourishing internal market. These enterprises have obtained substantial profits thanks to the government's contacts, and their technologies have become a symbol of the state's capacity to control and deal with its own citizens. The capacity of the implementations in Xinjiang has

---

[124] K. L. X. Wong and A. S. Dobson, *We're just data: Exploring China's social credit system in relation to digital platform ratings cultures in Westernised democracies'*, Global Media and China, Vol 4, No 2, 2019, pp. 220-232.

projected these enterprises to the vertex of the security of global surveillance technology, despite being object to critiques and international sanctions. The Chinese government, in its internal narrative, has justified the measures of surveillance and re-education in Xinjiang as necessary to fight extremism and promote economic development. This perspective has been widely and generally accepted by the majority of Chinese Han, partly thanks to the significant control of the government on the internal informatics environment. State media have emphasized the development of infrastructures and the creation of workplaces in the region, contributing to reinforcing the sustainment of governmental policies among the general population. The discussion on the effects of these policies on the Uighur population and on other ethnic minorities has been largely repressed, leading to a distorted public understanding inside China[125]. On the other hand, the international response has been marked by strong critiques. Journalistic investigations and campaigns of human rights organizations have enlightened the conditions inside these camps and the broad surveillance apparatus, arousing a strong condemnation at the international level. Several Western governments, international organizations, and activist groups have labeled China's actions as a serious violation of human rights. This led to the imposition of sanctions against Chinese functionaries and technology enterprises involved in the surveillance and repression in Xinjiang[126]. Despite the pressure and international control, the sophisticated surveillance network of Xinjiang persists. What is even scarier, is that the technology developed and refined in Xinjiang is not only still in use, but it is also commercialized in other countries. This raises concerns about the export of such surveillance capacities and the potential of other governments to adopt similar methods of social control.

A fundamental aspect in China is the social credit system (SCS), misunderstood as a single monolithic score for every citizen, it is in reality an articulated network of interconnected political and technological initiatives, finalized to shape individual and corporate behaviors[127]. At its core, it gathers social, behavioral, and digital data, combined with AI, to classify citizens and enterprises,

---

[125] Stefanie Kam, Michael Clarke, *Securitization, surveillance and 'de-extremization' in Xinjiang*, International Affairs, Volume 97, Issue 3, May 2021, Pages 625–642, https://academic.oup.com/ia/article/97/3/625/6219662?login=false

[126] R. Roberts, '*The biopolitics of China's "war on terror" and the exclusion of the Uyghurs*', Critical Asian Studies, Vol 50, No 2, 2018, pp. 232-258. https://es.scribd.com/document/404121659/The-Biopolitics-of-China-s-War-on-Terror-and-the-Exclusion-of-the-Uyghurs

[127] F. Liang, V. Das, N. Kostyuk, and M. M. Hussain, '*Constructing a data-driven society: China's social credit system as a state surveillance infrastructure*, Policy & internet, Vol 10, No 4, 2018, pp. 415-453; D. Mac Sithigh and M. Siems, 'The Chinese social credit system: A model for other countries?', The Modern Law Review, Vol 82, No 6, 2019, pp. 1034-1071.

rewarding or punishing them based on various parameters. It is composed of numerous pilot projects managed by municipalities and private companies, with the intention of integrating them into a cohesive national framework. Each project faces specific social and economic behaviours. For example, some programs monitor financial credibility, tracking reimbursements of loans or financial fraud. In SCS, government and enterprises operate in a complementary way. The government relies on the innovation of the private sector to deal with enormous datasets, whereas companies benefit from the legitimacy and normative framework provided by the state. This symbiosis aims to create a more efficient and reactive social credit system. Private companies own vast quantities of data on consumers' behaviours and, developing their own credit systems, not only satisfy their own commercial interests but also contribute to the national initiative of social credit. The collection of data goes beyond public registers and it includes financial transactions, medical records, occupational status, and conformity to civil duties. This information offers a multidimensional profile of the public, private, and financial life of each citizen. Citizens can contribute with data, both voluntarily or as part of compulsory interactions with governmental services. For instance, the presentation of information to obtain licenses, social services, or participation in community activities can be tracked and considered in the system. The surveillance extends also to public transport, where systems monitor in real-time the conformity of individuals to laws. The key to this system's influence resides in its capacity to correlate data coming from various sources. The online behavior of a person can be correlated to the physical activities registered to provide a holistic vision. The integration of data involves the use of sophisticated algorithms and constantly refined machine learning mechanisms, which analyze large sets of data to identify patterns, make predictions, and generate scores[128].

---

[128] K. L. X. Wong and A. S. Dobson, We're just data: Exploring China's social credit system in relation to digital platform ratings cultures in Westernised democracies', Global Media and China, Vol 4, No 2, 2019, pp. 220-232.

CONCLUSIONS

In conclusion of this thesis, it is fundamental to understand not only the technological evolutions of artificial intelligence, but also its deep strategic and social impacts on sectors, and the growing dependence on it. Over the last years, and still today, AI has become an essential resource, present in every aspect of our daily life, from the management of financial flows to healthcare assistance, public security, and the functioning of critical infrastructures. Its pervasiveness has reached a level never seen before, transforming radically the way people interact with technologies and daily administration tools. Indeed, AI systems have already been integrated into the decision-making process of enterprises, governmental institutions, and even in the daily lives of citizens, through virtual assistants, personalized recommendations, and health monitoring systems, to name a few examples. This growing integration of AI is not destined to cease; on the contrary, it is expected to expand and become more and more sophisticated. The forecast is that, over the next years, our interaction with this technology will become more intimate and intrusive. However, while offering numerous advantages, its pervasiveness raises doubts about how these capacities are integrated into social and institutional structures. We are facing, therefore, a big paradox: on one hand, the easiness, and efficiency that AI brings into our lives, on the other hand, the growing vulnerability to which we expose ourselves. The pervasiveness of AI implicates, for the first time, a big part of our lives is mediated by algorithms that do not always understand and that can influence our choices in unpredictable ways.

In the reflection of the efficiency of AI in contrasting terrorism, it is evident that the integration of these advanced technologies has led to significant and tangible results in the sector, as indicated in many reports and studies, that underline how the implementation of advanced systems of data analysis and machine learning have allowed the identification and prevention of terrorist attacks in real-time, demonstrating that the AI can operate as a strategic ally in the fight against terrorism. The predictive analysis technologies, indeed, are capable of monitoring accurately large-scale suspected communications, abnormal behaviors, and risk models. This offers security agencies an unprecedented tool of predictive analysis, that can turn out to be crucial for the management of threats and prevention of attacks, revealing a significant potential in the improvement of public security. The integration of technologies, not only facilitates the gathering and interpretation of data, but it also allows to generate more prompt and targeted responses to situational crises. Additionally, the Artificial Intelligence Prediction and Counterterrorism paper by the Chatham House[129] provides further insights on the role of AI in contrasting terrorism. The report shows how the pogress of automated analysis has made possible a more efficient identification of suspects and illicit activities. Sophisticated algorithms of facial recognition and natural language processing, widely used for digital surveillance, have improved notably the capacities of armed forces to respond to increasingly sophisticated global threats. However, while these tools represent a precious resource for public security, they are not free from critiques and ethical concerns.

As underlined in the report AI100 by Stanford[130], this technological evolution requires an in-depth reflection on how an increasingly AI-dominated social structure is being built. It is not only a matter of technological adaptability, but also a redefinition of relations among citizens, institutions, and technologies themselves. With the growth of AI dependency, societies must prepare for a reality where intelligent technologies are not only simple tools anymore, but active actors in daily dynamics. The consequences of such infiltration demand particular attention to the governance of technologies, to guarantee that the AI operates as an extension of the human capacities, rather than as an autonomous force that defines our lives. The necessity of establishing clear norms and standards for the use of AI has become, therefore, a crucial matter, not only to protect the privacy and safety of citizens but also to ensure that technology advances responsibly, preserving fundamental ethical values. This collective responsibility is essential to navigate in the future, in a way that is more and more shaped by the massive and indiscriminate adoption of AI. In this context, it is interesting to reflect on the movie "Her", which offers a dramatic and extremely current depiction of the pervasiveness of AI in our lives. The plot develops around the main character who develops a romantic relationship with an operative system equipped with artificial intelligence, whose voice, although lacking a physical body, becomes a fundamental element of emotional connection. The movie, despite being extremely narrative, enlightens a possible unsettling direction that society could move toward AI, in its attempt to facilitate daily life, might end up isolating individuals, creating a sort of dependency that substitutes the genuine human interactions with artificial experiences. Although the film represents an extreme case, it reflects a growing concern:

---

[129] Chatham House. (2019). *Artificial intelligence, prediction and counterterrorism*. Retrieved from https://www.chathamhouse.org/sites/default/files/2019-08-07-AICounterterrorism.pdf

[130] Stanford University. (2016). *One hundred year study on artificial intelligence: Report of the 2016-2017 study panel*. Retrieved from https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf

alongside the integration of technology into our existence, there is the risk of people being stuck in unidirectional relations with machines, compromising the quality of human interactions. Ai represents not only a risk for the abuses by governments and corporations but constitutes Also an intrinsic threat to individuals themselves. Its ubiquity and the power it confers can lead to situations where technology, instead of being a simple tool, becomes an agent of control and manipulation, amplifying vulnerabilities and social inequalities.

The question of whether Artificial intelligence represents an advantage for security or a menace to humanity raises deep concerning and controversial interrogatives that are worth a careful analysis. The question highlights a crucial conflict in today's society, as governments rely more on AI systems for numerous sectors, but especially national security, potentially jeopardizing basic freedoms and rights. The widespread use of AI in surveillance and law enforcement prompts ethical questions about how governments might use these technologies to oversee and regulate their populations. The use of algorithms by security agencies in predictive policing increases the risks of bias and discrimination, resulting in individuals being unjustly targeted due to flawed data and algorithmic analysis, ultimately infringing on their rights and liberties. Additionally, the incorporation of AI in military activities brings about additional complications. Governments may see autonomous systems as a way to improve strategic advantage, but relying on machine decision-making in warfare raises important ethical concerns about responsibility and the importance of human life. As AI technologies progress, the delicate balance between ensuring security and protecting human rights becomes more challenging. The danger of such systems being abused for authoritarian reasons is significant: a government focuses on monitoring in the name of protection risks creating an atmosphere of fear and suspicion among its citizens. Moreover, the swift progress of AI may result in a type of societal regulation that exceeds conventional methods. Citizens may sense the burden of continuous scrutiny, restricting their ability to freely express themselves and engage in social and political activities. The potential impact of intelligent systems monitoring individuals could cause them to self-censor, which could weaken democratic involvement and civic participation. The question that comes up is: *when does the pursuit of security start infringing on the rights it aims to safeguard?* In the end, as artificial intelligence continues to evolve, its true legacy in counterterrorism will not be measured solely by its technological achievements but by the human values it upholds and the trust it fosters.

BIBLIOGRAPHY

Albrecht, J. P., & Giebel, E. (2020). *The Impact of AI on Human Rights, Democracy, and the Rule of Law*. Retrieved from https://allai.nl/wp-content/uploads/2020/06/The-Impact-of-AI-on-Human-Rights-Democracy-and-the-Rule-of-Law-draft.pdf

Amnesty International. (2020). *Public statement: Amnesty International calls for a ban on the use of facial recognition technology for mass surveillance*.

Angwin, J., Larson, J., Mattu, K., & Kirchner, L. (2016). *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Association for Computing Machinery. (n.d.). *John McCarthy - A.M. Turing Award winner. ACM Turing Award*. Retrieved from https://amturing.acm.org/award_winners/mccarthy_1118322.cfm

Bharadiya, J. (2023). Artificial intelligence in transportation systems: A critical review. ResearchGate. Retrieved from https://www.researchgate.net/publication/371282928_Artificial_Intelligence_in_Transportation_Systems_A_Critical_Review

Biometrics Institute. (n.d.). What is biometrics? FAQs. Retrieved October 7, 2024, from https://www.biometricsinstitute.org/what-is-biometrics/faqs/

Bora, M., Kumar, K., Kaur, A., & Sonkar, R. (2021). Crowd abnormal behaviour detection using deep learning. Journal of Visual Communication and Image Representation, 78, 103091. https://doi.org/10.1016/j.jvci.2021.103091

BOULANIN, V., SAALMAN, L., TOPYCHKANOV, P., SU, F., & CARLSSON, M. P. (2020). *AI and the military modernization plans of nuclear-armed states*. In Artificial Intelligence, Strategic Stability and Nuclear Risk (pp. 31–100). Stockholm International Peace Research Institute. http://www.jstor.org/stable/resrep25355.9

Boulanin, V. (Ed.). (2019). *The impact of artificial intelligence on strategic stability and nuclear risk*, Volume I, Euro-Atlantic perspectives. Stockholm International Peace Research Institute. https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Ethics and information technology, 15, 209-227.

Chaliand, G., & Blin, A. (Eds.). (2017). *The history of terrorism, from antiquity to ISIS* (Updated ed.). University of California Press. https://books.google.it/books/about/The_History_of_Terrorism.html?hl=es&id=U6swDwAAQBAJ&redir_esc=y

Chatham House. (2019). *Artificial intelligence, prediction and counterterrorism*. Retrieved from https://www.chathamhouse.org/sites/default/files/2019-08-07-AICounterterrorism.pdf

Chatham House. (2019). *AI and counterterrorism: The implications for security*. https://www.chathamhouse.org/sites/default/files/2019-08-07-AICounterterrorism.pdf

Charniak, E. (2022). *Statistical language learning*. Retrieved from https://hal.science/hal-04561644/document

Chaturvedi, A., & Mukherjee, A. (2020). *Cross-national analysis of global security discourse using word embeddings*. Retrieved from https://preprints.apsanet.org/engage/api-gateway/apsa/assets/orp/resource/item/5ecd3ff1fd965c0018b47287/original/cross-national-analysis-of-global-security-discourse-using-word-embeddings.pdf

Chen, Q., & Bridges, R. A. (2017, December). *Automated behavioral analysis of malware: A case study of wannacry ransomware*. In 2017 16th IEEE International Conference on machine learning and applications *(ICMLA)* (pp. 454-460). IEEE.

Colson, D. (2017). *Propaganda and the Deed: Anarchism, Violence and the Representational Impulse*. American Studies, 55/56, 163–186. http://www.jstor.org/stable/44982624

David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, & Morgan Briggs; With a foreword by Lord Tim Clement-Jones, *ARTIFICIAL INTELLIGENCE, HUMAN RIGHTS, DEMOCRACY, AND THE RULE OF LAW*, May 2024. The Alan Turing Institute. https://edoc.coe.int/en/artificial-intelligence/10206-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-a-primer.html

Davis, A. L. (2021). *Artificial Intelligence and the Fight Against International Terrorism*. American Intelligence Journal, *38*(2), 63–73. https://www.jstor.org/stable/27168700

Daza, R., Guillen, L., & Tovar, E. (2022). *Artificial intelligence in the fight against terrorism: A systematic literature review*. IEEE Access, 10, 2345–2362. https://ieeexplore.ieee.org/document/10328769

De Hert, P., & Papageorgiou, A. (2017). *Privacy and Data Protection*. European Data Protection Law Review, 3(1), 1-9. Retrieved from https://www.switchlegal.nl/wp-content/uploads/2019/10/Privacy-and-Data-Protection-edpl-3-2017.pdf

Devitt, A., & Ahmad, K. (2013). Is there a language of sentiment? An analysis of lexical resources for sentiment analysis. Language Resources and Evaluation, 47(2), 475-511. http://www.jstor.org/stable/42636377

Doe, J., & Smith, A. (2024). *Deep learning approaches to artificial intelligence ethics*. arXiv. https://arxiv.org/abs/2402.06075

EITCA Academy. (n.d.). *Quali sono i principali risultati di DeepMind's AlphaGo, AlphaZero e AlphaFold e come dimostrano il potenziale del deep learning in diversi ambiti?*. EITCA Academy. Retrieved from https://it.eitca.org/intelligenza-artificiale/eitc-ai-adl-deep-learning-avanzato/introduzione-eitc-ai-adl-deep-learning-avanzato/introduzione-ad-approcci-avanzati-di-machine-learning/ripasso-dell%27esame-introduzione-agli-approcci-avanzati-di-machine-learning/quali-sono-i-principali-risultati-di-deepminds-alphago-alphazero-e-alphafold-e-come-dimostrano-il-potenziale-del-deep-learning-in-diversi-ambiti/

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 24 September 2024 on Artificial Intelligence*. Official Journal of the European Union. Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Parliament. (2020, August 27). *What is artificial intelligence and how is it used?*. European Parliament. Retrieved from https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used

European Union. (2021). *Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online*. Official Journal of the European Union, L 172, 79–109. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2021:172:FULL&from=EN

Eubanks, V. (2020). *AI and Inequalities*. *Sociological Forum, 35*(3), 759-775. Retrieved from https://compass.onlinelibrary.wiley.com/doi/epdf/10.1111/soc4.12962

European Parliament. (2024). T*he impact of artificial intelligence on human rights.* [Study]. Retrieved from https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.https://books.google.it/books?hl=it&lr=&id=pn4pDwAAQBAJ&oi=fnd&pg=PP10&dq=virginia+Eubanks+automating+inequality:+how+high+tech+tools+profile,+police+and+punish+the+poor&ots=gF_SKheouk&sig=xWdNtdfQCLECidKyCO0aKn3V7fA&redir_esc=y#v=onepage&q=virginia%20Eubanks%20automating%20inequality%3A%20how%20high%20tech%20tools%20profile%2C%20police%20and%20punish%20the%20poor&f=false

Fadlalla, A., & Lin, C.-H. (2001). *An Analysis of the Applications of Neural Networks in Finance*. Interfaces, 31(4), 112–122. http://www.jstor.org/stable/25062724

FORVIS Mazars. (n.d.). *EU AI Act: Different risk levels of AI systems*. FORVIS Mazars. Retrieved from https://www.forvismazars.com/ie/en/insights/news-opinions/eu-ai-act-different-risk-levels-of-ai-systems

Fuchs, C., & Horak, E. (2021). *Social media: A critical introduction*. Retrieved from https://s3.amazonaws.com/document.issuu.com/210712082405-72d2016d2af24b239b48ebea8af8c169/original.file?AWSAccessKeyId=ASIATDDRE5J7Q7OKSFIZ&Signature=Xv7Wpk4rDZ7XhNZkyGclYtKPFlQ=&x-amz-security-token=FwoGZXIvYXdzEKz//////////wEaDNNM0SrA59eXHRr3xCLkAtzEBERNqyZOP4JNB8CDEeLew+Rt/KCMO7ORaYC97hwOEczNWkfC/cViR+DSXs4SobhYvHtbajxgmiAZ+AwxkvpA1OnCtZkgFB3jhKUUrJuy6LqTIJLXjnvPUGRXEoWWANqQP6kOnpCEKLcv/KT/CmY6GlOYALdWsqutXoYzvghie4gXJAInXu3/fVHPhvAmLOBmTSyTpe233KM9ro6pnYBsuLMmcjg9A9v6UdRJ5PQwkVjA94MWKhbpNJfi8WiHP1KoCQ2AD9O6yfLxOjWLzThtBDsjj0a/6RASfnZAK+bDidufxKyMMyy1ujtX+YAStJ0OG1Mhb9rAti6zOy/+6IaOGTzjF0G+N+8uE4B0ORVtjRd8+ucjL2jjxs0mNz3bpxYTWDKaKqGW2vacKgsZobTwWClI9BTDawEMoYUgv603I1QS5hDOkClIJeNtInbVAp3wajBxtZ5LHPqn89UaFw73jhoGKMPxurgGMim4uVTuymwhxYwh10og3X3BGuLXwRvqS09xcAL+comFh79Ny305FG5y0A==&Expires=1729021911

GNET. (2020). *Artificial intelligence and countering violent extremism: A primer* (V2). https://gnet-research.org/wp-content/uploads/2020/10/GNET-Report-Artificial-Intelligence-and-Countering-Violent-Extremism-A-Primer_V2.pdf

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . & Bengio, Y. (2014). *Generative adversarial nets*. Advances in Neural Information Processing Systems, 27. Retrieved from https://arxiv.org/pdf/1406.2661

Greiman, V. (2021). *Human Rights and Artificial Intelligence: A Universal Challenge*. Journal of Information Warfare, 20(1), 50–62. https://www.jstor.org/stable/27036518

Harmanpreet Kaur, et al. (2020). *Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning*. CHI 2020 Paper. Accessible at http://www-personal.umich.edu/~harmank/Papers/CHI2020_Interpretability.pdf

B. HOFFMANN, J. Ware, (2020, September 24). *Challenges for effective counterterrorism intelligence in the 2020s*. Lawfare. https://www.lawfaremedia.org/article/challenges-effective-counterterrorism-intelligence-2020s

Human Rights Watch. (2012). *Drone strikes: A review of the U.S. government's use of drones for targeted killings*.

ICCT. (2023). *Evolutions in counter-terrorism: The role of artificial intelligence*. International Centre for Counter-Terrorism. https://www.icct.nl/icct-journal-special-edition-evolutions-counter-terrorism

International Organization for Standardization. (n.d.). *Artificial intelligence - Machine learning*. ISO. Retrieved October 2, 2024, from https://www.iso.org/artificial-intelligence/machine-learning

H. Jain, A. Vikram, Mohana, A. Kashyap and A. Jain, "*Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications*," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)

JavaTpoint. (n.d.). *Application of AI*. JavaTpoint. Retrieved from https://www.javatpoint.com/application-of-ai

Johansson, F. (2014). *Detecting linguistic markers for radical violence in social media*. ResearchGate. https://www.researchgate.net/publication/260775393_Detecting_Linguistic_Markers_for_Radical_Violence_in_Social_Media

Jurafsky, D., & Martin, J. H. (2019). *Speech and language processing*. Retrieved from https://web.stanford.edu/~jurafsky/slp3/old_oct19/15.pdf

S. Kam, Michael Clarke, Securitization, surveillance and 'de-extremization' in Xinjiang, International Affairs, Volume 97, Issue 3, May 2021, Pages 625–642, https://doi.org/10.1093/ia/iiab038 https://academic.oup.com/ia/article/97/3/625/6219662?login=false

Korkmaz, O. (2024). *Artificial intelligence and counterterrorism*. SETA Foundation. https://www.setav.org/en/assets/uploads/2024/04/P73En.pdf

Kterzidou, M. (2020). *Fair trial: A comparative analysis of the right to a fair trial in Europe and the United States*. Journal of Jurisprudence, 31(3), 1-26.

Kowalski, M. (2020, March 12). *Ethics on the radar: exploring the relevance of ethics support in counterterrorism*. Retrieved from https://hdl.handle.net/1887/86282

Kumar, A. (2022, October 27). *AI behind AlphaGo: Machine learning and neural network*. USC Institute for Creative Technologies. https://illumin.usc.edu/ai-behind-alphago-machine-learning-and-neural-network/

Larsonneur, L. (2023). *Terrorism: Overview of the evolving threat and NATO's role in countering it*. NATO Parliamentary Assembly. https://www.nato-pa.int/download-file?filename=/sites/default/files/2023-01/014%20DSCTC%2022%20E%20rev.%202%20fin%20%20-%20TERRORISM%20-%20LARSONNEUR%20REPORT.pdf

Lawrence Livermore National Laboratory. (n.d.). *The birth of artificial intelligence* (AI) research. LLNL. Retrieved from https://st.llnl.gov/news/look-back/birth-artificial-intelligence-ai-research

Yann Lecun, Yoshua Bengio, Geoffrey Hinton, *Deep learning*. Nature, 2015, 521 (7553), pp. 436-444. https://hal.science/hal-04206682/document

F. Liang, V. Das, N. Kostyuk, and M. M. Hussain, '*Constructing a data-driven society: China's social credit system as a state surveillance infrastructure*, Policy & internet, Vol 10, No 4, 2018, pp. 415-453; D. Mac Sithigh and M. Siems, 'The Chinese social credit system: A model for other countries?', The Modern Law Review, Vol 82, No 6, 2019, pp. 1034-1071.

Lippmann, R. (1987). *An introduction to computing with neural nets*. IEEE ASSP Magazine, 4(2), 4-22. https://doi.org/10.1109/MASSP.1987.1165576

Liu, B. (2021). D*eep learning for natural language processing: A comprehensive guide to understanding deep learning for natural language processing*. Retrieved from https://link.springer.com/content/pdf/10.1007/978-981-16-3346-1.pdf

Manyika, J. (2022). *Getting AI Right: Introductory Notes on AI & Society*. *Daedalus*, *151*(2),5-27. https://www.jstor.org/stable/48662023

Meyer, D. (2000). *Generalized additive models: An introduction with R*. Retrieved from https://link.springer.com/content/pdf/10.1023/A:1007673816718.pdf

Mohan, V., & B. Siddhartha. (2018). *Text mining: Open source tokenization tools - An analysis*. Retrieved from https://www.researchgate.net/profile/Vijayarani-Mohan/publication/329800669_Text_Mining_Open_Source_Tokenization_Tools_An_Analysis/links/5e4d03d4299bf1cdb935885a/Text-Mining-Open-Source-Tokenization-Tools-An-Analysis.pdf

Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, Chris Kuhlman, Achla Marathe, Liang Zhao, Ting Hua, Feng Chen, Chang-Tien Lu, Bert Huang, Aravind Srinivasan, Khoa Trinh†, Lise Getoor, Graham Katz, Andy Doyle, Chris Ackermann, Ilya Zavorin, Jim Ford, Kristen Summers, Youssef Fayed, Jaime Arredondo, Dipak Gupta, David Mares. '*Beating the News' with EMBERS: Forecasting Civil Unrest using Open Source Indicators*. 24 August 2014. https://people.cs.vt.edu/naren/papers/kddindg1572-ramakrishnan.pdf (accessed 1 Oct. 2018).

Non-Decisional Statement by the National AI Advisory Committee (NAIAC), Working Group on Regulation and Executive Action. *Rationales, Mechanisms, and Challenges to Regulating AI: A Concise Guide and Explanation* https://ai.gov/wp-content/uploads/2023/07/Rationales-Mechanisms-Challenges-Regulating-AI-NAIAC-Non-Decisional.pdf

National Counter Terrorism Authority Pakistan (NACTA). (2021). *Information revolution and cyber warfare: Role of artificial intelligence in combating terrorist propaganda*. https://nacta.gov.pk/wp-content/uploads/2021/09/Information-Revolution-and-Cyber-Warfare-Role-of-Artificial-Intelligence-in-Combating-Terrorist-Propaganda.pdf

M. Naz, S., & Shakil, M. (2019). *Searching for extremist content online using the dark crawler and sentiment analysis*. Advances in Information Technology and Computational Science, 24(16). https://doi.org/10.1108/S1521-613620190000024016

Netlaw.bg. (n.d.). *Artificial intelligence to counter cyber-terrorism*. Retrieved October 7, 2024, from https://www.netlaw.bg/p/p/a/paper-ai-to-counter-cyber-terrorism-3657.pdf

Nicholas, G., *Shedding Light on Shadowbanning*, April 2022, Centre for democracy and technology. https://www.researchgate.net/publication/360732590_Shedding_Light_on_Shadowbanning

Noble's, S. U. (2019). Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018. why popular culture matters, 166. https://d1wqtxts1xzle7.cloudfront.net/58938056/PCSJ_Book_Review.pdf?1555544782=&response-content-disposition=inline%3B+filename%3DReview_Benacka_Elizabeth_Rhetoric_Humor.pdf&Expires=1729676916&Signature=cpogNdWrMcQexB7zenjVikXzpceS15swFyW-T9O3vr0BfBFD78nZdsptQJ5oXR7A3rsJzkhlRNEyxZP07~fKetqsdyzoCWHPEBsz-MGr18UfaKBNpR4ykvarjsuo0Syqj5HUOQFx2rrj~PMf6B8Bu3F0OdCXJiAtzTFDahr08FDh1GwkrviyvioXo0gWvewzfhTfurqOy4WyFmaJh5Ld49-xHMFnD5emkUZSLub0NALVMYX3NKydzcvNAsODyWkkmi6d6YgnEVscX3ypTI0GW7vhJcJVZoQmzdHKKbTeszlc9SUT8Dj~PLPTQgnW7e9gfpo9ZV8T~WX2h5tLv4l0gQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Odom, M. and Sharda, R. 1990, "*A neural network model for bankruptcy prediction*," Proceedings of the IEEE International Conference on Neural Networks, San Diego, California, pp .163-168. https://ieeexplore.ieee.org/abstract/document/5726669

Osservatori Digital Innovation. (n.d.). *Deep learning: Significato, esempi e applicazioni. Osservatori Digital Innovation*. Retrieved from https://blog.osservatori.net/it_it/deep-learning-significato-esempi-applicazioni

Osservatori Digital Innovation. (2023). *Natural Language Processing (NLP): Come funziona l’elaborazione del linguaggio naturale*. Osservatori.net. https://blog.osservatori.net/it_it/natural-language-processing-nlp-come-funziona-lelaborazione-del-linguaggio-naturale

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press. https://books.google.it/books?hl=it&lr=&id=ll3rBQAAQBAJ&oi=fnd&pg=PP8&dq=frank+Pasquale+the+black+box+society:+the+secret+algorithms+that+control+money+and+information&ots=Yf-2HxT6D4&sig=LBBbMbq4L6U_7qUy8J0FN7SdeNc&redir_esc=y#v=onepage&q=frank%20Pasquale%20the%20black%20box%20society%3A%20the%20secret%20algorithms%20that%20control%20money%20and%20information&f=false

Pauwels, E. (2020). *Artificial Intelligence and Data Capture Technologies in Violence and Conflict Prevention: Opportunities and Challenges for the International Community*. Global Center on Cooperative Security. http://www.jstor.org/stable/resrep27551

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2022). *Video-based Human Action Recognition using Deep Learning: A Review*. arXiv. https://doi.org/10.48550/arXiv.2208.03775

Pratt, A. N. (2011). T*errorism’s Evolution: Yesterday, Today, and Forever*. Connections, *10*(2), 1–34. http://www.jstor.org/stable/26310647

RALPH THIELE, *Artificial Intelligence – A key enabler of hybrid warfare*. March 2020, Hybrid CoE Working Paper 6, COI STRATEGY & DEFENCE. https://www.hybridcoe.fi/wp-content/uploads/2020/07/WP-6_2020_rgb-1.pdf

Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R. P., Saraf, P., & Wang, W. (2014). ‘*Beating the News’ with EMBERS: Forecasting Civil Unrest using Open Source Indicators*. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and

Rand Corporation. (n.d.). *Biometrics for identification and verification in counterterrorism*. Retrieved from https://www.rand.org/content/dam/rand/www/external/congress/terrorism/phase1/biometrics.pdf

RAND Testimony: Ochmanek, D. (2006). *Military operations against terrorist groups abroad: Implications for the United States Air Force* (Testimony No. CT-262). RAND Corporation. https://www.rand.org/content/dam/rand/pubs/testimonies/2006/RAND_CT262-1.pdf

Richbourg, R. F. (2018). *Deep Learning: Measure Twice, Cut Once*. Institute for Defense Analyses. http://www.jstor.org/stable/resrep36394

Rineheart, J. (2010). *Counterterrorism and Counterinsurgency*. Perspectives on Terrorism, 4(5), 31–47. http://www.jstor.org/stable/26298482

Richemond-Barak, D., & Feinberg, A. (2015). *The irony of the Iron Dome: intelligent defense systems, law, and security*. Harv. Nat'l Sec. J., 7, 469. https://d1wqtxts1xzle7.cloudfront.net/47704271/Irony_of_the_Iron_Dome_HNSJ-libre.pdf?1470081757=&response-content-disposition=inline%3B+filename%3DThe_Irony_of_the_Iron_Dome_Intelligent_D.pdf&Expires=1729248169&Signature=Jx67g1lrWZ8-AJlnSChdA4O2GbIqDldRnifq1tuVDGzTqkPuJQoRM6~rE6shQocxBLmKXVhBCfN5xEDcLcLKpo-xm-lX2mxvQ5q-Pb~-QM2bTw7A39xVd8VfOGaty6rZ2fkFRtWa3AGYysR3oJ7T7CABxiCDRhgtM7NplYozZtOQxfaEvIYBXx5ZkjWHlZ7w6y0-E0z4H70QO1gVQHuvf-V~t6FmKBylVbBCZxbDxwk5lpeDAyBkmQ-I1W6lK7LcbkP0LC3upmbfaqPOQqlyIOFO7nUu3hsNQ5S~5Lc5NUPxdN3SSm2oyN~-YV8d7d7HBW3Qun-kECqfy2KonQlUtA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Roche, G. and Leibold, J. (2022), *State Racism and Surveillance in Xinjiang (People's Republic of China)*. The Political Quarterly, 93: 442-450. https://doi.org/10.1111/1467-923X.13149

5. R. Roberts, 'T*he biopolitics of China's "war on terror" and the exclusion of the Uyghurs'*, Critical Asian Studies, Vol 50, No 2, 2018, pp. 232-258. https://es.scribd.com/document/404121659/The-Biopolitics-of-China-s-War-on-Terror-and-the-Exclusion-of-the-Uyghurs

Sipione, G. (n.d.). *Analisi delle dinamiche del terrorismo internazionale e strategie di contrasto*. Ministero dell'Interno. https://culturaprofessionale.interno.gov.it/FILES/docs/1260/TESTO%20INTEGRALE%20Sipione.pdf

Shams, A. (2023). *Right of Privacy and the Growing Scope of Artificial Intelligence*. Current Trends in Law and Society, Volume 3, Number 1, 2023, Pages 1 - 11 Retrieved from https://www.researchgate.net/publication/374045442_Right_of_Privacy_and_the_Growing_Scope_of_Artificial_Intelligence/fulltext/650af9f1c05e6d1b1c1ef58e/Right-of-Privacy-and-the-Growing-Scope-of-Artificial-Intelligence.pdf?origin=publicationDetail

Shukla, P. P., & Hasan, M. M. (2021). *A review of the ethical challenges in artificial intelligence systems*. Journal of Information Technology & Politics, 19(1), 26-46. https://doi.org/10.1177/15485129211073126

B, S., Siddhartha. (2021). *An interpretation of lemmatization and stemming in natural language processing*. Retrieved from https://www.researchgate.net/profile/Siddhartha-B-S/publication/348306833_An_Interpretation_of_Lemmatization_and_Stemming_in_Natural_Language_Processing/links/6048467f299bf1e078696a3a/An-Interpretation-of-Lemmatization-and-Stemming-in-Natural-Language-Processing.pdf

Silver, J. (2016, February 18). *Has a rampaging AI algorithm really killed thousands in Pakistan*? The Guardian. https://www.theguardian.com/science/the-lay-scientist/2016/feb/18/has-a-rampaging-ai-algorithm-really-killed-thousands-in-pakistan

Slobogin, C. (2010). *GOVERNMENT DRAGNETS*. Law and Contemporary Problems, 73(3), 107–143. http://www.jstor.org/stable/25766402

Stanford University. (2016). *One hundred year study on artificial intelligence: Report of the 2016-2017 study panel*. Retrieved from https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. Pp. 2-10 https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

Töws, M. (2023). *Artificial Intelligence and Human Rights: A Human-Centric Approach to AI Governance* [Master's thesis, University of Tübingen]. Retrieved from https://tobias-lib.ub.uni-tuebingen.de/xmlui/bitstream/handle/10900/65097/2301-volledige-tekst_tcm45-535187.pdf?sequence=1&isAllowed=y

Trend Micro Reserach, United Nationas Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), *Malicious Uses and Abuses of Artificial Intelligence*, https://documents.trendmicro.com/assets/white_papers/wp-malicious-uses-and-abuses-of-artificial-intelligence.pdf

TÜR, M. R. (2022). *Energy Supply Security and Artificial Intelligence Applications*. Insight Turkey, 24(3), 213–234. https://www.jstor.org/stable/48733385

United Nations Counter-Terrorism Centre & United Nations Interregional Crime and Justice Research Institute. (2021). **Countering terrorism online with artificial intelligence: Opportunities and challenges*.* United Nations Office of Counter-Terrorism. https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf

A.Turing. *Macchine calcolatrici e intelligenza* (1950) https://disf.org/files/macchine-calcolatrici-e-intelligenza.pdf

United Nations. (2021). *Analytical brief on biometrics*. Retrieved from https://www.un.org/securitycouncil/ctc/sites/www.un.org.securitycouncil.ctc/files/files/documents/2021/Dec/cted_analytical_brief_biometrics_0.pdf

US Department of Defense (DOD), Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy* (DOD: Washington, DC, June 2016), p. 61. https://apps.dtic.mil/sti/pdfs/AD1017790.pdf

United Nations Office of Counter-Terrorism, UN Counter-Terrorism Centre (UNCCT), & United Nations Interregional Crime and Justice Research Institute (UNICRI). (2021). *Countering terrorism online with artificial intelligence: An overview for law enforcement and counter-terrorism agencies in South Asia and South-East Asia*.

United Nations Office of Counter-Terrorism, UN Counter-Terrorism Centre (UNCCT), & United Nations, *Algorithms an terrorism: the malicious use of artificial Intelligence for terrorist purposes*. https://unicri.it/sites/default/files/2021-06/Malicious%20Use%20of%20AI%20-%20UNCCT-UNICRI%20Report_Web.pdf

Valenti G., Annovi C., Di Liddo M., Definire il terrorismo per supportare la prevenzione e il contrasto alla radicalizzazione. Giugno 2023. https://www.esteri.it/wp-content/uploads/2023/09/CeSI_Definire_il_terrorismo.pdf

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). *Attention is All you Need. Neural Information Processing Systems*. https://arxiv.org/pdf/1706.03762

World Health Organization. (2021). *Artificial intelligence is changing the health sector*. In WHO Consultation Towards the Development of guidance on ethics and governance of artificial intelligence for health: Meeting report Geneva, Switzerland, 2–4 October 2019 (pp. 3–7). World Health Organization. http://www.jstor.org/stable/resrep35680.7

K. L. X. Wong and A. S. Dobson, *We're just data: Exploring China's social credit system in relation to digital platform ratings cultures in Westernised democracies*', Global Media and China, Vol 4, No 2, 2019, pp. 220-232. https://journals.sagepub.com/doi/pdf/10.1177/2059436419856090